CAUCHY INTEGRAL AND CALDERÓN-ZYGMUND OPERATORS ON NONHOMOGENEOUS SPACES

F. NAZAROV, S. TREIL, AND A. VOLBERG

1. Preliminaries

The goal of this paper is to consider the boundedness of singular integral operators with Calderón-Zygmund kernels in $L^2(\mu)$, where μ is an arbitrary measure without atoms in $\mathbb{R}^n$. We learned about this problem from Mark Melnikov.

We concentrate on the case $n = 2$ for the sake of simplicity only. Let μ be a positive measure without atoms on the complex plane $\mathbb{C}$. We assume that the measure μ is regular, i.e.,

$$\mu(D_r) \leq Cr$$

for any disk D_r of radius r. We call the attention of the reader to the fact that the above condition is necessary for the boundedness in $L^2(\mu)$ of a wide class of singular integral operators with Calderón-Zygmund kernels, see [1]. In particular this is so for the Cauchy integral operator $\mathcal{C} = \mathcal{C}_\mu$, with which we are going to start.

The paper continues the line of research started in [2, 3], where we treated Calderón–Zygmund operators in non-classical situations.

Acknowledgment: The authors are very grateful to Mike Frazier for reading this paper and making many valuable remarks. Without his kind help the text would be much less comprehensible for an inexperienced reader.

The Cauchy integral operator $\mathcal{C} = \mathcal{C}_\mu$ is defined by

$$(\mathcal{C}f)(z) = \int \frac{f(\xi)}{\xi - z} d\mu(\xi) .$$

One of the main results of the paper (see Theorem 1.1 below) is that the Cauchy integral operator is bounded in $L^2(\mu)$ if and only if it is bounded on characteristic functions of squares.

The question arises of how to interpret the operator $\mathcal{C}_\mu$, because it is not clear why the integral converges in any reasonable sense for $z \in \operatorname{supp} \mu$.

One way to understand the result is as follows. Consider the family of the truncated operators $\mathcal{C}^\varepsilon$,

$$(\mathcal{C}^\varepsilon f)(z) = \int\limits_{\varepsilon < |\xi - z| < 1/\varepsilon} \frac{f(\xi)}{\xi - z} d\mu(\xi)$$

(clearly, there is no problem interpreting $\mathcal{C}^\varepsilon$, and all $\mathcal{C}^\varepsilon$ are bounded operators in $L^2(\mu)$). And the result is that all $\mathcal{C}^\varepsilon$ are uniformly (in ε) bounded if and only if they are uniformly bounded on characteristic functions of squares.

For technical reasons it is more convenient to perform another "regularization" in such a fashion that the resulting regularized operators are Calderón–Zygmund operators. For example, one can define $\mathcal{C}^\varepsilon$ as the convolution with the kernel

$$K_\varepsilon(re^{i\theta}) = \begin{cases} r^{-1}e^{-i\theta}, & \varepsilon \le r \le 1/\varepsilon; \\ \varepsilon^{-1}e^{-i\theta}, & 0 \le r \le \varepsilon; \\ (2\varepsilon - \varepsilon^2 r)e^{-i\theta}, & 1/\varepsilon \le r \le 2/\varepsilon; \\ 0, & r > 2/\varepsilon \end{cases}$$

(see Figure 1).

Clearly (under the assumption that the measure μ is regular, i.e., $\mu(D_r) \le Cr$, which we have throughout the paper), the above regularized operators are uniformly bounded if and only if the truncated operators are[1]. So it does not matter which operators $\mathcal{C}^\varepsilon$ to consider, and it is technically more convenient to consider the second type of regularization.

Note that the result holds for any "reasonable" regularization (we do not want to specify here what does "reasonable" mean) of $\mathcal{C}$. For example, one can pick as $\mathcal{C}^\varepsilon$ the convolution with the kernel

$$K_\varepsilon(z) = \frac{\overline{z}}{|z|^2 + \varepsilon} - \frac{\overline{z}}{|z|^2 + 1/\varepsilon}.$$

Another way to interpret the result is to define the operator $\mathcal{C}$ axiomatically, similarly to the way it is usually done in the theory of Calderón–Zygmund operators. Namely, one can consider an abstract operator $\mathcal{C}$ defined at least on characteristic functions of squares and antisymmetric ($\langle \mathcal{C}f, g\rangle = -\langle f, \mathcal{C}g\rangle$) with respect to the standard *linear* duality $\langle f, g\rangle = \int fg d\mu$ (this is not a typo, there is no complex conjugation here).

We also assume that the operator has the kernel $1/(\xi - z)$ in the sense that for any functions f and g with disjoint compact supports,

$$\langle \mathcal{C}f, g\rangle = \iint \frac{f(\xi)g(z)}{\xi - z} d\mu(\xi) d\mu(z).$$

Let χ_Q denote the characteristic function of the set Q. For a square Q let $\ell(Q)$ denote its size, i.e., the length of its side.

Now we are ready to state one of our main results. While the paper was being prepared, we learned that this theorem was independently proved by Xavier Tolsa by a completely different method. As far as we understand, the specific properties of the Cauchy kernel are used heavily in [5]. On the other hand, the reader will see from Theorem 1.2 that our approach does not use any specifics of the Cauchy kernel.

[1] to see that it is enough to apply the Schur test to the kernel of the difference

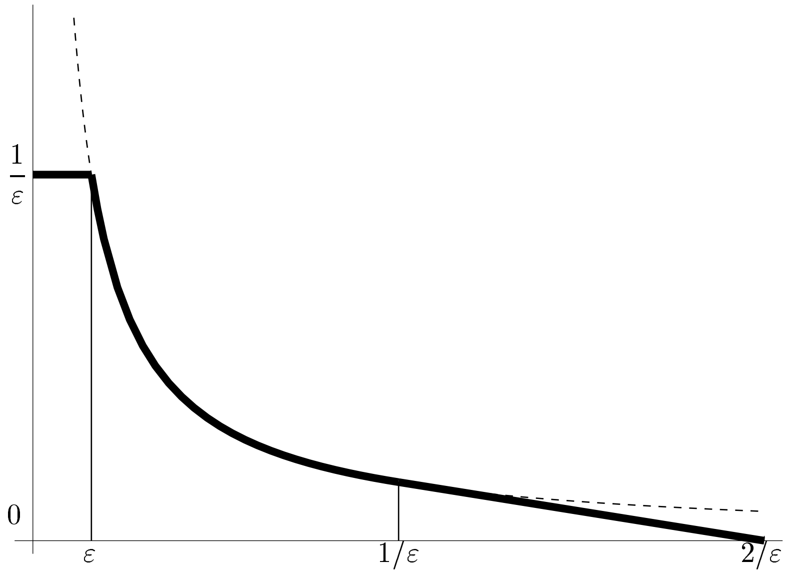

FIGURE 1. One of possible regularizations of the Cauchy kernel. The graph of $|K_\varepsilon(z)|$ vs. $|z|$ is displayed. On the interval $[1/\varepsilon, 2/\varepsilon]$ the function is linear, and the derivatives at the point $1/\varepsilon$ agree

Theorem 1.1. *Let μ be a regular measure, i.e., $\mu(Q) \le \ell(Q)$ for any square $Q \subset \mathbb{C}$. Then the following statements are equivalent*

1. *The Cauchy operator $\mathcal{C} = \mathcal{C}_\mu$ is bounded in $L^2(\mu)$;*
2. *For any square Q, the estimate $\int_{\mathbb{C}} |\mathcal{C}\chi_Q|^2 d\mu \le C\|\chi_Q\|^2_{L^2(\mu)} = C\mu(Q)$ holds;*
3. *For any square Q, the estimate $\int_{2Q} |\mathcal{C}\chi_Q|^2 d\mu \le C'\|\chi_Q\|^2_{L^2(\mu)} = C'\mu(Q)$ holds;*
4. *For any square Q, the estimate $\int_{Q} |\mathcal{C}\chi_Q|^2 d\mu \le C''\|\chi_Q\|^2_{L^2(\mu)} = C''\mu(Q)$ holds.*

Moreover, the norm of $\mathcal{C}$ can be estimated from above by a constant depending only on C (respectively on C' or C'').

Let us notice that the last condition is equivalent to Melnikov's celebrated condition on the local curvature of measure, see [7], [6].

We will prove an even more general theorem about general Calderón–Zygmund operators. Let T be a Calderón–Zygmund operator with kernel

$K(z, \xi)$:

$$Tf(z) := \int K(z, \xi) f(\xi) d\mu(\xi) .$$

We assume that the kernel satisfies the following standard assumptions:

1. $|K(s, t)| \le \dfrac{Const}{|s-t|}$;
2. There exists $\varepsilon > 0$ such that

$$|K(s, t) - K(s_0, t)|, \ |K(t, s) - K(t, s_0)| \le Const \frac{|s-s_0|^\varepsilon}{|t-s_0|^{1+\varepsilon}}$$

   whenever $|t - s_0| \ge 2|s - s_0|$.

Again as with the Cauchy operator $\mathcal{C}$, there are two ways to interpret the operator T. One is to assume that the operator T is a "limit"[2] of "nice" operators T_ε with "nice" kernels ("nice" here means that the operators T_ε are bounded and the integrals are well defined), and we are interested in getting a uniform bound on the norms of the operators T_ε. Another way is again to define the operator T axiomatically as an abstract operator defined at least on characteristic functions of squares with the kernel K in the sense that for functions f and g with disjoint compact supports

$$\langle Tf, g\rangle = \iint K(z, \xi) f(\xi) g(z) d\mu(\xi) d\mu(z).$$

For a Calderón–Zygmund operator T let T' be its adjoint with respect to the duality $\langle f, g\rangle = \int fg d\mu$, i.e. let T' be an operator with kernel $K'(\xi, z) := K(z, \xi)$. Here if we have a family of "nice" operators T_ε there is no question how to define the adjoint ones. If the operator T is defined axiomatically, we know only its bilinear form (at least on linear combinations of characteristic functions of squares), and that is enough for our purposes.

The next theorem holds in $\mathbb{R}^n$. We prove it in $\mathbb{R}^2$ only for the sake of simplicity.

Theorem 1.2. *Let μ be a regular measure, i.e., $\mu(Q) \le \ell(Q)$ for any square $Q \subset \mathbb{C}$. Then the following statements are equivalent:*

1. *The operator T is bounded in $L^2(\mu)$;*
2. *For any square Q, the estimates $\int_{\mathbb{C}} |T\chi_Q|^2 d\mu \le C\|\chi_Q\|^2_{L^2(\mu)} = C\mu(Q)$ and $\int_{\mathbb{C}} |T'\chi_Q|^2 d\mu \le C\|\chi_Q\|^2_{L^2(\mu)} = C\mu(Q)$ hold;*
3. *For any square Q, the estimates $\int_{2Q} |T\chi_Q|^2 d\mu \le C'\|\chi_Q\|^2_{L^2(\mu)} = C'\mu(Q)$ and $\int_{2Q} |T'\chi_Q|^2 d\mu \le C'\|\chi_Q\|^2_{L^2(\mu)} = C'\mu(Q)$ hold;*
4. *For any square Q, the estimates $\int_Q |T\chi_Q|^2 d\mu \le C''\|\chi_Q\|^2_{L^2(\mu)} = C''\mu(Q)$ and $\int_Q |T'\chi_Q|^2 d\mu \le C''\mu(Q)$ hold, and for any two squares Q and R with disjoint interiors,*

$$\langle T\chi_Q, \chi_R\rangle \le C''\|\chi_Q\|_{L^2(\mu)}\|\chi_R\|_{L^2(\mu)} = C''\mu(Q)^{1/2}\mu(R)^{1/2}.$$

[2] We do not assume here that the limit exists.

Moreover, the norm of T can be estimated from above by a constant depending only on C (respectively on C' or C'') and the parameters in the definition of the Calderón–Zygmund kernel.

Note that trivially $1 \Longrightarrow 2 \Longrightarrow 3 \Longrightarrow 4$. Also it is not difficult to show that $3 \Longrightarrow 2$.

Indeed, property 1 of Calderón–Zygmund kernels implies that $|T\chi_Q(z)| \leq C\mu(Q)/\operatorname{dist}(z,Q)$, and therefore

$$\begin{aligned}
\int\limits_{\mathbb{C}\setminus 2Q} |T\chi_Q|^2 d\mu &\leq \int\limits_0^{K\mu(Q)/\ell(Q)} \mu(\{z \in \mathbb{C}\setminus Q : |T\chi_Q(z)| > t\}) d(t^2) \\
&\leq \int\limits_0^{K\mu(Q)/\ell(Q)} \mu(\left\{z : \operatorname{dist}(z,Q) < \frac{C\mu(Q)}{t}\right\}) 2t dt \\
&\leq \int\limits_0^{K\mu(Q)/\ell(Q)} \frac{C\mu(Q)}{t} 2t dt \leq C\frac{\mu(Q)^2}{\ell(Q)} \leq C\mu(Q).
\end{aligned}$$

So it does not matter whether we integrate $|T\chi_Q|^2$ over the whole complex plane $\mathbb{C}$ or only over the square $2Q$.

To prove Theorem 1.2 we need only to show that $4 \Longrightarrow 1$. Theorem 1.1 will follow from it if we show that condition 4 of Theorem 1.1 implies condition 4 of Theorem 1.2. This is given by the following lemma.

Lemma 1.3. *Let the measure μ be regular and Q and R be two disjoint squares. Then*

$$\left|\langle \mathcal{C}\chi_Q, \chi_R\rangle\right| \leq C\mu(Q)^{1/2}\mu(R)^{1/2}.$$

Proof. Without loss of generality one can assume that the square Q lies in the (closed) lower half plane $\operatorname{Im} z \leq 0$ and that R lies in the open upper half plane $\operatorname{Im} z > 0$.

Note that the restrictions of the measure μ to the squares Q and R are Carleson measures for the lower and upper half planes, respectively[3].

We claim that the function $f := \mathcal{C}\chi_Q$ belongs to the Hardy space H^2 on the upper half plane. Indeed, f is clearly analytic in the open upper half plane. Take any function g on the line $i\varepsilon + \mathbb{R}$ satisfying $\int_{-\infty}^{\infty} |g(x+i\varepsilon)|^2 dx \leq$

[3]There is a little detail here. Formally, Carleson measures live inside the domain, and in our case there could be some measure on the real line. But the regularity condition implies that the part of the measure living on the real line is majorized by the 1-dimensional Lebesgue measure, and therefore for measures $\mu|_Q$ and $\mu|_R$ the Carleson Embedding Theorem holds.

1. Then

$$\int_{-\infty}^{\infty} f(x+i\varepsilon)g(x+i\varepsilon)dx = \int_{-\infty}^{\infty}\Big(\int_Q \frac{1}{\xi-(x+i\varepsilon)}d\mu(\xi)\Big)g(x+i\varepsilon)dx$$
$$= \int_Q\Big(\int_{-\infty}^{\infty}\frac{g(x+i\varepsilon)}{\xi-(x+i\varepsilon)}dx\Big)d\mu(\xi).$$

The function

$$G(\xi) := \int_{-\infty}^{\infty}\frac{g(x+i\varepsilon)}{\xi-(x+i\varepsilon)}dx$$

belongs to the Hardy class H^2 on the lower half plane $\operatorname{Im} z < 0$, and its norm is bounded by an absolute constant. Therefore

$$\Big|\int_{-\infty}^{\infty} f(x+i\varepsilon)g(x+i\varepsilon)dx\Big| = \Big|\int_Q G(\xi)d\mu(\xi)\Big|$$
$$\le \Big(\int_Q |G(\xi)|^2 d\mu(\xi)\Big)^{1/2}\mu(Q)^{1/2} \le C\mu(Q)^{1/2}$$

(the last inequality follows from the Carleson Embedding Theorem, see [4]).

Therefore the function f belongs to H^2 on the upper half plane, and its norm is bounded by $C\mu(Q)^{1/2}$. Applying again the Carleson Embedding Theorem, we get

$$|\langle \mathcal{C}\chi_Q, \chi_R\rangle| = \Big|\int_R f(z)d\mu(z)\Big|$$
$$\le \Big(\int_R |f(z)|^2 d\mu(z)\Big)^{1/2}\mu(R)^{1/2} \le C\mu(Q)^{1/2}\mu(R)^{1/2}.$$

□

Remark. To prove Lemma 1.3 for the truncated Cauchy operator $\mathcal{C}^\varepsilon$,

$$\mathcal{C}^\varepsilon f(z) = \int_{|\xi-z|>\varepsilon}\frac{f(\xi)}{\xi-z}d\mu(\xi)$$

it is convenient to consider another regularization $\mathcal{C}_\varepsilon$ of the operator $\mathcal{C}$ given by

$$\mathcal{C}_\varepsilon f(z) = \int\limits_{\operatorname{Im}\xi\le 0}\frac{f(\xi)}{\xi-z-i\varepsilon}d\mu(\xi), \qquad \operatorname{Im} z > 0$$

(the operator maps functions supported by the closed lower half plane $\overline{\mathbb{C}}_-$ to functions on the open upper half plane $\mathbb{C}_+$). Applying the Schur test, one can easily see that the difference $\mathcal{C}^\varepsilon - \mathcal{C}_\varepsilon$ is a bounded operator from the $L^2(\overline{\mathbb{C}}_-;\mu)$ to $L^2(\mathbb{C}_+;\mu)$. And to prove the lemma for $\mathcal{C}_\varepsilon$ we just need to move the square R up by ε and apply the lemma for $\mathcal{C}$.

Applying Lemma 1.3 for $\mathcal{C}^\varepsilon$ and then for $\mathcal{C}^{1/\varepsilon}$, we see that it holds for the differences $\mathcal{C}^\varepsilon - \mathcal{C}^\varepsilon$, which are exactly the truncated operators considered in the beginning of the paper (where we integrated over the set $\varepsilon < |\xi - z| < 1/\varepsilon$).

2. Martingale difference decomposition of $L^2(\mu)$ and the matrix of a Calderón–Zygmund operator

Fix a dyadic lattice $\mathcal{D}$ in $\mathbb{C}$. Just for our convenience we will consider only lattices constructed of squares with sides 2^k, $k \in \mathbb{Z}$ (we consider squares of all sizes, not only with a fixed k).

Denote by E_k the averaging operator over dyadic squares of size (length of the side) 2^k, namely $E_k f(z) = \mu(Q)^{-1} \int_Q f d\mu$, where Q is a dyadic square of size 2^k containing z (for the sake of definiteness, we consider squares of the form $[a, b) \times [c, d)$). If Q is a square of size 2^k, we denote by $E_Q f$ the restriction of $E_k f$ to Q: $E_Q f = (\mu(Q)^{-1} \int_Q f d\mu)\chi_Q = \chi_Q E_k f$.

Let $\Delta_k := E_{k-1} - E_k$. Again for a dyadic square Q of size 2^k, denote by $\Delta_Q f$ the restriction of $\Delta_k f$ to Q. Clearly for any $f \in L^2(\mu)$, the functions $\Delta_Q f$, $Q \in \mathcal{D}$, are orthogonal to each other and for any fixed n

$$\|f\|^2_{L^2(\mu)} = \sum_{Q \in \mathcal{D}, \ell(Q) \le 2^n} \|\Delta_Q f\|^2 + \sum_{Q \in \mathcal{D}, \ell(Q) = 2^n} \|E_Q f\|^2.$$

In the following two lemmas we assume that φ_Q and ψ_R are $L^2(\mu)$-functions supported by squares Q and R, respectively, and orthogonal to constants:

$$\int \varphi_Q d\mu = \int \psi_R d\mu = 0.$$

We also assume the following normalization of φ_Q and ψ_R:

$$\|\varphi_Q\|_{L^2(\mu)} = \mu(Q)^{-1/2}, \qquad \|\psi_R\|_{L^2(\mu)} = \mu(R)^{-1/2}.$$

In the future we plan to apply the lemmas below to the corresponding normalizations of the functions $\Delta_Q f$ and $\Delta_R g$, where $f, g \in L^2(\mu)$.

Lemma 2.1. *Let Q, R be two squares, $\ell(Q) \le \ell(R)$, and $d := \operatorname{dist}(Q, R) > 0$ (here we do not assume that the squares belong to the same dyadic lattice). Let $\varphi_Q, \psi_R \in L^2(\mu)$ be functions supported by the squares Q and R, respectively, and normalized by $\|\varphi_Q\|_{L^2(\mu)} = \mu(Q)^{-1/2}$, $\|\psi_R\|_{L^2(\mu)} = \mu(R)^{-1/2}$. Suppose also that φ_Q is orthogonal to constants. Then*

$$\left|\langle T\varphi_Q, \psi_R\rangle\right| \le C \frac{\ell(Q)^\varepsilon}{d^{1+\varepsilon}}.$$

Proof. First of all, notice that the Cauchy inequality implies $\|\varphi_Q\|_{L^1(\mu)} \le \|\varphi_Q\|_{L^2(\mu)}\mu(Q)^{1/2} = 1$, and similarly $\|\psi_R\|_{L^1(\mu)} \le 1$. Now we get

$$\begin{aligned}\left|\langle T\varphi_Q, \psi_R\rangle\right| &= \left|\iint K(t,s)\varphi_Q(s)\psi_R(t)d\mu(s)d\mu(t)\right| \\ &= \left|\iint \left[K(t,s) - K(t,s_0)\right]\varphi_Q(s)\psi_R(t)d\mu(s)d\mu(t)\right| \\ &\le C\iint \frac{|s-s_0|^\varepsilon}{|t-s_0|^{1+\varepsilon}}|\varphi_Q(s)|\cdot|\psi_R(t)|d\mu(s)d\mu(t) \\ &\le C\frac{\ell(Q)^\varepsilon}{d^{1+\varepsilon}}\|\varphi_Q\|_{L^1(\mu)}\cdot\|\psi_R\|_{L^1(\mu)} \le C\frac{\ell(Q)^\varepsilon}{d^{1+\varepsilon}}.\end{aligned}$$

□

Definition 2.2. Let $\alpha = \varepsilon/(2\varepsilon+2)$, and so $\alpha + \alpha\varepsilon = \varepsilon/2$. Let r be some positive integer to be fixed later. Consider a pair of disjoint squares Q and R. Suppose for definiteness that $\ell(Q) \le \ell(R)$. We will call this pair *singular* if $\operatorname{dist}(Q,R) \le \ell(Q)^\alpha\ell(R)^{1-\alpha}$, and *essentially singular* if, in addition, $\ell(Q) \le 2^{-r}\ell(R)$.

Definition 2.3. Let $D(Q,R)$ denote the so-called *long distance* between squares:

$$D(Q,R) := \operatorname{dist}(Q,R) + \ell(Q) + \ell(R)\,.$$

Lemma 2.4. *Let T be a Calderón–Zygmund operator and let $\varphi_Q, \psi_R \in L^2(\mu)$ be functions supported by the squares Q and R, respectively, and normalized by $\|\varphi_Q\|_{L^2(\mu)} = \mu(Q)^{-1/2}$, $\|\psi_R\|_{L^2(\mu)} = \mu(R)^{-1/2}$. Suppose also that $\ell(Q) \le \ell(R)$ and that φ_Q is orthogonal to constants. Then*

$$\left|\langle T\varphi_Q, \psi_R\rangle\right| \le C\frac{\ell(Q)^{\varepsilon/2}\ell(R)^{\varepsilon/2}}{D(Q,R)^{1+\varepsilon}},$$

provided that $Q \cap R = \varnothing$ and the pair Q, R is not singular.

Proof. If $\operatorname{dist}(Q,R) \ge \ell(R)$, then $D(Q,R) \le 3\operatorname{dist}(Q,R)$; thus, the estimate from Lemma 2.1 implies

$$\left|\langle T\varphi_Q, \psi_R\rangle\right| \le C\frac{\ell(Q)^\varepsilon}{D(Q,R)^{1+\varepsilon}} \le C\cdot\frac{\ell(Q)^{\varepsilon/2}\ell(R)^{\varepsilon/2}}{D(Q,R)^{1+\varepsilon}}.$$

Now let us suppose that $\operatorname{dist}(Q,R) \le \ell(R)$, but the pair Q, R is not singular. That means

$$\operatorname{dist}(Q,R) \ge \ell(Q)^\alpha\ell(R)^{1-\alpha}.$$

The estimate of Lemma 2.1 and the identity $\alpha + \alpha\varepsilon = \varepsilon/2$ imply

$$\left|\langle T\varphi_Q, \psi_R\rangle\right| \le \frac{C\cdot\ell(Q)^\varepsilon}{\ell(Q)^{\varepsilon/2}\ell(R)^{1+\varepsilon/2}} = \frac{C\cdot\ell(Q)^{\varepsilon/2}\ell(R)^{\varepsilon/2}}{\ell(R)^{1+\varepsilon}} \le C\frac{\ell(Q)^{\varepsilon/2}\ell(R)^{\varepsilon/2}}{D(Q,R)^{1+\varepsilon}}.$$

□

Note that if we do not normalize the functions φ_Q and ψ_R, the estimate from Lemma 2.4 can be rewritten as

$$\left|\langle T\varphi_Q, \psi_R\rangle\right| \le C\frac{\ell(Q)^{\varepsilon/2}\ell(R)^{\varepsilon/2}}{D(Q,R)^{1+\varepsilon}}\mu(Q)^{1/2}\mu(R)^{1/2}\|\varphi_Q\|_{L^2(\mu)}\|\psi_R\|_{L^2(\mu)}.$$

The following theorem shows that the matrix $\{T_{Q,R}\}_{Q,R\in\mathcal{D}}$ defined by

$$T_{Q,R} := \frac{\ell(Q)^{\varepsilon/2}\ell(R)^{\varepsilon/2}}{D(Q,R)^{1+\varepsilon}}\mu(Q)^{1/2}\mu(R)^{1/2},$$

generates a bounded operator in ℓ^2.

Theorem 2.5. *Let the measure μ satisfy $\mu(Q) \le C\ell(Q)$ for all squares Q. Then for the matrix $\{T_{Q,R}\}_{Q,R\in\mathcal{D}}$ defined above, one has*

$$\sum_{Q,R\in\mathcal{D}} T_{Q,R}x_Q \cdot y_R \le C\Big(\sum_{Q\in\mathcal{D}} x_Q^2\Big)^{1/2}\Big(\sum_{Q\in\mathcal{D}} y_Q^2\Big)^{1/2}$$

for any sequences of non-negative numbers $\{x_Q\}, \{y_Q\} \in \ell^2$.

Proof. The symmetry of Q and R implies that it is enough to consider only the sum over Q, R such that $\ell(Q) \le \ell(R)$. So we can just assume that $T_{Q,R} = 0$ if $\ell(Q) > \ell(R)$.

Let us "slice" the matrix $\{T_{Q,R}\}_{Q,R\in\mathcal{D}}$. Namely, for any $n = 0, 1, 2, 3, \dots$, define the matrix $\{T^{(n)}_{Q,R}\}_{Q,R\in\mathcal{D}}$ by putting

$$T^{(n)}_{Q,R} = \begin{cases} T_{Q,R}, & \ell(Q) = 2^{-n}\ell(R); \\ 0, & \text{otherwise.} \end{cases}$$

If we show that the norms of the operators $T^{(n)}$ decrease as a geometric progression, i. e., that

$$\sum_{Q,R\in\mathcal{D}} T^{(n)}_{Q,R}x_Q \cdot y_R \le 2^{-n\beta}C\left(\sum_{Q\in\mathcal{D}} x_Q^2\right)^{1/2}\left(\sum_{Q\in\mathcal{D}} y_Q^2\right)^{1/2}$$

for some $\beta > 0$, then we are done.

We can split the matrices $T^{(n)}$ into layers $T^{(n,k)}$, where

$$T^{(n,k)}_{Q,R} = \begin{cases} T^{(n)}_{Q,R}, & \ell(R) = 2^k \\ 0, & \text{otherwise} \end{cases}$$

Clearly, the layers $T^{(n,k)}$ of $T^{(n)}$ do not interfere, therefore it is enough to estimate each layer separately. So, it is enough to show that for any sequences of non-negative $x = \{x_Q\}_{Q\in\mathcal{D}}$, $y = \{y_Q\}_{Q\in\mathcal{D}} \in \ell^2$,

$$\langle T^{(n,k)}x, y\rangle = \sum_{\substack{Q,R\in\mathcal{D}\\ \ell(Q)=2^{k-n},\ell(R)=2^k}} T^{(n,k)}_{Q,R}x_Q y_R$$
$$\le 2^{-n\beta}C\Big(\sum_{\substack{Q\in\mathcal{D}\\ \ell(Q)=2^{k-n}}} x_Q^2\Big)^{1/2}\Big(\sum_{\substack{Q\in\mathcal{D}\\ \ell(Q)=2^{k}}} y_Q^2\Big)^{1/2}.$$

One can rewrite the matrix $T^{(n,k)}$ as an integral operator. Namely, if we define

$$X := \sum_{Q:\ell(Q)=2^{k-n}} \mu(Q)^{-1/2}x_Q\chi_Q, \qquad Y := \sum_{Q:\ell(Q)=2^{k}} \mu(Q)^{-1/2}y_Q\chi_Q,$$

then

$$\|X\|^2_{L^2(\mu)} = \sum_{\ell(Q)=2^{k-n}} x_Q^2, \qquad \|Y\|^2_{L^2(\mu)} = \sum_{\ell(Q)=2^{k}} y_Q^2.$$

Now the estimate we need can be rewritten as

$$\sum_{\substack{\ell(Q)=2^{k-n}\\ \ell(R)=2^k}} T^{(n,k)}_{Q,R}x_Q\cdot y_R = \iint K_k^{(n)}(s,t)X(s)\cdot Y(t)d\mu(s)d\mu(t)$$
$$\le C\|X\|_{L^2(\mu)}\|Y\|_{L^2(\mu)},$$

where the kernel $K_k^{(n)}(s,t)$ is defined by

$$K_k^{(n)}(s,t) = \sum_{\substack{\ell(Q)=2^{k-n}\\ \ell(R)=2^k}} T_{Q,R}\mu(Q)^{-1/2}\mu(R)^{-1/2}\chi_Q(s)\chi_R(t).$$

Note that for each pair s,t, the sum has only one non-zero term, so the kernel $K_k^{(n)}(s,t)$ can be easily estimated:

$$K_k^{(n)}(s,t) \le C2^{-n\varepsilon/2}\cdot\frac{2^{k\varepsilon}}{(2^k+|t-s|)^{1+\varepsilon}} = C2^{-n\varepsilon/2}\mathcal{K}_k(t-s),$$

where $\mathcal{K}_k(s) = 2^{k\varepsilon}/(2^k+|s|)^{1+\varepsilon}$. Integrating the distribution function, one can show that

$$\int \mathcal{K}_k(s)d\mu(s) \le Const.$$

So, by the Schur Lemma the integral operators with kernels $\mathcal{K}_k(s)$ are uniformly bounded, therefore the norms of the operators $T^{(n,k)}$ (and hence of $T^{(n)}$) decrease as a geometric progression, and we are done. $\square$

3. Paraproducts and the estimate of $\langle T\varphi_Q, \psi_R\rangle$ when $Q \subset R$

As usual in the theory of singular integral operators to estimate $\langle T\varphi_Q, \psi_R\rangle$ when $Q \subset R$, one can use the so-called *paraproducts*. The classical construction will not work in our case and we will slightly modify it.

3.1. **Paraproducts.** Let r be a positive integer to be defined later (it is the same number we used in the definition of essentially singular pairs, see Definition 2.2). We define a *paraproduct* $\Pi = \Pi_{T'}$ by

$$\Pi f := \sum_{R\in\mathcal{D}} \sum_{\substack{Q\subset R:\ell(Q)=2^{-r}\ell(R)\\ \partial Q\cap\partial R=\varnothing}} E_R f \cdot \Delta_Q T'\mathbf{1}.$$

If we are working with a "nice" operator T, then $T'\mathbf{1}$ is well defined. Note that even if $T'\mathbf{1}$ is not well defined, we still can define $\Delta_Q T'\mathbf{1}$ by duality as the function f satisfying

$$\langle f, g\rangle = \langle \mathbf{1}, T\Delta_Q g\rangle \qquad \forall g \in L^2(\mu).$$

Let us study the matrix of Π. Let $Q, R \in \mathcal{D}$. Let φ_Q and ψ_R be functions supported by Q and R, constant on each of the four smaller subsquares and orthogonal to constants.

Notice that $\langle \varphi_Q, \Pi\psi_R\rangle$ is non-zero only if $Q \subset R$, $\ell(Q) < 2^{-r}\ell(R)$, and $\partial Q \cap \partial R' = \varnothing$, where $R' \subset R$ is the square of size $2^r\ell(Q)$ containing Q.

In this case

$$\langle \varphi_Q, \Pi\psi_R\rangle = \langle \varphi_Q, \Delta_Q T'\mathbf{1}\rangle\psi_R(s) = \langle T\varphi_Q, \mathbf{1}\rangle\psi_R(s) \tag{3.1}$$

where s is an arbitrary point in the square R'. Note that instead of R' we can take here the dyadic square of size $\ell(R)/2$ that contains Q, because ψ_R is constant on squares of size $\ell(R)/2$.

Theorem 3.1. *If for any square Q*

$$\int_Q |T'\chi_Q|^2 d\mu \le C\mu(Q),$$

then the paraproduct Π is bounded in $L^2(\mu)$.

To prove the theorem, we need the following lemma.

Lemma 3.2. *If for any square Q*

$$\int_Q |T'\chi_Q|^2 d\mu \le C\mu(Q),$$

then the following "Carleson measure condition"

$$\sum_{\substack{Q\subset R\\ \ell(Q)\le 2^{-r}\ell(R)\\ \partial Q\cap\partial R=\varnothing}} \|\Delta_Q T'\mathbf{1}\|^2_{L^2(\mu)} \le C\|\chi_R\|^2_{L^2(\mu)} = C\mu(R) \qquad \forall R \in \mathcal{D}$$

holds.

Proof. Let S be a dyadic square and let $S' = 1.1S$. Since T' is bounded on characteristic functions of squares,

$$\sum_{Q\subset S} \|\Delta_Q T'\chi_{S'}\|^2_{L^2(\mu)} \le \|T'\chi_{S'}\|^2_{L^2(\mu)} \le C\|\chi_{S'}\|^2_{L^2(\mu)} = C\mu(1.1S).$$

Let us now estimate $\sum_{Q\in S} \|\Delta_Q T'(\mathbf{1}-\chi_{S'})\|^2_{L^2(\mu)}$. Fix a square Q and let $f \in L^2(\mu)$, $\|f\| = 1$. Clearly,

$$\langle f, \Delta_Q T'(\mathbf{1}-\chi_{S'})\rangle = \langle T\Delta_Q f, \mathbf{1}-\chi_{S'}\rangle = \int_{\mathbb{C}\setminus S'} \big(T\Delta_Q f\big)(z)\,d\mu(z).$$

Taking into account that the Cauchy inequality implies $\|\Delta_Q f\|_{L^1(\mu)} \le \|\Delta_Q f\|_{L^2(\mu)}\mu(Q)^{1/2} \le C\mu(Q)^{1/2}$, it is easy (using the properties of Calderón–Zygmund kernels) to obtain the following estimate for $z \notin Q$:

$$\big|\big(T\Delta_Q f\big)(z)\big| \le \frac{C\cdot \ell(Q)^\varepsilon}{\operatorname{dist}(z,Q)^{1+\varepsilon}}\mu(Q)^{1/2}.$$

Integrating the distribution function, we get

$$\big|\langle f, \Delta_Q T'(\mathbf{1}-\chi_{S'})\rangle\big| \le \int_{\mathbb{C}\setminus S'} \big|\big(T\Delta_Q f\big)(z)\big|\,d\mu(z) \le C'\cdot \frac{\ell(Q)^\varepsilon}{\ell(S)^\varepsilon}\mu(Q)^{1/2}.$$

Since f is arbitrary, we get

$$\|\Delta_Q T'(\mathbf{1}-\chi_{S'})\|^2_{L^2(\mu)} \le C\cdot \frac{\ell(Q)^{2\varepsilon}}{\ell(S)^{2\varepsilon}}\mu(Q).$$

Taking the sum over all $Q \subset S$, we get

$$\sum_{Q\in S} \|\Delta_Q T'(\mathbf{1}-\chi_{S'})\|^2_{L^2(\mu)} \le C\mu(S) \le C\mu(S'),$$

and finally,

$$\sum_{Q\in S} \|\Delta_Q T'\mathbf{1}\|^2_{L^2(\mu)} \le C\mu(1.1S). \tag{3.2}$$

Now we apply the estimate (3.2) to the squares $S \subset R$ of size $2^{-r}\ell(R)$ such that $\partial S\cap\partial R = \varnothing$, then to the squares of size $2^{-(r+1)}\ell(R)$, $\partial S\cap\partial R = \varnothing$ that were not covered on the previous step, then to the squares of size $2^{-(r+2)}\ell(R)$ ... (see Figure 2). Clearly for all such S, we have $1.1S \subset R$ and each point of R is covered by no more than 4 squares $1.1S$ (see again Fig. 2), and we are done. □

Proof of Theorem 3.1. This theorem is essentially the Carleson Embedding Theorem. Since the vectors $\Delta_Q T'\mathbf{1}$, $Q \in \mathcal{D}$, are orthogonal, to prove that the paraproduct Π is bounded, we need to show that

$$\sum_{R\in\mathcal{D}} \sum_{\substack{Q\subset R:\ell(Q)=2^{-r}\ell(R)\\ \partial Q\cap\partial R=\varnothing}} |f_R|^2\cdot \|\Delta_Q T'\mathbf{1}\|^2_{L^2(\mu)} \le C\|f\|^2_{L^2(\mu)},$$

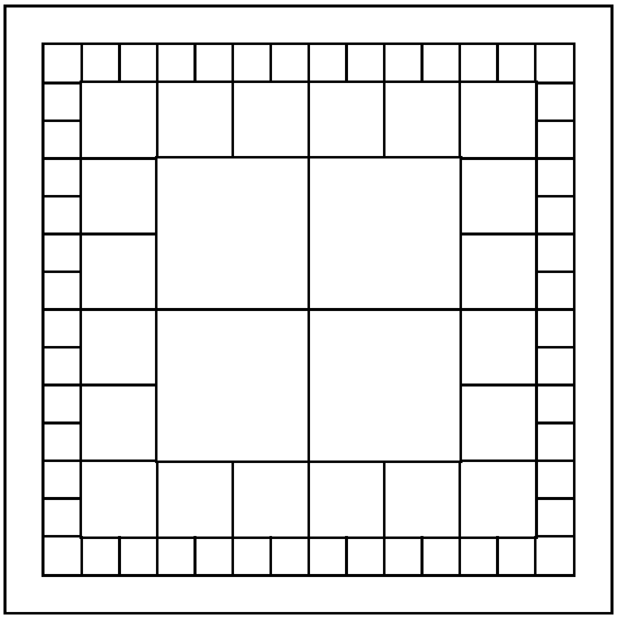

FIGURE 2. Whitney type decomposition of the square R. There are four squares S of size $2^{-r}\ell(R)$ (here $r=2$), around are squares of size $2^{-(r+1)}\ell(R)$, then squares of size $2^{-(r+2)}\ell(R)$.

where f_R denotes the average of f over the square R: $f_R = \mu(R)^{-1}\int_R f\,d\mu$ (note the difference with $E_R f$: f_R is just a number and $E_R f$ is a function: $E_R f = f_R \cdot \chi_R$).

Denote

$$a_R := \sum_{\substack{Q\subset R:\ell(Q)=2^{-r}\ell(R)\\ \partial Q\cap\partial R=\varnothing}} \|\Delta_Q T'\mathbf{1}\|^2_{L^2(\mu)}\,.$$

Then to prove the theorem, we need to show that

$$\sum_{R\in\mathcal{D}} |f_R|^2 a_R \le C\|f\|^2_{L^2(\mu)}\,.$$

Note that it is enough to prove this inequality only for non-negative functions. So from now on, we assume that $f\ge 0$.

Lemma 3.2 implies that the numbers a_R satisfy the Carleson measure condition

$$\sum_{Q\subset R} a_Q \le C'\mu(R).$$

So, we have reduced the problem to the following version of the Carleson Embedding Theorem: *if the numbers* a_Q *satisfy* $\sum_{Q\subset R} a_Q \le \mu(R)$ *for every*

$R \in \mathcal{D}$ (we can divide a_Q by C'), *then for any* $f \geq 0$, *we have*

$$\sum_{R \in \mathcal{D}} f_R^2 a_R \leq C \|f\|_{L^2(\mu)}^2 . \tag{3.3}$$

Now there exist many different ways to prove this and similar results. We present one more way, based on the Bellman function.

Consider the function $\mathcal{B}$ of three real variables

$$\mathcal{B}(F, f, A) = 4 \left(F - \frac{f^2}{1 + A} \right)$$

defined on the set

$$0 \leq A \leq 1, \qquad 0 \leq f \leq F^{1/2}$$

(note, that it is a convex subset of $\mathbb{R}^3$). It is easy to check that the function $\mathcal{B}$ is concave ($d^2\mathcal{B}$ is non-positive definite) and that $\dfrac{\partial \mathcal{B}}{\partial A} \geq f^2$.

Let $\mu_k \geq 0$ and $\mu := \sum \mu_k$. Let the triples (F, f, A) and (F_k, f_k, A_k) be in the domain of $\mathcal{B}$ and satisfy

$$\mu F = \sum \mu_k F_k, \qquad \mu f = \sum \mu_k f_k, \qquad \mu A - \sum \mu_k A_k =: a \geq 0.$$

Then

$$\begin{aligned} \frac{a}{\mu} f^2 \leq \mathcal{B}(F, f, A) - \mathcal{B}(F, f, A - \frac{a}{\mu}) &= \mathcal{B}(F, f, A) - \mathcal{B}(F, f, \frac{1}{\mu} \sum \mu_k A_k) \\ &\leq \mathcal{B}(F, f, A) - \frac{1}{\mu} \sum \mu_k \mathcal{B}(F_k, f_k, A_k) \end{aligned}$$

(the first inequality follows from $\dfrac{\partial \mathcal{B}}{\partial A} \geq f^2$, and the second one follows from $d^2\mathcal{B} \leq 0$). We rewrite the resulting inequality as

$$a f^2 \leq \mu \mathcal{B}(F, f, A) - \sum \mu_k \mathcal{B}(F_k, f_k, A_k) \tag{3.4}$$

and we will call it the *main inequality.*

Consider a dyadic square R, and let R_k, $k = 1, 2, 3, 4$, be the four dyadic squares of size $\ell(R)/2$ contained in R.

Denote $\mu = \mu(R)$, $\mu_k = \mu(R_k)$. Let us apply the main inequality to

$$F = F_R = \mu(R)^{-1} \int_R f^2 d\mu, \qquad f = f_R = \mu(R)^{-1} \int_R f d\mu,$$
$$A = A_R = \mu(R)^{-1} \sum_{Q \in R} a_Q$$

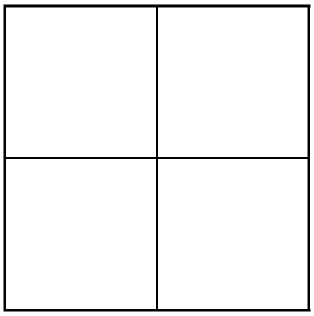

FIGURE 3. A square R and its *skeleton* (the boundary of the square and the cross).

with F_k, f_k, and A_k being the corresponding averages[4] for the squares R_k. In this case $a = a_R$. All the assumptions are satisfied and we get

$$a_R f_R^2 \le \mu(R)\mathcal{B}(F_R, f_R, A_R) - \sum_{k=1}^{4} \mu(R_k)\mathcal{B}(F_{R_k}, f_{R_k}, A_{R_k}).$$

Let us now write the same inequality for each of the four squares R_k, then for the squares in R_k, etc. If we take the sum of all such inequalities for n levels, we get

$$\sum_{\substack{Q\subset R\\ \ell(Q)>2^{-n}\ell(R)}} a_Q f_Q^2 \le \mu(R)\mathcal{B}(F_R, f_R, A_R) - \sum_{\substack{Q\subset R\\ \ell(Q)=2^{-n}\ell(R)}} \mu(Q)\mathcal{B}(F_Q, f_Q, A_Q).$$

Since $0 \le \mathcal{B}(F, f, A) \le 4F$ on the domain of $\mathcal{B}$,

$$\sum_{Q\subset R} a_Q f_Q^2 \le \mu(R)\mathcal{B}(F_R, f_R, A_R) \le 4\mu(R)F = 4\int_R f^2 d\mu$$

which implies (3.3), and therefore the theorem. □

3.2. **Estimates of $\langle T\varphi_Q, \psi_R\rangle$.** Since we know that the paraproduct Π is bounded, we only need to estimate the matrix $\langle (T - \Pi')\varphi_Q, \psi_R\rangle$.

For a square R let its *skeleton* $\operatorname{Sk} R$ be the union of the boundaries of the four subsquares of R of size $\ell(R)/2$, see Fig. 3.

We will call a pair of squares Q, R with $\ell(Q) \le \ell(R)$ *singular* if

$$\operatorname{dist}(Q, \operatorname{Sk} R) \le \ell(Q)^\alpha \ell(R)^{1-\alpha},$$

where $\alpha = \varepsilon/(2\varepsilon + 2)$, and so $\alpha + \alpha\varepsilon = \varepsilon/2$.

[4]Clearly, all the triples are in the domain. The inequality $f \le F^{1/2}$ is just the Cauchy–Schwartz inequality.

We also will call a pair of squares Q, R with $\ell(Q) \le \ell(R)$, *essentially singular* if, in addition, $\ell(Q) < 2^{-r}\ell(R)$. Note that the definitions are consistent with the ones we had for disjoint Q and R.

Suppose $Q \subset R$, $\ell(Q) < 2^{-r}\ell(R)$, and the pair Q, R is not singular. Let φ_Q and ψ_R be functions supported by Q and R, respectively, constant on each of the four subsquares of Q (respectively, R) and orthogonal to constants. Let us normalize φ_Q, ψ_R in $L^2(\mu)$:

$$\|\varphi_Q\|^2_{L^2(\mu)} = 1, \qquad \|\psi_R\|^2_{L^2(\mu)} = 1$$

Let R_k, $k = 1, \dots, 4$, be the dyadic subsquares of R of size $\ell(R)/2$. Suppose also for definiteness that $Q \subset R_1$.

We can write $\psi_R = \sum_{k=1}^4 c_k \chi_{R_k}$, where c_k are some constants.

We know (see (3.1)) that if $Q \subset R_1$, then $\langle \Pi' \varphi_Q, \psi_R\rangle = \langle T\varphi_Q, \mathbf{1}\rangle c_1$. Then

$$\begin{aligned}
\Big|\langle (T - \Pi')\varphi_Q, \psi_R\rangle\Big| &= \Big|\langle T\varphi_Q, \psi_R - c_1\mathbf{1}\rangle\Big| \\
&\le |c_1| \cdot \Big|\langle T\varphi_Q, \chi_{R_1} - \mathbf{1}\rangle\Big| + \sum_{k=2}^4 \Big|\langle T\varphi_Q, c_k\chi_{R_k}\rangle\Big|
\end{aligned}$$

The first term is easy to estimate. For $z \in \mathbb{C} \setminus Q$ we can write

$$|(T\varphi_Q)(z)| \le C\frac{\ell(Q)^\varepsilon}{\operatorname{dist}(z,Q)^{1+\varepsilon}} \cdot \|\varphi_Q\|_{L^1(\mu)} \le C\frac{\ell(Q)^\varepsilon}{\operatorname{dist}(z,Q)^{1+\varepsilon}} \cdot \mu(Q)^{1/2}$$

Therefore writing the integral below via the distribution function, one can get

$$|\langle T\varphi_Q, \chi_{R_1} - \mathbf{1}\rangle| \le \int_{\mathbb{C}\setminus R_1} |T\varphi_Q| d\mu \le C\frac{\ell(Q)^\varepsilon}{\operatorname{dist}(Q,\partial R_1)^\varepsilon} \cdot \mu(Q)^{1/2}.$$

Since $\|\psi_R\|_{L^2(\mu)} = 1$, we have $|c_1| \le \mu(R_1)^{-1/2}$ and therefore

$$|c_1| \cdot \Big|\langle T\varphi_Q, \chi_{R_1} - \mathbf{1}\rangle\Big| \le C\frac{\ell(Q)^\varepsilon}{\operatorname{dist}(Q,\partial R_1)^\varepsilon} \cdot \left(\frac{\mu(Q)}{\mu(R_1)}\right)^{1/2}.$$

The pair Q, R is not singular, which means that

$$\operatorname{dist}(Q, \partial R_1) \ge \ell(Q)^\alpha \ell(R)^{1-\alpha} \ge \ell(Q)^{1/2}\ell(R)^{1/2},$$

and therefore

$$|c_1| \cdot \Big|\langle T\varphi_Q, \chi_{R_1} - \mathbf{1}\rangle\Big| \le C \cdot \left(\frac{\ell(Q)}{\ell(R)}\right)^{\varepsilon/2} \left(\frac{\mu(Q)}{\mu(R_1)}\right)^{1/2}.$$

To estimate $\langle T\varphi_Q, c_k\chi_{R_k}\rangle$, $k = 2, 3, 4$, we can use Lemma 2.4. It implies (if we take into account that in our case $D(Q,R) \asymp \ell(R)$) that

$$\begin{aligned}
|\langle T\varphi_Q, c_k\chi_{R_k}\rangle| &\le C \cdot \frac{\ell(Q)^{\varepsilon/2}}{\ell(R)^{1+\varepsilon/2}} \mu(Q)^{1/2}\mu(R_k)^{1/2} \\
&\le C \cdot \left(\frac{\ell(Q)}{\ell(R)}\right)^{\varepsilon/2} \left(\frac{\mu(Q)}{\ell(R)}\right)^{1/2} \\
&\le C \cdot \left(\frac{\ell(Q)}{\ell(R)}\right)^{\varepsilon/2} \left(\frac{\mu(Q)}{\mu(R_1)}\right)^{1/2}
\end{aligned}$$

So we have proved the following lemma.

Lemma 3.3. *Suppose that $Q \subset R$, $\ell(Q) \le 2^{-r}\ell(R)$ and that the pair Q, R is not singular. Let φ_Q and ψ_R be functions supported by Q and R, respectively, constant on each of the four subsquares of Q (respectively, R) and orthogonal to constants. Let also R_1 be the subsquare of R of size $\ell(R)/2$ containing Q. Then for the Calderón–Zygmund operator T,*

$$\Big|\langle (T - \Pi')\varphi_Q, \psi_R\rangle\Big| \le C \cdot \left(\frac{\ell(Q)}{\ell(R)}\right)^{\varepsilon/2} \left(\frac{\mu(Q)}{\mu(R_1)}\right)^{1/2} \|\varphi_Q\|_{L^2(\mu)} \|\psi_R\|_{L^2(\mu)} .$$

3.3. **Estimate of $\langle Tf, g\rangle$ for "good" functions.** Let a dyadic lattice $\mathcal{D}$ be fixed. A dyadic square Q is called "bad" (for the lattice) if there exists a bigger square R such that the pair Q, R is essentially singular; otherwise the square is called "good".

A function $f \in L^2(\mu)$ is called "good" if for any "bad" square Q, we have $\Delta_Q f = 0$.

Theorem 3.4. *Let μ be a regular measure, and let T be a Calderón–Zygmund operator such that T and T' are bounded on characteristic functions of squares (in the sense of condition 4 of Theorem 1.2). Then for any two "good" functions f and g*

$$|\langle Tf, g\rangle| \le C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)} .$$

Proof. We need to estimate the sum $\langle Tf, g\rangle = \sum_{Q,R\in\mathcal{D}} \langle T\Delta_Q f, \Delta_R g\rangle$. Due to the symmetry, it is enough to estimate only the sum over $\ell(Q) \le \ell(R)$.

First of all, notice that since T is bounded on characteristic functions of squares,

$$|\langle T\Delta_Q f, \Delta_R g\rangle| \le C\|\Delta_Q f\|_{L^2(\mu)}\|\Delta_R g\|_{L^2(\mu)} .$$

Therefore

$$\begin{aligned}
&\sum_{\substack{2^{-r}\ell(R)\le\ell(Q)\le\ell(R)\\ \operatorname{dist}(Q,R)\le\ell(R)}} |\langle T\Delta_Q f, \Delta_R g\rangle| \\
&\le C\Big(\sum_{Q\in\mathcal{D}} \|\Delta_Q f\|^2_{L^2(\mu)}\Big)^{1/2}\Big(\sum_{R\in\mathcal{D}} \|\Delta_R g\|^2_{L^2(\mu)}\Big)^{1/2} = C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)}
\end{aligned}$$

(finitely many bounded diagonals).

On the other hand, Lemma 2.4 and Theorem 2.5 imply that

$$\sum_{\substack{2^{-r}\ell(R)\le\ell(Q)\le\ell(R)\\ \operatorname{dist}(Q,R)\ge\ell(R)}} |\langle T\Delta_Q f,\Delta_R g\rangle|$$

$$\le C\Big(\sum_{Q\in\mathcal{D}}\|\Delta_Q f\|^2_{L^2(\mu)}\Big)^{1/2}\Big(\sum_{R\in\mathcal{D}}\|\Delta_R g\|^2_{L^2(\mu)}\Big)^{1/2} = C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)}.$$

It remains to estimate the sum

$$\sum_{\ell(Q)\le 2^{-r}\ell(R)}\langle T\Delta_Q f,\Delta_R g\rangle = \sum_{\substack{Q\subset R\\ \ell(Q)\le 2^{-r}\ell(R)}}\ldots + \sum_{\substack{Q\cap R=\varnothing\\ \ell(Q)\le 2^{-r}\ell(R)}}\ldots\ .$$

The second sum can be estimated by Lemma 2.4 and Theorem 2.5:

$$\sum_{\substack{Q\cap R=\varnothing\\ \ell(Q)\le 2^{-r}\ell(R)}} |\langle T\Delta_Q f,\Delta_R g\rangle|$$

$$\le \sum_{\substack{Q\cap R=\varnothing\\ \ell(Q)\le 2^{-r}\ell(R)}} C\frac{\ell(Q)^{\varepsilon/2}\ell(R)^{\varepsilon/2}}{D(Q,R)^{1+\varepsilon}}\mu(Q)^{1/2}\mu(R)^{1/2}\|\Delta_Q f\|_{L^2(\mu)}\|\Delta_R g\|_{L^2(\mu)}$$

$$\le C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)}$$

(since the functions f, g are "good", entries $\langle T\Delta_Q f,\Delta_R g\rangle$ corresponding to essentially singular pairs Q, R are zero, and all others can be estimated as above, see Lemma 2.4)

To estimate the first sum, notice that Π has a very special "triangular" matrix. Namely, in the sum $\langle f,\Pi g\rangle = \sum_{Q,R}\langle\Delta_Q f,\Pi\Delta_R g\rangle$ only the terms with $Q\subset R, \ell(Q)\le 2^{-r}\ell(R)$ may be non-zero. Thus

$$\sum_{\substack{Q\subset R\\ \ell(Q)\le 2^{-r}\ell(R)}}\langle T\Delta_Q f,\Delta_R g\rangle = \sum_{\substack{Q\subset R\\ \ell(Q)\le 2^{-r}\ell(R)}}\langle (T-\Pi')\Delta_Q f,\Delta_R g\rangle + \langle f,\Pi g\rangle.$$

We know that the paraproduct Π is bounded, so we have to estimate the sum.

Now let us use again the fact that f is "good". Taking into account the estimate of $\langle (T-\Pi')\Delta_Q f,\Delta_R g\rangle$ for non-singular pairs Q,R, which follows from Lemma 3.3, we get that it is enough to show that

$$\sum_{\substack{Q\subset R\\ \ell(Q)<2^{-r}\ell(R)}}\left(\frac{\ell(Q)}{\ell(R)}\right)^{\varepsilon/2}\left(\frac{\mu(Q)}{\mu(R_1)}\right)^{1/2}\|\Delta_Q f\|_{L^2(\mu)}\|\Delta_R g\|_{L^2(\mu)}$$

$$\le C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)},$$

where $R_1 = R_1(Q,R)$ is the subsquare of R of size $\ell(R)/2$ containing Q. In other words, it is enough to show that the matrix $\{T_{Q,R}\}_{Q,R\in\mathcal{D}}$,

$$T_{Q,R} = \begin{cases} \left(\dfrac{\ell(Q)}{\ell(R)}\right)^{\varepsilon/2} \left(\dfrac{\mu(Q)}{\mu(R_1)}\right)^{1/2}, & Q\subset R,\ \ell(Q) < 2^{-r}\ell(R); \\ 0, & \text{otherwise,} \end{cases}$$

generates a bounded operator in ℓ^2.

Let us "slice" the matrix $\{T_{Q,R}\}_{Q,R\in\mathcal{D}}$. Namely, for $n = r+1, r+2, r+3, \dots$, define the matrix $\{T^{(n)}_{Q,R}\}_{Q,R\in\mathcal{D}}$ by

$$T^{(n)}_{Q,R} = \begin{cases} T_{Q,R}, & \ell(Q) = 2^{-n}\ell(R); \\ 0, & \text{otherwise.} \end{cases}$$

If we show that the norms of the operators $T^{(n)}$ decrease as a geometric progression, i.e., that

$$\sum_{Q,R\in\mathcal{D}} T^{(n)}_{Q,R}|x_Q|\cdot|y_R| \le 2^{-n\beta}C\left(\sum_{Q\in\mathcal{D}}|x_Q|^2\right)^{1/2}\left(\sum_{Q\in\mathcal{D}}|y_Q|^2\right)^{1/2}$$

for some $\beta > 0$, then we are done.

We can split the matrices $T^{(n)}$ into layers $T^{(n,k)}$, where

$$T^{(n,k)}_{Q,R} = \begin{cases} T^{(n)}_{Q,R}, & \ell(Q) = 2^k \\ 0, & \text{otherwise} \end{cases}$$

Clearly, the layers $T^{(n,k)}$ of $T^{(n)}$ do not interfere; therefore it is enough to estimate each layer separately.

Note that the "rows" $\{T^{(n,k)}_{Q,R} : Q\subset R\}$ (R is fixed, $\ell(R) = 2^{k+n}$) are uniformly (in $\mathbb{R}$) bounded in ℓ^2:

$$\begin{aligned}\sum_{\substack{Q:Q\subset R\\ \ell(Q)=2^k}} (T^{(n,k)}_{Q,R})^2 &\le C\cdot\left(\frac{\ell(Q)}{\ell(R)}\right)^{\varepsilon}\sum_{\substack{R_1:R_1\subset R\\ \ell(R_1)=\ell(R)/2}}\ \sum_{\substack{Q:Q\subset R_1\\ \ell(Q)=2^k}}\frac{\mu(Q)}{\mu(R_1)} \\ &= 4C\cdot\left(\frac{\ell(Q)}{\ell(R)}\right)^{\varepsilon} = 4C2^{-n\varepsilon}.\end{aligned}$$

Note that the supports of the "rows" of $T^{(n,k)}$ are pairwise disjoint. Therefore the rows do not interfere, and so the norm of $T^{(n,k)}$ is bounded by $C2^{-n\varepsilon/2}$. We are done.

The sum over the pairs such that $\ell(R) < \ell(Q)$ is estimated similarly, and the theorem is proved. $\square$

4. Averaging trick and pulling ourselves by hair

4.1. **Random dyadic lattice.** After we have proved the estimate for "good" functions, the question arises: 'What should we do about the "bad" ones?' And the surprising answer is — nothing, just ignore them! The point is that "bad" intervals are extremely rare, so we do not have to worry about them.

Let us explain why.

To illustrate the idea, first consider the case when we have some *a priori* estimate on the norm of T. This happens, for example, if we want to prove that all regularizations T_ε are uniformly bounded, and we know that each T_ε is bounded. For simplicity we also assume that for any dyadic lattice $\mathcal{D}$ any function $f \in L^2(\mu)$ can be decomposed as $f = \sum_{Q\in\mathcal{D}} \Delta_Q f$.

Fix a grid of squares of size 2^{-n_0}. Clearly, if n_0 is sufficiently large, one can find functions f and g in $L^2(\mu)$, constant on the squares of the grid, and such that $\|f\| = \|g\| = 1$ and

$$\Big|\langle Tf, g\rangle\Big| \geq \frac{1}{2}\|T\|.$$

Based on the given grid of squares, let us construct a *random* dyadic lattice. We do not have any choice constructing dyadic squares smaller than those we already have: we just split each of the given squares into 4 equal subsquares, then split each new square etc. The interesting part begins when we consider bigger squares. There are 4 possibilities for how to position squares of size $2\cdot 2^{-n_0}$, so let us assign probability $1/4$ to each of them. For each of the choices there are 4 possibilities for putting squares of size $4\cdot 2^{-n_0}$, so again assign probability $1/4$ to each of them. Continuing the procedure, we get the desired random dyadic lattice.

For a given dyadic square Q in the base (one of the squares of size 2^{-n_0} we started from) the probability that there exists a bigger square R of size $2^{-n_0}2^{k+1}$ such that the pair Q, R is singular, can be estimated from above by $4(2^{-k\alpha} + 2^{-k}) \leq 8\cdot 2^{-k\alpha}$. So the probability that there exists a bigger square R such that the pair Q, R is essentially singular can be estimated by $p_r = \sum_{k=r}^{\infty} 8\cdot 2^{-k\alpha} = 8\cdot 2^{-r\alpha}/(1-2^{-\alpha})$.

Similarly, we pick a dyadic lattice and take a square Q of size greater than 2^{-n_0} there. Consider the conditional probability (under the condition that the square Q belongs to the random dyadic lattice) that there exists a bigger square R such that the pair Q, R is essential singular. Clearly it can be estimated by the same constant p_r.

4.2. **Pulling ourselves by hair.** For a function f constant on all squares of the grid, let us split it as

$$f = \sum_{\substack{Q\in\mathcal{D}\\ \ell(Q)\geq 2^{-n_0}}} \Delta_Q f = \sum_{Q:Q \text{ is bad}} \Delta_Q f + \sum_{Q:Q\text{ is good}} \Delta_Q f = f_{\text{bad}} + f_{\text{good}}$$

(of course, the decomposition depends on the choice of the lattice Q).

Let us estimate the mathematical expectation $\mathbb{E}\|f_{\text{bad}}\|^2_{L^2(\mu)}$ (taken over the random dyadic lattices constructed above). To do that, let us consider (for a fixed dyadic lattice $\mathcal{D}$) the so-called *square function* $S(z)$ defined for $z \in \mathbb{C}$ by

$$Sf(z) = S_{\mathcal{D}} f := \sum_{Q:Q\ni z} \|\Delta_Q f\|^2_{L^2(\mu)} \mu(Q)^{-1} \chi_Q \,.$$

Clearly, $\int_{\mathbb{C}} Sf(z)d\mu(z) = \|f\|^2_{L^2(\mu)}$.

Consider the average square function $\mathbb{E}Sf(z)$ (for each z take the mathematical expectation over all dyadic lattices). Changing the order of integration, one can see that $\int_{\mathbb{C}} \mathbb{E}Sf(z)d\mu(z) = \|f\|^2_{L^2(\mu)}$.

The (conditional) probability that a square Q is bad is at most p_r, so

$$\mathbb{E}Sf_{\text{bad}}(z) \le p_r \mathbb{E}Sf(z).$$

Since

$$\mathbb{E}\|f_{\text{bad}}\|^2 = \mathbb{E}\Big(\int Sf_{\text{bad}} d\mu\Big) = \int \mathbb{E}Sf_{\text{bad}}\, d\mu \le p_r \int_{\mathbb{C}} \mathbb{E}Sf\, d\mu = p_r\|f\|^2_{L^2(\mu)},$$

we get $\mathbb{E}\|f_{\text{bad}}\|^2 \le p_r\|f\|^2_{L^2(\mu)}$.

The probability that $\|f_{\text{bad}}\|^2_{L^2(\mu)} \ge 4p_r\|f\|^2_{L^2(\mu)}$ cannot be more than $1/4$, therefore with probability $3/4$ we have $\|f_{\text{bad}}\|_{L^2(\mu)} \le 2\sqrt{p_r}\|f\|_{L^2(\mu)}$.

So, if we have *two* functions f and g, then with probability at least $1/2$ we have simultaneously

$$\|f_{\text{bad}}\|_{L^2(\mu)} \le 2\sqrt{p_r}\|f\|_{L^2(\mu)}, \qquad \|g_{\text{bad}}\|_{L^2(\mu)} \le 2\sqrt{p_r}\|g\|_{L^2(\mu)}.$$

Now pick a dyadic lattice such that for $\delta = 2\sqrt{p_r}$, we have

$$\|f_{\text{bad}}\|_{L^2(\mu)} \le \delta\|f\|_{L^2(\mu)} \qquad \text{and} \qquad \|g_{\text{bad}}\|_{L^2(\mu)} \le \delta\|g\|_{L^2(\mu)}\,.$$

We can always pick such a lattice because, as we have shown above, a random lattice fits with probability at least $1/2$.

First of all, let us recall that we have the estimate

$$|\langle Tf_{\text{good}}, g_{\text{good}}\rangle| \le C\|f_{\text{good}}\|_{L^2(\mu)}\|g_{\text{good}}\|_{L^2(\mu)}\,.$$

We can write

$$\begin{aligned}|\langle Tf, g\rangle| &\le |\langle Tf_{\text{good}}, g\rangle| + |\langle Tf_{\text{bad}}, g\rangle| \\ &\le |\langle Tf_{\text{good}}, g_{\text{good}}\rangle| + |\langle Tf_{\text{good}}, g_{\text{bad}}\rangle| + |\langle Tf_{\text{bad}}, g\rangle|.\end{aligned}$$

We have

$$|\langle Tf_{\text{good}}, g_{\text{good}}\rangle| \le C\|f_{\text{good}}\|_{L^2(\mu)}\|g_{\text{good}}\|_{L^2(\mu)} \le C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)} \le C,$$
$$|\langle Tf_{\text{good}}, g_{\text{bad}}\rangle| \le \|T\|\delta, \qquad |\langle Tf_{\text{bad}}, g\rangle| \le \|T\|\delta,$$

because $\|f_{\text{bad}}\|_{L^2(\mu)} \le \delta$, $\|f_{\text{good}}\|_{L^2(\mu)} \le \|f\|_{L^2(\mu)} \le 1$, and the same is true for g. Therefore, since $|\langle Tf, g\rangle| \ge \|T\|/2$,

$$\frac{1}{2}\|T\| \le C + 2\delta\|T\|.$$

If we pick r large enough so that $\delta \leq 1/8$, we can conclude that $\|T\| \leq 4C$ and we are done.

4.3. **What to do if we do not have an a priori estimate of $\|T\|$.** The first idea would be to restrict the operator T to a subspace where we know it is bounded.

For example, let us fix some dyadic grid of squares of size 2^{-n_0} and some big square. If we consider functions that are supported by this large square and are constant on each square of the grid, then the restriction of the operator T to such functions will be bounded (maybe its norm will be very large), because the space is finite-dimensional. If we prove an estimate for the norm (or the quadratic form) of the "restricted" operator T that is independent of how fine the grid is and how large the square is, then we are done.

At first glance it seems that the reasoning from Section 4.2 based on splitting functions into "bad" and "good" parts works fine, but there is a small, but unpleasant detail: the functions $f_{\rm bad}$, $g_{\rm bad}$ are not necessarily supported by the same square as f and g. And for that reasoning to work, the functions $f_{\rm bad}$, $g_{\rm bad}$ should remain in the same class as f and g.

So we have to change the reasoning a little. We have a grid of squares of size 2^{-n_0} fixed. For a square R of size 2^{n_0} that is a union of squares of the grid, let M_R be the norm of the restriction of the operator $E_{-n_0}T$ to the space of all functions supported by the square R and constant on the squares of the grid. Let $M = \sup_R M_R$ where the supremum is taken over all such squares R. Clearly $M < \infty$.

Pick a square R and functions $f, g \in L^2(\mu)$ with $\|f\| = \|g\| = 1$, where f is supported by the square R and both are constant on the squares of the grid such that

$$\Big|\langle Tf, g\rangle\Big| \geq \frac{1}{2}M.$$

For the function f, write the decomposition

$$f = \sum_{Q:\,\ell(Q)\leq\ell(R)} \Delta_Q f + \sum_{Q:\,\ell(Q)=\ell(R)} E_Q f,$$

and denote

$$f_{\rm bad} = \sum_{\substack{Q:\,\ell(Q)\leq\ell(R)\\ Q \text{ is bad}}} \Delta_Q f,$$

$$f_{\rm good} = f - f_{\rm bad} = \sum_{\substack{Q:\,\ell(Q)\leq\ell(R)\\ Q \text{ is good}}} \Delta_Q f + \sum_{Q:\,\ell(Q)=\ell(R)} E_Q f.$$

Again, as above, the decomposition $f = f_{\rm good} + f_{\rm bad}$ depends on the choice of the random dyadic lattice.

Acting as in the case with an a priori estimate (see Section 4.2), we get a dyadic lattice such that for $\delta = 2\sqrt{p_r}$,

$$\|f_{\text{bad}}\|_{L^2(\mu)} \le \delta\|f\|_{L^2(\mu)} \qquad \text{and} \qquad \|g_{\text{bad}}\|_{L^2(\mu)} \le \delta\|g\|_{L^2(\mu)}\,.$$

Note that although the function f_{bad} is not supported by the square R, it is supported by the square $3R$. And it is easy to see that the norm of the restriction of $E_{-n_0}T$ to the space of all functions supported by $3R$ and constant on the squares of the grid is at most $\sqrt{3^2}M = 3M$.

As in Section 4.2, one can write

$$|\langle Tf, g\rangle| \le |\langle Tf_{\text{good}}, g_{\text{good}}\rangle| + |\langle Tf_{\text{good}}, g_{\text{bad}}\rangle| + |\langle Tf_{\text{bad}}, g\rangle|,$$

and

$$|\langle Tf_{\text{good}}, g_{\text{good}}\rangle| \le C\|f_{\text{good}}\|_{L^2(\mu)}\|g_{\text{good}}\|_{L^2(\mu)} \le C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)} \le C,$$
$$|\langle Tf_{\text{good}}, g_{\text{bad}}\rangle| \le 3M\delta, \qquad |\langle Tf_{\text{bad}}, g\rangle| \le 3M\delta.$$

Therefore

$$\frac{1}{2}M \le C + 6\delta M.$$

If we pick r large enough so that $\delta \le 1/24$, we can conclude that $M \le 4C$ and we are done.

References

[1] Guy David, *Wavelets and singular integrals on curves and surfaces*, Lecture Notes in Mathematics, vol. 1465, Springer-Verlag, Berlin, 1991.

[2] S. Treil and A. Volberg, *Wavelets and the angle between past and future*, J. Funct. Anal. **143** (1997), no. 2, 269–308.

[3] F. L. Nazarov and S. R. Treil, *The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis*, Algebra i Analiz **8** (1996), no. 5, 32–162.

[4] J. B. Garnett, "Bounded analytic functions," Acad. Press, NY, 1981.

[5] X. Tolsa, *L^2-boundedness of the Cauchy integral operator for continuous measures*, submitted.

[6] Pertti Mattila, Mark S. Melnikov, and Joan Verdera, *The Cauchy integral, analytic capacity, and uniform rectifiability*, Ann. of Math. (2) **144** (1996), no. 1, 127–136.

[7] Mark S. Melnikov and Joan Verdera, *A geometric proof of the L^2 boundedness of the Cauchy integral on Lipschitz graphs*, Internat. Math. Res. Notices **1995**, no. 7, 325–331.

Address:

Department of Mathematics
Michigan State University
East Lansing, Michigan 48824, USA
e-mail: fedja@math.msu.edu
treil@math.msu.edu
volberg@math.msu.edu