WEIGHTED NORM INEQUALITIES FOR CALDERÓN–ZYGMUND OPERATORS

By

Aleksandr B. Reznikov

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Mathematics — Doctor of Philosophy

2014

ABSTRACT

WEIGHTED NORM INEQUALITIES FOR CALDERÓN–ZYGMUND OPERATORS

By

Aleksandr B. Reznikov

Given a Calderón-Zygmund operator T and two weights u and v, we study sufficient conditions for this operator to be bounded from the space $L^p(u)$ to $L^p(v)$. We also study sharp bounds for the corresponding norm. Further, we study a question about conditions for boundedness of all Calderón-Zygmund operators from $L^p(u)$ to $L^p(v)$. We do it in the Euclidian setting and in metric spaces.

Finally, we study the limiting case $p = 1$ and the case $u = v$, when the operator has a chanse to be weakly bounded, i.e. bounded from the space $L^1(u)$ to the space $L^{1,\infty}(u)$.

In particular, we disprove the "A_1 conjecture", prove the "A_2 conjecture" in the metric space setting, prove the "bump conjecture" for $p = 2$; moreover, we state the "separated bump conjecture" and prove it in several particular cases.

To my mom, wife and daughter — three most wonderful women.

ACKNOWLEDGMENTS

This section is the most pleasant to write. During my graduate study at MSU I got, besides some nice results, a lot of great friends, collegues and, most important, got married. Though some problems remain unsolved and some things remain undone, I could not wish better five years for myself.

First of all, I would like to express my utmost gratitude to Alexander Volberg, who, in these five years, became for me more than just an academic father. His helpful, respectful and friendly attitude towards graduate students are absolutely unique, and I never felt sorry for my choice of the university and of the advisor. The problems that he gave me, and the conferences and workshops that he encouraged me to attend, were extremely useful for my mathematical career.

I would also like to mention other MSU faculty and staff, that mentored and helped me during these five years. Our secretaries Barbara Miller and Leslie Aitcheson, graduate directors Zhengfang Zhou and Casim Abbas, professors Aaron Levin, Sheldon Newhouse, Vladimir Peller, Michael Shapiro, Ignacio Uriarte-Tuero, Kirill Vanninskii, Jon Wolfson, Dapeng Zhan and Jane Zimmerman made my academic life a lot more pleasant, challenging and interesing.

My research was also greatly affected by collegues outside MSU. It was a great pleasure to interact with David Cruz-Uribe, Paul Hagelstein, Viktor Havin, Steve Hoffman, Alex Iosevich, Carlos Kenig, Michael Lacey, Fedor Nazarov, Carlos Perez, Stefanie Petermichl, Maria Reguera, Sergei Treil, Cristoph Thiele, Vasilii Vasyunin, Igor Verbitskii.

It would be a pity not to mention my collegues and friends Alya Beznosova, Matt Bond, Nick Boros, Liping Chen, Asya Diop, Misha Dubashinskii, Paata Ivanishvili, Misha Ivanov, Borya Krylov, Seryozha Matveenko, Seryozha Nikitin, Nick Pattakos, Guillermo Rey, Asya Slutskaya, Sasha Vallander, Liangmin Zhou.

I would like to thank Vladimir Shlapentokh for organizing a lot of interesting seminars and conversations. They stimulated me a lot, and I do think that they were very important for my development. Special thanks to participants of these meetings: Lyuba and Vladimir Shlapentokhs, Viktoriya and Mark Dykmans, Zhanna and Yuriy Gandelsmans, Irina and Michael Shapiro, Olga and Alexander Volbergs. The level of these meetings is extremely high, and they were inspiring both for "academic" and "real" life.

Finally, it is my utmost pleasure to thank for all support my family: my grandfather Grisha Mints, my grandmother Inna Mints, my mother Anna Reznikova, my parents in law Tatiana and Viktor Shushkovs, and my wife Natalia Shushkova.

TABLE OF CONTENTS

Chapter 1

Preliminary notation, definitions and theorems

1.1 Preliminary notation and definitions

We begin with some notation and definitions that are needed throughout this thesis.

Definition 1 (Weight)**.** By a weight in $\mathbb{R}^n$ we call a funtion w, which is positive almost everywhere and locally integrable with respect to the Lebesgue measure dx.

Definition 2 (Calderón-Zygmund kernel)**.** A function $K\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called a Calderón-Zygmund kernel if there exist positive numbers C and ε such that the following conditions are satisfied:

$$|K(x,y)| \leqslant \frac{C}{|x-y|^n} \tag{1.1}$$

$$|K(x,y) - K(x',y)| \leqslant C\frac{|x-x'|^\varepsilon}{|x-y|^{n+\varepsilon}} \quad \text{if} \quad |x-x'| < \frac{1}{2}|x-y| \tag{1.2}$$

$$|K(x,y) - K(x,y')| \leqslant C\frac{|y-y'|^\varepsilon}{|x-y|^{n+\varepsilon}} \quad \text{if} \quad |y-y'| < \frac{1}{2}|x-y|. \tag{1.3}$$

Definition 3 (Calderón-Zygmund Operator)**.** An operator T is called a Calderón-Zygmund operator, if there exists a Calderón-Zygmund Kernel K, such that

$$\text{For any } f \in C_0^\infty, \text{ and any } x \notin \operatorname{supp} f : Tf(x) = \int K(x,y)f(y)dy \tag{1.4}$$

$$T \text{ is a bounded operator from } L^2(dx) \text{ to } L^2(dx). \tag{1.5}$$

An example of a Calderón-Zygmund operator in dimension one is Hilbert transform H with kernel $K(x,y) = \frac{1}{x-y}$. An example of a Calderón-Zygmund operator in dimension n is Rietz transform with vector-valued kernel $K(x,y) = \frac{x-y}{|x-y|^{n+1}}$.

Notation 1. For a set $Q \subset \mathbb{R}^n$ and a function φ we denote

$$\langle \varphi \rangle_Q = \frac{1}{|Q|} \int_Q \varphi(x)dx,$$

where $|Q|$ is the Lebesgue measure of the set Q.

Definition 4 (The A_p class)**.** Let $1 < p < \infty$. A weight w belongs to a class A_p on a cube I (where I is allowed to be equal to $\mathbb{R}^n$) if the following holds:

$$\text{For every cube } J \subset I \colon \langle w \rangle_J \langle w^{-\frac{1}{p-1}} \rangle_J^{p-1} \leqslant Q.$$

The best constant Q is called the A_p *characteristic* of w and is denoted by $[w]_p$.

Definition 5 (The RH_p class)**.** Let $1 < p < \infty$. A weight w belongs to a class RH_p on a cube I (where I is allowed to be equal to $\mathbb{R}^n$) if the following holds:

$$\text{For every cube } J \subset I \colon \langle w^p \rangle_J^{\frac{1}{p}} \leqslant R \langle w \rangle_J.$$

The best constant R is called the RH_p *characteristic* of w and is denoted by $[w]_{RH_p}$.

As one can see, these definitions do not work for $p = 1$ and $p = \infty$. However, if we carefully consider the limit of the left-hand side, we get the following.

Definition 6 (The limiting cases)**.**

1. A weight w belongs to a class A_1 on a cube I (where I is allowed to be equal to $\mathbb{R}^n$) if the following holds:

$$\text{For every cube } J \subset I\colon \langle w \rangle_J \leqslant Q \inf_J w.$$

The best constant Q is called the A_1 *characteristic* of w and is denoted by $[w]_1$.

2. A weight w belongs to a class A_∞ on a cube I (where I is allowed to be equal to $\mathbb{R}^n$) if the following holds:

$$\text{For every cube } J \subset I\colon \langle w \rangle_J \leqslant Q e^{\langle \log(w) \rangle_J}.$$

The best constant Q is called the A_∞ *characteristic* of w and is denoted by $[w]_\infty$.

3. A weight w belongs to a class RH_1 on a cube I (where I is allowed to be equal to $\mathbb{R}^n$) if the following holds:

$$\text{For every cube } J \subset I\colon \langle \frac{w}{\langle w \rangle_J} \log \frac{w}{\langle w \rangle_J} \rangle_J \leqslant R.$$

The best constant R is called the RH_1 *characteristic* of w and is denoted by $[w]_{RH_1}$.

Next definitions will concern so-called “dyadic models” for Calderón-Zygmund operators. We make use of these models for weighted estimates of the operators.

Notation 2. If I is a cube in $\mathbb{R}^n$ then by $\ell(I)$ we denote the sidelength of I.

Definition 7 (Dyadic grid)**.** A dyadic grid $\mathcal{D}$ is a union of collections of cubes $\mathcal{D}_k$, such that:

1. For a fixed k cubes in $\mathcal{D}_k$ do not intersect, and for $I \in \mathcal{D}_k$: $\ell(I) = 2^{-k}$.

2. For any cube $I \in \mathcal{D}_{k+1}$ there exists a unique cube $J \in \mathcal{D}_k$, such that $I \subset J$. In this situation I is called a dyadic child of J; J is called the dyadic father of I.

Elements of $\mathcal{D}$ are called dyadic cubes.

Definition 8 (Carleson sequense)**.** Fix a dyadic grin $\mathcal{D}$. A sequense $\{a_I\}_{I\in\mathcal{D}}$ is called C-Carleson, if for any interval $J \in \mathcal{D}$ the following inequality holds:

$$\sum_{I\subset J} a_I|I| \leqslant C|J|.$$

Definition 9 (Generalized Haar function)**.** Given a dyadic cube J, h_J is a (generalized) Haar function associated to a cube J if

$$h_J(x) = \sum_{I\in ch(J)} c_I \chi_J(x),$$

where $ch(J)$ is the set of dyadic children of J and $|c_I| \leq 1$.

Definition 10 (Generalized Haar shift)**.** We say that an operator S has a Haar shift kernel of complexity (m, n) if

$$Sf(x) = \sum_J S_J(f),$$

where

$$S_J(f) = \frac{1}{|J|} \sum_{\substack{I, I' \subset J \\ \ell(I) = 2^{-n}\ell(J) \\ \ell(I') = 2^{-m}\ell(J)}} (f, h_I) h_{I'}$$

and h_I and $h_{I'}$ are generalized Haar functions associated to the cubes I and I' respectively. We say that S is a Haar shift of complexity (m, n) if it has a Haar shift kernel of complexity (m, n), and it is bounded on $L^2(dx)$.

Definition 11 (Positive shift)**.** We say that an operator S is a positive shift, if there exists a 2-Carleson sequense $\{a_I\}_{I \in \mathcal{D}}$, such that

$$Sf(x) = \sum_{J \in \mathcal{D}} a_J \langle f \rangle_J \chi_J.$$

Definition 12 (Orlitz norm)**.** Given a Young function Φ we define the corresponding Orlitz norm on a cube J by

$$\|f\|_{J,\Phi} = \|f\|_{L^\Phi(J)} = \|f\|_{L^\Phi_J} = \inf \left\{ \lambda > 0 : \frac{1}{|J|} \int_J \Phi\big(|f(x)|/\lambda\big)\, dx \leq 1 \right\}.$$

Finally, we translate several notions to the metric space setting.

Definition 13 (Doubling metric space)**.** A space X with a metric ρ is called doubling if there exists a measure μ on X and a constant C, such that for any ball $B(x, r) = \{y \colon \rho(x, y) < r\}$ the following holds:

$$\mu(B(x, 2r)) \leqslant C\mu(B(x, r)).$$

In what follows all metric spaces are assumed to be doubling.

Definition 14 (A Calderón-Zygmund kernel)**.** Let $\lambda(x, r)$ be a positive function, increasing and doubling in r, i.e. $\lambda(x,2r) \leqslant C\lambda(x,r)$, where C does not depend on x and r. $K(x,y)\colon X \times X \to \mathbb{R}$ is a Calderón-Zygmund kernel, associated to a function λ, if there exist positive numbers C, ε, such that

$$|K(x,y)| \leqslant C \min\left(\frac{1}{\lambda(x,\rho(x,y))}, \frac{1}{\lambda(y,\rho(x,y))}\right), \tag{1.6}$$

$$|K(x,y) - K(x',y)| \leqslant C\frac{\rho(x,x')|^{\varepsilon}}{\rho(x,y)^{\varepsilon}\lambda(x,\rho(x,y))}, \quad \rho(x,y) \geqslant C\rho(x,x'), \tag{1.7}$$

$$|K(x,y) - K(x,y')| \leqslant C\frac{\rho(y,y')^{\varepsilon}}{\rho(x,y)^{\varepsilon}\lambda(y,\rho(x,y))}, \quad \rho(x,y) \geqslant C\rho(y,y'). \tag{1.8}$$

Definition 15 (Calderón-Zygmund operator)**.** Let λ and K be as in the previous definition. Let μ be a measure on X, such that $\mu(B(x,r)) \leqslant C\lambda(x,r)$, where C does not depend on x and r. We say that T is a Calderón-Zygmund operator with kernel K if

$$T \text{ is bounded } L^2(\mu) \to L^2(\mu), \tag{1.9}$$

$$Tf(x) = \int K(x,y)f(y)d\mu(y), \ \forall x \notin \operatorname{supp} f d\mu. \tag{1.10}$$

Definition 16 (A_2 weights in metric spaces)**.** Let μ be a doubling measure. We say that a weight w belongs to $A_{2,\mu}$ if

$$[w]_{2,\mu} = \sup_{x,r} \frac{1}{\mu(B(x,r))} \int\limits_{B(x,r)} w d\mu \cdot \frac{1}{\mu(B(x,r))} \int\limits_{B(x,r)} w^{-1} d\mu < \infty.$$

The measure μ will always be fixed and we will supress the subindex μ.

1.2 Some known theorems

In this section we collect the theorems that are known and that we will use without proofs.

Theorem 1.2.1 (Hytönen's decomposition). *In a space* $\mathbb{R}^d$*, set* $\mathcal{D}^0 = \{2^{-k}([0,1)^d + m) \colon k \in \mathbb{Z}, m \in \mathbb{Z}^d\}$*. For a binary family* $\omega = (\omega_j)_{j\in\mathbb{Z}}$*,* $\omega_j \in \{0,1\}$*, set* $I + \omega = I + \sum_{j\colon 2^{-j} < \ell(I)} 2^{-j}\omega_j$*. Denote* $\mathcal{D}^\omega = \{I + \omega \colon I \in \mathcal{D}^0\}$*. For the canonical probability on set of binary sequences, and for a Calderón-Zygmund operator* T *it is true that*

$$(Tf, g) = c(T) \cdot \mathbb{E}_\omega \sum_{i\geqslant 0, j\geqslant 0} \tau_{ij}(S^{ij}_\omega f, g), \quad f, g \in C_0^\infty,$$

where $\tau_{ij} \leqslant c2^{-i+j}$*, and* S^{ij}_ω *is a generalized dyadic shift of complexity* (i,j)*.*

In the Section 4.3 we prove a version of this theorem for metric spaces. The next theorem is another way to estimate a Calderón-Zygmund operator.

Theorem 1.2.2 (Lerner's decomposition). *For a Calderón-Zygmund operator* T *and a function Banach space* X *it is true that*

$$\|T\|_X \leqslant C(T, X) \cdot \sup \|S\|_X,$$

where S *is a positive dyadic shift with a* 2*-Carleson sequense* $\{a_I\}$*. The supremum is taken over all dyadic grids on* $\mathbb{R}^d$ *and all* 2*-Carleson sequences.*

The next theorem shows how to estimate such operators.

Theorem 1.2.3. *Let S be a positive Haar shift of complexity (m,n). Then*

$$\|S(\cdot\sigma)\|_{L^p(\sigma)\to L^p(u)} \leqslant \tau\|M(\cdot\sigma)\|_{L^p(\sigma)\to L^p(u)} + \sup_Q \frac{\|\chi_Q S(\chi_Q\sigma)\|_{L^p(u)}}{\sigma(Q)^{\frac{1}{p}}} + \sup_Q \frac{\|\chi_Q S^*(\chi_Q u)\|_{L^{p'}(\sigma)}}{u(Q)^{\frac{1}{p'}}}, \quad (1.11)$$

where M is the Hardy-Littlewood maximal function.

Chapter 2

Two weight estimates

2.1 Main results

Suppose T is a Calderón-Zygmund operator on $\mathbb{R}$, and u, v are weights. By T_u we denote the operator that acts as $T_u(f) = T(fu)$. The question one asks it the following:

What conditions should weights u and v have to assure that the operator T_u is bounded from $L^2(u)$ to $L^2(\sigma)$?

Thus questions got some attention recently in the works of F. Nazarov, S. Treil, A. Volberg, A. Lerner, M. Lacey, E. Sawyer, C.Y. Shen, I. Uriarte-Tuero, D. Cruz-Uribe, C. Perez, J.M. Martell.

This question was considered for individual operators: Haar shift, see [NTV2, NTV3] and Hilbert Transform, see [NTV4, LSUT, LSSUT, Lac1, Lac2]. The conditions in these questions were formulated in terms of these individual operators.

In our papers [CURV, NRTV1, NRTV2, NRV1] we consider the "bump" approach to this problem. We give a condition on weights u, σ which assures that for **any** Calderón-Zygmund Operator T, T_u is bounded from $L^2(u)$ to $L^2(\sigma)$. In [CURV] we also have some L^p results.

The "bump" condition appears from the famous Sarason conjecture.

Conjecture. If there exists a number R, such that for any interval I we have $P_u(z) \cdot P_v(z) \leqslant$

R, then the Hilbert Transform H_u is bounded from $L^2(u)$ to $L^2(\sigma)$. Here

$$P_u(z) = \int\limits_{\mathbb{R}} \frac{\mathrm{Im} z}{(\mathrm{Re} z - t)^2 + (\mathrm{Im} z)^2} u(t) dt.$$

This conjecture is known to be **false**, see [NV] or [LSUT] for a counterexample.

The bump approach appeared in works of C. Fefferman, [F], Chang-Wilson-Wolf, [CWW]. For more history we refer the reader to the book [CUMP1]. In fact, the A_2 condition for a weight w reads as $\langle w \rangle_I \langle w^{-1} \rangle_I \leqslant Q$. In our setting we have two weights u, v, and the condition $\langle u \rangle_I \langle v \rangle_I \leqslant Q$ is even weaker than the one in the Sarason conjecture (just take $z = c_I + |I| i$, where c_I is the center of the interval I). We notice that $\langle u \rangle_I$ is the squared $L^2(\frac{dx}{|I|})$ norm of the function $u^{\frac{1}{2}}$ on the interval I. We try to consider a stronger norm. Precisely, let $A\colon [0, \infty) \to [0, \infty)$ be a Young function. In [CURV] we prove the following theorems.

Theorem 2.1.1 (Separated bump conjecture). *Suppose $A(t) = t^2 \log^{1+\varepsilon}(t)$. Then if for any interval I*

$$\|u^{\frac{1}{2}}\|_{I,A} \langle \sigma \rangle_I^{\frac{1}{2}} + \|\sigma^{\frac{1}{2}}\|_{I,A} \langle u \rangle_I^{\frac{1}{2}} \leqslant Q,$$

then the operator T_u is bounded from $L^2(u)$ to $L^2(\sigma)$.

The same is true for $A(t) = t^2 \log(t) (\log\log(t))^{1+\varepsilon}$ for sufficiently big ε.

Theorem 2.1.2 (Weak separated bump conjecture). *Suppose $A(t) = t^2 \log^{1+\varepsilon}(t)$. Then if for any interval I*

$$\|v^{\frac{1}{2}}\|_{I,A} \langle u \rangle_I^{\frac{1}{2}} \leqslant Q,$$

then the operator T_u is bounded from $L^2(u)$ to $L^{2,\infty}(v)$.

The same is true for $A(t) = t^2 \log(t) (\log\log(t))^{1+\varepsilon}$ for sufficiently big ε.

Theorem 2.1.3 (Bump conjecture)**.** *Suppose* $\Phi(t)$ *is a young function, such that* $\frac{1}{\Phi(t)}$ *is integrable at* ∞*. Then if for any interval* I

$$\|u\|_{I,\Phi} \cdot \|v\|_{I,\Phi} \leqslant Q,$$

then the operator T_w *is bounded from* $L^2(u)$ *to* $L^2(v)$*.*

Notice that while the bump conjecture is proven in full, the separated bump conjecture is proven only for some functions A. In what follows we give the precise conditions on A for which we can prove the result.

2.2 The stopping time approach to the separated bump conjecture

We start with the following definition.

Definition 17. We will say that a Young function $\bar{A}$ satisfies the $B_{p'}$ condition, $1 < p < \infty$, if for some $c > 0$,

$$\int_c^\infty \frac{\bar{A}(t)}{t^{p'}} \frac{dt}{t} < \infty.$$

If A and $\bar{A}$ are doubling (i.e., if $A(2t) \leq CA(t)$, and similarly for $\bar{A}$), then $\bar{A} \in B_p$ if and only if

$$\int_c^\infty \left(\frac{t^p}{A(t)}\right)^{p'-1} \frac{dt}{t} < \infty.$$

Examples of such bumps are

$$A(t) = t^p \log(e+t)^{p-1+\delta}, \quad \bar{A}(t) \approx \frac{t^{p'}}{\log(e+t)^{1+\delta'}}, \tag{2.1}$$

$$B(t) = t^{p'} \log(e+t)^{p'-1+\delta}, \qquad \bar{B}(t) \approx \frac{t^p}{\log(e+t)^{1+\delta''}}\,, \tag{2.2}$$

where $\delta > 0$, $\delta' = \delta/(p-1)$, $\delta'' = \delta/(p'-1)$ Given p, $1 < p < \infty$, let A and B be Young functions such that $\bar{A} \in B_{p'}$ and $\bar{B} \in B_p$. Our main condition on weights u and σ will be

$$\sup_Q \langle u\rangle_Q^{1/p} \|\sigma^{1/p'}\|_{B,Q} < \infty, \tag{2.3}$$

$$\sup_Q \|u^{1/p}\|_{A,Q} \langle\sigma\rangle_Q^{1/p'} < \infty. \tag{2.4}$$

Remark 1. By the properties of the Luxemburg norm we have that either condition implies the two-weight A_p condition:

$$\sup_Q \langle u\rangle_Q^{1/p} \langle\sigma\rangle_Q^{1/p'} < \infty. \tag{2.5}$$

Recall that the Hardy-Littlewood maximal operator is defined to be

$$Mf(x) = \sup_{Q \ni x} \langle |f| \rangle_Q = \sup_{Q \ni x} \|f\|_{1,Q}.$$

Given a Young function A, we define the Orlicz maximal operator M_A by

$$M_A f(x) := \sup_{Q \ni x} \|f\|_{A,Q}.$$

The following result is due to Pérez [Pe] (see also [1]).

Theorem 2.2.1. *Fix p, $1 < p < \infty$, and let A be a Young function such that $A \in B_p$. Then $M_A : L^p \to L^p$.*

The B_p condition is also sufficient for a two-weight norm inequality for the Hardy-Littlewood maximal operator. This result is also due to Pérez [Pe, 1].

Theorem 2.2.2. *Fix p, $1 < p < \infty$, and let B be a Young function such that $\bar{B} \in B_p$. If the pair of weights (u, σ) satisfies*

$$\sup_Q \langle u \rangle_Q^{1/p} \|\sigma^{1/p'}\|_{B,Q} < \infty, \tag{2.6}$$

then

$$\|M(f\sigma)\|_{L^p(u)} \leq C\|f\|_{L^p(\sigma)}. \tag{2.7}$$

By the decomposition theorem of Lerner, to prove the separated bump conjecture it will suffice to prove that they hold for a positive dyadic shift. More precisely we will prove the following.

Theorem 2.2.3. *Given p, $1 < p < \infty$, suppose A and B are log-bumps of the form* (2.1)*,* (2.2)*, and the pair of weights (u, σ) satisfies* (2.3) *and* (2.4)*. Given any positive dyadic shift S it holds that*

$$\|S(f\sigma)\|_{L^p(u)} \leq C\|f\|_{L^p(\sigma)}.$$

Theorem 2.2.4. *Given p, $1 < p < \infty$, suppose A is a log-bump of the form* (2.1)*, and the pair of weights (u, σ) satisfies* (2.4)*. Given any positive dyadic shift S it holds that*

$$\|S(f\sigma)\|_{L^{p,\infty}(u)} \leq C\|f\|_{L^p(\sigma)}.$$

By the Theorem 1.2.3 is suffices to estimate $\|\chi_I S(\chi_I \sigma)\|_{L^p(u)}$. The dual estimate will be the same.

Fix a cube Q_0; using the notation from the definition of a Haar shift, we have that

$$\chi_{Q_0} S(\chi_{Q_0}\sigma) = \sum_{R\subset Q_0} S_R(\sigma) + \chi_{Q_0} \sum_{R,\, Q_0 \subsetneq R} S_R(\chi_{Q_0}\sigma) \leqslant \sum_{R\subset Q_0} S_R(\sigma) + \chi_{Q_0}\langle\sigma\rangle_{Q_0}. \quad (2.8)$$

The second inequality is straightforward: see, for instance, [H, HyLa, HLM+, HPTV]. As we noted above, the pair (u, σ) satisfies the two-weight A_p condition (2.5). Therefore, the $L^p(u)$ norm of the second term is bounded by

$$\|\chi_{Q_0}\|_{L^p(u)}\langle\sigma\rangle_{Q_0} = \langle u\rangle_{Q_0}^{1/p}\langle\sigma\rangle_{Q_0}^{1/p'}\sigma(Q_0)^{1/p} \leq C\sigma(Q_0)^{1/p}\,.$$

To estimate the $L^p(u)$ norm of the first term, we form the following decomposition (see [HyLa]):

$\mathcal{K} = \mathcal{K}_i = \{Q \subset Q_0 \colon \ell(Q) = 2^{i+\tau n}\}$, $n \in \mathbb{Z}$;

$\mathcal{K}_a = \{Q \in \mathcal{K} \colon 2^a \leqslant \langle u\rangle_Q^{\frac{1}{p}}\langle\sigma\rangle_Q^{\frac{1}{p'}} < 2^{a+1}\}$;

$\mathcal{P}_0^a$ = all maximal cubes in $\mathcal{K}_a$;

$\mathcal{P}_n^a = \left\{\text{maximal cubes } P' \subset P \in \mathcal{P}_{n-1}^a, \text{ such that } P' \in \mathcal{K}_a,\ \langle\sigma\rangle_{P'} > 2\langle\sigma\rangle_P\right\}$;

$\mathcal{P}^a = \bigcup_{n\geqslant 0} \mathcal{P}_n^a$.

Following the terminology from [LPR], we call members of $\mathcal{P}^a$ *principal cubes.*

Hereafter we suppress the index i; this will give us a sum with $\tau+1$ terms. Given $Q \in \mathcal{K}_a$, let $\Pi(Q)$ denote the minimal principal cube that contains it, and define

$$\mathcal{K}_a(P) = \{Q \in \mathcal{K}_a : \Pi(Q) = P\}.$$

We will estimate the $L^p(u)$ norm of the first sum on the right-hand side of (2.8) using the exponential decay distributional inequality originated in [LPR]. Below, S is any positive generalized Haar shift that is bounded on unweighted L^2. In particular, we will take S to be one of the positive Haar shifts $S_{\mathcal{L}}$ from above.

We need the following notation. For a family $\mathcal{S}$ we denote

$$S_{\mathcal{Q}} = \sum_{Q \in \mathcal{Q}} S_Q.$$

The following distributional inequality holds.

Theorem 2.2.5. *There exists a constant c, depending only on the dimension and the unweighted L^2 norm of the shift, such that for any $P \in \mathcal{P}^a$,*

$$u\left(x \in P : |S_{\mathcal{K}_a(P)}(\sigma)| > t\frac{\sigma(P)}{|P|}\right) \lesssim e^{-c\,t}\,u(P).$$

It follows from Theorem 2.2.5 that for some positive constant c,

$$\|\sum_{R \subsetneq Q} S_R(\sigma)\|_{L^p(u)} \leqslant C\tau \sum_a \left(\sum_{P \in \mathcal{P}^a} u(P)\left(\frac{\sigma(P)}{|P|}\right)^p\right)^{\frac{1}{p}}. \tag{2.9}$$

We sketch the proof of (2.9) following the beautiful calculations in [HyLa]:

$$\sum_{R \subsetneq Q} S_R(\sigma) = \sum_{i=0}^{\tau} \sum_a \sum_{P \in \mathcal{P}^a} S_{\mathcal{K}^a(P)}(\sigma)\,,$$

and so

$$\| \sum_{R \subsetneq Q} S_R(\sigma) \|_{L^p(u)} \leqslant (\tau+1) \sum_a \| \sum_{P \in \mathcal{P}^a} S_{\mathcal{K}^a(P)}(\sigma) \|_{L^p(u)} \,.$$

Fix a. Using Fubini's theorem we write

$$\begin{aligned}
&\| \sum_{P \in \mathcal{P}^a} S_{\mathcal{K}^a(P)}(\sigma) \|_{L^p(u)} \\
&\qquad = \left(\int \left(\sum_j \sum_{P \in \mathcal{P}^a} \chi_{\{S_{\mathcal{K}^a(P)}(\sigma) \in (j, j+1] \frac{\sigma(P)}{|P|}\}} S_{\mathcal{K}^a(P)}(\sigma)(x) \right)^p u(x)\, dx \right)^{1/p} \\
&\qquad \leq \sum_j (j+1) \left(\int \left[\sum_{P \in \mathcal{P}^a} \chi_{\{S_{\mathcal{K}^a(P)}(\sigma) \in (j, j+1] \frac{\sigma(P)}{|P|}\}} \frac{\sigma(P)}{|P|} \right]^p u(x)\, dx \right)^{1/p} .
\end{aligned}$$

By the choice of the stopping cubes $P \in \mathcal{P}^a$ we have that

$$\left[\sum_{P \in \mathcal{P}^a} \chi_{\{S_{\mathcal{K}^a(P)}(\sigma) \in (j, j+1] \frac{\sigma(P)}{|P|}\}} \frac{\sigma(P)}{|P|} \right]^p \lesssim \sum_{P \in \mathcal{P}^a} \chi_{\{S_{\mathcal{K}^a(P)}(\sigma) \in (j, j+1] \frac{\sigma(P)}{|P|}\}} \left(\frac{\sigma(P)}{|P|} \right)^p .$$

Let us explain it. Take a point x and nested cubes $P_0 \supset P_1 \supset \ldots \supset P_N$, $P_k \in \mathcal{P}^a_{n_k}$, $x \in P_k$, and $x \in \{S_{\mathcal{K}^a(P)}(\sigma) \in (j, j+1] \frac{\sigma(P)}{|P|}\}$ (i.e., for which the terms in sums above are non-zero). By definition of $\mathcal{P}^a_{n_k}$ we have $\langle \sigma \rangle_{P_k} > 2 \langle \sigma \rangle_{P_{k-1}}$. The inequality may be much better if we skip several generations, but 2 in the right-hand side is guaranteed. We have a sequense $y_k = \frac{\sigma(P_k)}{|P_k|}$. The inequality says that

$$\left(\sum_k y_k \right)^p \leqslant C \; \sum_k y_k^p.$$

Notice that $y_k > 2 y_{k-1}$. Thus,

$$\sum_k y_k \leqslant \sum_{k=0}^N \frac{1}{2^{N-k}} y_N \lesssim y_N.$$

Therefore,

$$\left(\sum_k y_k\right)^p \lesssim y_N^p \leqslant \sum_k y_k^p.$$

This beautiful observation from [HyLa] lets us write

$$\|\sum_{P\in\mathcal{P}^a} S_{\mathcal{K}^a(P)}(\sigma)\|_{L^p(u)}$$
$$\lesssim \sum_j (j+1)\Big(\sum_{P\in\mathcal{P}^a}\left(\frac{\sigma(P)}{|P|}\right)^p u(S_{\mathcal{K}^a(P)}(\sigma)\in(j,j+1]\frac{\sigma(P)}{|P|})\Big)^{1/p}.$$

Then by the distributional inequality from Theorem 2.2.5:

$$\|\sum_{P\in\mathcal{P}^a} S_{\mathcal{K}^a(P)}(\sigma)\|_{L^p(u)} \lesssim \sum_j (j+1)e^{-cj/p}\Big(\sum_{P\in\mathcal{P}^a}\left(\frac{\sigma(P)}{|P|}\right)^p u(P)\Big)^{1/p}.$$

This gives us (2.9).

It is at this point in the proof that we can no longer assume that our pair of weights (u,σ) satisfies the general A_p bump condition and we must instead make the more restrictive assumption that we have log bumps. Before doing so, however, we want to show how the proof goes and where the problem arises for general bumps. We will then give the modification necessary to make this argument work for log bumps.

Define the sequence

$$\mu_Q = \begin{cases} |P|, & Q = P\,,\text{for some cube } P\in\mathcal{P}^a \\ 0, & \text{otherwise;}\end{cases}$$

then the inner sum in (2.9) becomes

$$\sum_{Q\subset Q_0} \frac{u(Q)}{|Q|}\left(\frac{\sigma(Q)}{|Q|}\right)^p \mu_Q.$$

But by Hölder's inequality in the scale of Orlicz spaces,

$$\frac{\sigma(Q)}{|Q|} = \langle \sigma^{\frac{1}{p}}\sigma^{\frac{1}{p'}}\rangle_Q \leqslant C\|\sigma^{\frac{1}{p'}}\|_{Q,B}\|\sigma^{\frac{1}{p}}\|_{Q,\bar{B}} \leqslant \|\sigma^{\frac{1}{p'}}\|_{Q,B} \inf_{x\in Q} M_{\bar{B}}(\sigma^{\frac{1}{p}}\chi_Q). \tag{2.10}$$

Therefore, by (2.3),

$$\sum_{Q\subset Q_0} \frac{u(Q)}{|Q|}\left(\frac{\sigma(Q)}{|Q|}\right)^p \mu_Q \leqslant K^p \sum_{Q\subset Q_0} \mu_Q \inf_{x\in Q} M_{\bar{B}}(\sigma^{\frac{1}{p}}\chi_Q)^p. \tag{2.11}$$

To complete the proof we need two lemmas. The first can be found in [LPR], formula (3.3).

Lemma 2.2.6. $\{\mu_Q\}$ *is a Carleson sequence.*

The second is a folk theorem; a proof can be found in [MP].

Lemma 2.2.7. *If* $\{\mu_Q\}$ *is a Carleson sequence, then*

$$\sum_{Q\subset Q_0} \mu_Q \inf_Q \chi_{Q_0} F(x) \lesssim \int_{Q_0} F(x)dx.$$

Combining these two lemmas with Theorem 2.2.1 (since $\bar{B} \in B_p$) we see that

$$\sum_Q \frac{u(Q)}{|Q|} \left(\frac{\sigma(Q)}{|Q|}\right)^p \mu_Q \leqslant K^p \sum_{Q\,,\,Q\subset Q_0} \mu_Q \inf_{x\in Q} M_{\bar{B}}(\sigma^{\frac{1}{p}}\chi_{Q_0})^p$$
$$\lesssim K^p \|M_{\bar{B}}(\sigma^{\frac{1}{p}}\chi_{Q_0})\|^p_{L^p(dx)} \lesssim K^p \|\sigma^{\frac{1}{p}}\chi_{Q_0}\|^p_{L^p(dx)} = K^p\sigma(Q_0).$$

This would complete the proof except that we must now sum over a, and in (2.9) this sum goes from $-\infty$ to the logarithm of the two-weight A_p constant of the pair (u, σ). We cannot evaluate this sum unless we can modify the above argument to yield a decay constant in a. In the one-weight argument in [HyLa] the authors could use the fact that the parameter a run from 0 to the logarithm of A_p constant: this follows since by Hölder's inequality the A_p constant of any weight is at least 1. In the two-weight case the A_p constant can be arbitrarily small, and therefore we must sum over infinitely many values of a. We are able to get the desired decay constant only by assuming that we are working with log bumps.

We modify the above argument as follows. Essentially, we will use the properties of log bumps to replace $\bar{B}$ with a slightly larger Young function. Define $B_0(t) = t^{p'}\log(e+t)^{p'-1+\frac{\delta}{2}}$; then we again have that $\bar{B}_0 \in B_p$. Instead of (2.11) we will prove that there exists γ, $0 < \gamma < 1$, such that

$$\sum_{Q\subset Q_0} \frac{u(Q)}{|Q|} \left(\frac{\sigma(Q)}{|Q|}\right)^p \mu_Q \leqslant C_\gamma K^{(1-\gamma)p} 2^{a\gamma p} \sum_{Q\subset Q_0} \mu_Q \inf_{x\in Q} M_{\bar{B}_0}(\sigma^{\frac{1}{p}}\chi_{Q_0})^p. \tag{2.12}$$

Given inequality (2.12), we can repeat the argument above, but we now have the decay term $2^{a\gamma p}$ which allows us to sum in a and get the desired estimate.

To prove (2.12) suppose for the moment that there exists γ such that

$$\|\sigma^{\frac{1}{p'}}\|_{Q,B_0} \le C_1 \|\sigma^{\frac{1}{p'}}\|_{Q,B}^{1-\gamma} \|\sigma^{\frac{1}{p'}}\|_{L^{p'}(Q,dx/|Q|)}^{\gamma}. \tag{2.13}$$

Given this, fix a cube $Q \in \mathcal{P}^a$—we can do this since otherwise $\mu_Q = 0$. Then

$$\begin{aligned}
&\frac{u(Q)}{|Q|}\left(\frac{\sigma(Q)}{|Q|}\right)^p \\
&\qquad \lesssim \langle u\rangle_Q \|\sigma^{1/p'}\|_{B_0,Q}^p \|\sigma^{1/p}\|_{\bar{B}_0,Q}^p \\
&\qquad \lesssim \langle u\rangle_Q \|\sigma^{1/p'}\|_{B,Q}^{(1-\gamma)p} \|\sigma^{1/p'}\|_{L^{p'}(Q,dx/|Q|)}^{\gamma p} \|\sigma^{1/p}\|_{\bar{B}_0,Q}^p \\
&\qquad = (\langle u\rangle_Q^{1/p} \|\sigma^{1/p'}\|_{B,Q})^{(1-\gamma)p} \cdot (\langle u\rangle_Q^{1/p} \|\sigma^{1/p'}\|_{L^{p'}(Q,dx/|Q|)})^{\gamma p} \cdot \|\sigma^{1/p}\|_{\bar{B}_0,Q}^p \\
&\qquad \lesssim K^{(1-\gamma)p} \cdot (\langle u\rangle_Q^{1/p} \langle \sigma\rangle_Q^{1/p'})^{\gamma p} \cdot \|\sigma^{1/p}\|_{\bar{B}_0,Q}^p \\
&\qquad \lesssim K^{(1-\gamma)p} \cdot 2^{a\gamma p} \cdot \|\sigma^{1/p}\|_{\bar{B}_0,Q}^p \\
&\qquad \lesssim K^{(1-\gamma)p} \cdot 2^{a\gamma p} \cdot \inf_{x\in Q} M_{\bar{B}_0}(\sigma^{\frac{1}{p}}\chi_{Q_0})^p.
\end{aligned}$$

Inequality (2.12) now follows immediately.

Therefore, to complete the proof we must establish (2.13). By the rescaling properties of the Luxemburg norm [1, Section 5.1], the right-hand side of this inequality is equal to

$$\|\sigma^{\frac{1-\gamma}{p'}}\|_{C,Q} \|\sigma^{\frac{\gamma}{p'}}\|_{p'/\gamma,Q},$$

where $C(t) = B(t^{\frac{1}{1-\gamma}})$. Therefore, by the generalized Hölder's inequality in Orlicz spaces ([1, Lemma 5.2]), inequality (2.13) holds if for all $t > 1$,

$$C^{-1}(t)t^{\frac{\gamma}{p'}} \lesssim B_0^{-1}(t). \tag{2.14}$$

A straightforward calculation (see [1, Section 5.4]) shows that

$$C^{-1}(t) = B^{-1}(t)^{1-\gamma} \approx \frac{t^{\frac{1-\gamma}{p'}}}{\log(e+t)^{\frac{1-\gamma}{p}+\frac{\delta(1-\gamma)}{p'}}}, \qquad B_0^{-1}(t) \approx \frac{t^{\frac{1}{p'}}}{\log(e+t)^{\frac{1}{p}+\frac{\delta}{2p'}}}.$$

By equating the exponents on the logarithm terms, we see that (2.14) holds if we take

$$\gamma = \frac{\delta}{2(p'-1+\delta)}.$$

Therefore, with this value of γ inequality (2.13) holds, and this completes our proof.

For the convenience of the reader we give a direct proof of (2.13); this computation will also be used below in Section 2.2.2. The desire inequality obviously follows from the following lemma.

Lemma 2.2.8. *Given a probability measure μ, let f be a non-negative measurable function. Let B, B_0 be logarithmic bumps as in (2.2) with $\delta = \tau$ and $\delta = \frac{\tau}{2}$ respectively. Then there exists an absolute constant C and $\gamma = \gamma(p', \tau) > 0$ such that*

$$\|f\|_{B_0,\mu} \le C\, \|f\|_{B,\mu}^{1-\gamma}\, \|f\|_{L^{p'}(\mu)}^{\gamma}. \tag{2.15}$$

Proof. We will actually show that $\gamma = \frac{1}{2+(p'-1)\frac{2}{\tau}}$. Define $\Delta := \int |f|^{p'}\, d\mu$. Since inequality (2.15) is homogeneous, we may assume without loss of generality that

$$\|f\|_{B,\mu} = 1. \tag{2.16}$$

Moreover, we may assume that $\Delta \le 1$: otherwise (2.15) can be achieved by choosing C

sufficiently large. Let $\epsilon < 1$ and K be constants; we will determine their precise value (in this order) below. Then we have that

$$
\begin{aligned}
&\int \frac{f^{p'}}{\epsilon^{p'}} \log\left(e + \frac{f}{\epsilon}\right)^{p'-1+\frac{\tau}{2}} d\mu \\
&\quad \le \int_{\{f \le K\epsilon\}} \cdots + \int_{\{f \ge K\epsilon\}} \cdots \\
&\quad \le \frac{\Delta}{\epsilon^{p'}} [\log(e+K)]^{p'-1+\frac{\tau}{2}} + \int_{\{f \ge K\epsilon\}} \frac{f^{p'}}{\epsilon^{p'}} \frac{\log(e + \frac{f}{\epsilon})^{p'-1+\tau}}{[\log(e+K)]^{\frac{\tau}{2}}} d\mu \\
&\quad \le \frac{\Delta}{\epsilon^{p'}} [\log(e+K)]^{p'-1+\frac{\tau}{2}} + \int \frac{f^{p'}}{\epsilon^{p'}} \frac{\log(\frac{e}{\epsilon} + \frac{f}{\epsilon})^{p'-1+\tau}}{[\log(e+K)]^{\frac{\tau}{2}}} d\mu \\
&\quad \le \frac{\Delta}{\epsilon^{p'}} [\log(e+K)]^{p'-1+\frac{\tau}{2}} \\
&\qquad + \frac{1}{\epsilon^{p'} [\log(e+K)]^{\tau/2}} \left[\int f^{p'} \log^{p'-1+\tau}(e+f)\, d\mu + \int f^{p'} \log(\epsilon^{-1})^{p'-1+\tau} d\mu \right] \\
&\quad \le \frac{\Delta}{\epsilon^{p'}} [\log(e+K)]^{p'-1+\frac{\tau}{2}} + \frac{1}{\epsilon^{p'} [\log(e+K)]^{\tau/2}} \left[1 + \Delta \log(\epsilon^{-1})^{p'-1+\tau} \right].
\end{aligned}
$$

In the last line we used (2.16). Fix ϵ so that

$$\Delta = (\epsilon^{p'})^{1+c},$$

where $c = 1 + (p'-1)\frac{2}{\tau}$. In other words,

$$\epsilon = (\Delta^{1/p'})^{\gamma} = \|f\|_{L^{p'}(\mu)}^{\gamma}, \; \gamma = \frac{1}{1+c}.$$

Now choose K so that

$$[\log(e+K)]^{\tau/2} \approx \epsilon^{-p'};$$

then

$$[\log(e+K)]^{p'-1+\tau/2} \approx (\epsilon^{-p'})^{1+(p'-1)\frac{2}{\tau}} =: (\epsilon^{-p'})^c.$$

If we substitute these values into the above calculation, we see that the right hand side is dominated by a constant. Hence, by the definition of the Luxemburg norm,

$$\|f\|_{B_0,\mu} \leq C\,\epsilon = C\,\|f\|^{\gamma}_{L^{p'}(\mu)}.$$

This completes the proof. □

Remark 2. The conjugate testing condition can be verified similarly. The adjoint S^* is also a Haar shift, and so we can apply the distribution inequality from Theorem 2.9 to it. Also, the second sum in (2.8) will have the same pointwise estimate (exchanging σ and v) if we replace S with S^*.

Remark 3. In the proof of the first testing condition we only used the bump condition (2.3); to prove the second testing condition we use the second bump condition (2.4).

2.2.1 Proof of Theorem 2.2.4

The proof of the weak-type inequality uses essentially the same argument as above; here we sketch the changes required. We repeat the argument, replacing the $L^p(u)$ norm with the $L^{p,\infty}(u)$ norm. Since the pair (u, σ) satisfies the two-weight A_p condition we have the well known inequality that

$$\|M(f\sigma)\|_{L^{p,\infty}(u)} \leq C\|f\|_{L^p(\sigma)},$$

where the constant C depends only on the A_p constant and the dimension. Therefore it remains to estimate the $L^{p,\infty}(u)$ norm of $S_{\mathcal{L}}(|f|\sigma)$. However, from Hytönen, *et al.* [HLM+, Theorem 4.3] we have the following analog of Theorem 1.2.3.

Theorem 2.2.9. *Let S be a positive Haar shift of complexity (m, n). Then*

$$\|S(\cdot\sigma)\|_{L^p(\sigma)\to L^{p,\infty}(u)} \leqslant \tau\|M(\cdot\sigma)\|_{L^p(\sigma)\to L^{p,\infty}(u)} + \sup_Q \frac{\|\chi_Q S^*(\chi_Q u)\|_{L^{p'}(\sigma)}}{u(Q)^{\frac{1}{p'}}}.$$

Given Theorem 2.2.9 the argument now proceeds exactly as before, using the bump condition (2.4) to bound the testing condition. This completes the proof.

2.2.2 Proof for the log log-bumps

In this section we consider bumps of the following form.

$$A(t) = t^p \log(e+t)^{p-1} \log\log(e^e+t)^{p-1+\delta} \quad \bar{A}(t) \approx \frac{t^{p'}}{\log(e+t)\log\log(e^e+t)^{1+\delta'}}, \tag{2.17}$$

$$B(t) = t^{p'} \log(e+t)^{p'-1} \log\log(e^e+t)^{p'-1+\delta}, \quad \bar{B}(t) \approx \frac{t^{p}}{\log(e+t)\log\log(e^e+t)^{1+\delta''}}, \tag{2.18}$$

where $\delta > 0$. Our proofs are very similar to the proofs given in previous sections, so we will describe the principle changes. As before, we need to prove the following theorems for positive dyadic shifts.

Theorem 2.2.10. *Given p, $1 < p < \infty$, suppose A and B are loglog-bumps of the form* (2.17), (2.18) *with $\delta > 0$* **sufficiently large**, *and the pair of weights (u, σ) satisfies* (2.3) *and* (2.4). *Given any positive dyadic shift S, $\|S(f\sigma)\|_{L^p(u)} \leq C|f\|_{L^p(\sigma)}$.*

Theorem 2.2.11. *Given p, $1 < p < \infty$, suppose A is a loglog-bump of the form* (2.17) *with $\delta > 0$* **sufficiently large**, *and the pair of weights (u, σ) satisfies* (2.4). *Given any positive dyadic shift S, $\|S(f\sigma)\|_{L^{p,\infty}(u)} \leq C|f\|_{L^p(\sigma)}$.*

We will prove Theorem 2.2.10 by modifying the proof of Theorem 2.2.3 above; Theorem 2.2.11 is proved similarly. The main step is to adapt Lemma 2.2.8 to work with loglog-bumps. Let B be as in (2.18), and define B_0 similarly but with δ replaced by $\delta/2$. Then arguing almost exactly as we did in the proof of Lemma 2.2.8, we have that

$$\|f\|_{B_0,\mu} \le C\|f\|_{B,\mu}\, \varepsilon\Big(\frac{\|f\|_{L^{p'}(\mu)}}{\|f\|_{B,\mu}}\Big), \tag{2.19}$$

where $\varepsilon(t) = (\log \frac{C}{t})^{-\kappa}$, $C = C(p,\delta)$, and $\kappa = \kappa(p,\delta)$ with $\kappa > 1$ if δ is large enough. For the detailed proof for wider range of bumps see [NRV], Section 5.2.2.

Given (2.19) we have that

$$\begin{aligned}
&\frac{u(Q)}{|Q|}\Big(\frac{\sigma(Q)}{|Q|}\Big)^p \\
&\qquad \le C\langle u\rangle_Q \|\sigma^{1/p'}\|^p_{B_0,Q}\|\sigma^{1/p}\|^p_{\bar{B}_0,Q} \\
&\qquad \le C\langle u\rangle_Q \|\sigma^{1/p'}\|^p_{B,Q}\, \varepsilon\Big(\frac{\langle\sigma\rangle_Q^{1/p'}}{\|\sigma^{1/p'}\|_{B,Q}}\Big)^p \|\sigma^{1/p}\|^p_{\bar{B}_0,Q} \\
&\qquad \le C\left(\frac{\langle u\rangle_Q^{1/p}\|\sigma^{1/p'}\|_{B,Q}}{\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}}\right)^p \varepsilon\Big(\frac{\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}}{\langle u\rangle_Q^{1/p}\|\sigma^{1/p'}\|_{B,Q}}\Big)^p (\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'})^p\|\sigma^{1/p}\|^p_{\bar{B}_0,Q}\,.
\end{aligned}$$

To complete the proof, we need a good bound in a for the product of first three terms. Moreover, it is enough to get a good bound for negative and very big in absolute value a. Thus, we can think that $\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}$ is very small.

Consider a function

$$\varphi(t) = t\varepsilon(\frac{1}{t}).$$

Then

$$\frac{\langle u\rangle_Q^{1/p}\|\sigma^{1/p'}\|_{B,Q}}{\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}}\varepsilon\Big(\frac{\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}}{\langle u\rangle_Q^{1/p}\|\sigma^{1/p'}\|_{B,Q}}\Big)=\varphi\left(\frac{\langle u\rangle_Q^{1/p}\|\sigma^{1/p'}\|_{B,Q}}{\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}}\right).$$

Set $t_0=\frac{\langle u\rangle_Q^{1/p}\|\sigma^{1/p'}\|_{B,Q}}{\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}}$. Then $c\leqslant t_0\leqslant C\frac{1}{\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}}$. The right-hand side of the last inequality is very big. Moreover, the function $\varphi(t)=t(\log(Ct))^{-\varkappa}$ is increasing near ∞. Therefore,

$$\varphi(t_0)\leqslant C\varphi\left(\frac{1}{\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}}\right)\leqslant C_1\varphi(2^{-a}).$$

Therefore, since on all cubes $P\in\mathcal{P}^a$ we have $\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}\sim 2^a$, we get

$$C\left(\frac{\langle u\rangle_Q^{1/p}\|\sigma^{1/p'}\|_{B,Q}}{\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}}\right)^p\varepsilon\Big(\frac{\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'}}{\langle u\rangle_Q^{1/p}\|\sigma^{1/p'}\|_{B,Q}}\Big)^p(\langle u\rangle_Q^{1/p}\langle\sigma\rangle_Q^{1/p'})^p\|\sigma^{1/p}\|^p_{\bar B_0,Q}$$

$$\leqslant C_2\varepsilon(2^a)^p\|\sigma^{1/p}\|^p_{\bar B_0,Q}.$$

Using the Carleson property of the sequence μ_Q, we get

$$\sum_{Q\subset Q_0}\mu_Q\frac{u(Q)}{|Q|}\left(\frac{\sigma(Q)}{|Q|}\right)^p\leqslant C_2\varepsilon(2^a)^p\int M_{\bar B_0}(\sigma^{1/p}\chi_{Q_0})^p dx\leqslant C_3\varepsilon(2^a)^p\sigma(Q_0).$$

Thus, returning to the formula (2.9), we get

$$\sum_a\left(\sum_{P\in\mathcal{P}^a}u(P)\left(\frac{\sigma(P)}{|P|}\right)^p\right)^{\frac1p}\leqslant\sigma^{\frac1p}(Q_0)\sum_{a<0}\varepsilon(2^a).$$

By the formula tor ε, the series converge if $\varkappa>1$.

2.3 The Bellman function approach to the bump conjecture

2.3.1 Main result

In this section we will work with Orlitz functions Φ, such that $\frac{1}{\Phi}$ is integrable near infinity. We aim to prove the following theorem.

Theorem 2.3.1. *Let functions Φ_1 and Φ_2 be as above. Let the weights v, w satisfy*

$$\sup_I \|v\|_{\Phi_1,Q}\|w\|_{\Phi_2,Q} < \infty; \tag{2.20}$$

here the supremum is taken over all cubes I.

Then for any bounded Calderón-Zygmund operator T the operator T_v is bounded from $L^2(v)$ to $L^2(w)$.

2.3.2 Orlicz norms and distribution functions

Orlicz norm is not very convenient to work with, so we would like to replace it by something more tractable.

2.3.2.1 A lower bound for the Orlicz norm

We start with the remark that notation $\int_0 f(t)dt < \infty$ means that the function is integrable near zero. Similarly, $\int^\infty f(t)dt < \infty$ means that the function is integrable near infinity.

Let Φ be a continuous non-negative increasing convex function such that $\Phi(0) = 0$ and $\int^{+\infty} \frac{dt}{\Phi(t)} < +\infty$. Define $\Psi(s)$ parametrically by $\Psi(s) = \Phi'(t)$ when $s = \frac{1}{\Phi(t)\Phi'(t)}$ $(t > 0)$. Then $\Psi(s)$ is positive and decreasing for $s > 0$ and $s\Psi(s)$ is increasing. Moreover $\int_0 \frac{ds}{s\Psi(s)} <$

$+\infty$. Indeed, using our parametrization we can rewrite the last integral as

$$\int^{+\infty} \left(\frac{1}{\Phi(t)} + \frac{\Phi''(t)}{\Phi'(t)^2} \right) dt\,.$$

The first integral converges by our assumption and the second integrand has a bounded near $+\infty$ antiderivative $\frac{-1}{\Phi'(t)}$.

Let $w \geq 0$ on $I \subset \mathbb{R}^n$. Define the normalized distribution function N of w by

$$N(t) = N_I^w(t) = \frac{1}{|I|}\, |\{x \in I : w(x) > t\}| \tag{2.21}$$

Lemma 2.3.2. *Let $\Psi : (0,1] \to \mathbb{R}_+$ be a decreasing function such that the function $s \mapsto s\Psi(s)$ is increasing. Let Φ be a Young function and let*

$$\Psi(s) \leq C\Phi'(t) \qquad \textit{where} \quad s = \frac{1}{\Phi(t)\Phi'(t)}$$

for all sufficiently large t. Then for $N = N_I^w$

$$\mathbf{n}_\Psi(N) := \int_0^\infty N(t)\Psi(N(t))\, dt \leq C\|w\|_{L^\Phi(I)}\,. \tag{2.22}$$

Proof. The left hand side scales like a norm under multiplication by constants, so it is enough to show that if $\|w\|_{L^\Phi(I)} \leq 1$, i.e.,

$$\frac{1}{|I|}\int_I \Phi(w) = \int_0^\infty N(t)\Phi'(t)\, dt \leq 1$$

then $\mathbf{n}_\Psi(N)$ is bounded by a constant. Since $s\Psi(s)$ increases, we may have trouble only at

$+\infty$ It is clear that it suffices to estimate the integral over the set where $\Psi(N(t)) > \Phi'(t)$ but since Ψ is decreasing this means that $N(t) \leq C/(\Phi(t)\Phi'(t))$, so we get at most $\int^{+\infty} \Phi(t)^{-1} dt$ and we are done. □

Remark 4. In fact, for sufficiently regular Φ, the converse inequality

$$\|w\|_{L^\Phi(I)} \leq C \int_0^\infty N(t)\Psi(N(t))\, dt$$

holds for any positive decreasing integrable N. To see this, let us consider the family of Φ's such that $\Phi(t) = t\rho(t)$ and ρ is monotonically increasing and "logarithmically concave" in the sense that $\frac{t\rho'(t)}{\rho(t)}$ decreases monotonically when $t \to \infty$. We also assume of course that $\lim_{t\to\infty} \rho(t) = \infty$ and that $\rho(t) \geq 1$. Let $G(t) := N(t)\Psi(N(t))$. When t goes to infinity, N is monotonically decreasing to zero, and hence G is also monotonically decreasing (as $s\Psi(s)$ increases near zero).

Put $s^{-1} = \Phi(t)\Phi'(t) \asymp t\rho^2(t)$ (just because $\Phi'(t) \asymp \rho(t)$ by our "logarithmic concavity" of ρ assumption). Hence $s \geq c_1(t\rho^2(t))^{-1}$. Now Ψ is decreasing by definition, and this implies

$$\Psi(c_1(t\rho^2(t))^{-1}) \geq \Phi'(t) \asymp \rho(t) \geq c_2\rho(t)\,. \tag{2.23}$$

We now ask an addition to "logarithmic concavity", namely:

$$\frac{t\rho'(t)}{\rho(t)} \log \rho(t) \to 0 \text{ as } t \to \infty\,. \tag{2.24}$$

Denote $r(x) = \log(\rho(e^x))$. We required at the beginning that $\lim_{x\to\infty} r(x) = \infty$. The last inequality says in particular that $r'(x) = o(1)r(x)^{-1}$, and therefore, r' tends to zero at infinity. Thus $r(x) \le \frac{x}{3}$ for all large x. Keeping this in mind we continue.

Set $u = \frac{t\rho^2(t)}{c_1}$. Then

$$t = \frac{c_1 u}{\rho^2(t)}.$$

Thus, since ρ is an increasing to infinity function and we assume that t is sufficiently big, we get $\rho^2(t) \ge c_1$. Therefore,

$$t \le u,$$

and, thus,

$$t = \frac{c_1 u}{\rho^2(t)} \ge c_1 \frac{u}{\rho^2(u)}.$$

Hence, using (2.23), we get

$$\Psi(u^{-1}) \ge c_2\rho(t) \ge c_2\rho(c_1 \frac{u}{\rho^2(u)}) \,. \tag{2.25}$$

Next, we will prove the following inequality. Recall that $r(x) = \log(\rho(e^x))$. We claim that

$$\Delta(x) = r(x) - r(x - 2r(x) - c_0) \le C.$$

In fact, by the mean value theorem we have for certain $\xi \in (x - 2r(x) - c_0, x)$

$$\Delta(x) = (2r(x) + c_0)r'(\xi) = (2r(x) + c_0)\frac{\rho'(e^\xi)}{\rho(e^\xi)}e^\xi.$$

Since we assumed that $t\frac{\rho'(t)}{\rho(t)}$ is monotonically decreasing, we get (now using (2.24) in the second comparison below):

$$\Delta(x) \le (2r(x)+c_0)\frac{\rho'(e^{x-2r(x)-c_0})}{\rho(e^{x-2r(x)-c_0})}e^{\rho(e^{x-2r(x)-c_0})} =$$
$$= (2r(x)+c_0)\frac{o(1)}{\log\rho(e^{x-2r(x)-c_0})} = (2r(x)+c_0)\frac{o(1)}{r(x-2r(x)-c_0)} =$$
$$= o(1)\left(\frac{2r(x)+c_0}{r(x-2r(x)-c_0)} - 2 + 2\right) = o(1)\frac{2\Delta(x)+c_0}{r(x-2r(x)-c_0)} + o(1).$$

Finally, we use that $r(x-2r(x)-c_0)$ is separated from zero when x is big. Thus

$$\Delta(x) \le \Delta(x)o(1) + o(1).$$

This immediately implies $\Delta(x) = o(1)$ when $x \to \infty$, and thus

$$\Delta(x) \le C\,. \tag{2.26}$$

Let us now write what does it mean. In fact, by the definition of r and by (2.26), we can conclude that

$$C \ge r(x) - r(x-2r(x)-c_0) = \log\frac{\rho(e^x)}{\rho(e^{x-2\log\rho(e^x)-c_0})} = \log\frac{\rho(e^x)}{\rho(\frac{e^x}{c_3\rho^2(e^x)})}.$$

Thus, we get for all large u:

$$\rho(u) \le c_4\rho(\frac{u}{c_3\rho^2(u)})\,.$$

We chose $c_3 = c_1^{-1}$ and plug the above inequality into (2.25). Then we finally get

$$\Psi(u^{-1}) \ge c_5\rho(u)$$

If $N\Psi(N) = G$ then $c_6 G \geq N\rho(\frac{1}{N})$ by the previous inequality. Therefore, $N \leq c_6 G$ (we assumed that $\rho \geq 1$), and $N \leq \frac{c_6 G}{\rho(\frac{1}{N})} \leq \frac{c_6 G}{\rho(\frac{1}{c_6 G})}$. And we can continue the previous estimate: $N(t) \leq \frac{c_6 G(t)}{\rho(\frac{1}{c_6 G(t)})} \leq \frac{c_6 G(t)}{\rho(t)}$. We used here the fact that the integrability and monotonicity of G implies that $G(t) = o(\frac{1}{t})$, in particular, $G(t) < \frac{1}{c_6 t}$ for large t. But we already mentioned that $\Phi'(t) \leq c_7 \rho(t)$. Combining the last two inequalities, we get $N(t)\Phi'(t) \leq c_6 c_7 G(t)$, and we just obtained that $\int_0^\infty N(t)\Phi'(t)dt \leq c_4 c_5$.

2.3.2.2 Examples

In the above section only the behavior of Φ at $+\infty$ (equivalently, the behavior of Ψ near 0) was important, so we will concentrate our attention there.

Let $\Phi(t) = t(\ln t)^\alpha$, $\alpha > 1$ near ∞. Then

$$\Phi'(t) \sim (\ln t)^\alpha, \qquad \Phi(t)\Phi'(t) \sim t(\ln t)^{2\alpha},$$

so $\Psi(s) := (\ln(1/s))^\alpha$ satisfies the assumptions of Lemma 2.3.2: to see that we notice

$$\ln(\Phi(t)\Phi'(t)) \sim \ln t.$$

If $\Phi(t) = t \ln t (\ln\ln t)^\alpha$, $\alpha > 1$, then

$$\Phi'(t) \sim \ln t (\ln\ln t)^\alpha, \qquad \Phi(t)\Phi'(t) \sim t(\ln t)^2 (\ln\ln t)^{2\alpha}$$

and $\Psi(s) = \ln(1/s)(\ln\ln(1/s))^\alpha$ works. because again $\ln(\Phi(t)\Phi'(t)) \sim \ln t$.

Note that in both examples $\int_0 (s\Psi(s))^{-1} ds < \infty$.

The examples of Young functions with higher order logarithms are treated similarly.

2.3.3 Main result in new language

We restate our main result using the newly built functions $\Psi_{1,2}$. Let $\Psi_1, \Psi_2 : (0,1] \to \mathbb{R}_+$ be as above, i.e. for $i = 1, 2$, Ψ_i is decreasing, $s \mapsto s\Psi_i(s)$ is increasing and

$$\int_0^1 \frac{ds}{s\Psi_i(s)} < \infty.$$

Recall that for a weight w the normalized distribution function N_I^w is defined by (2.21)

Theorem 2.3.3. *Let the weights v, w satisfy*

$$\sup_I \mathbf{n}_{\Psi_1}(N_I^v)\mathbf{n}_{\Psi_2}(N_I^w) < \infty; \tag{2.27}$$

here the supremum is taken over all cubes I, and $\mathbf{n}_\Psi$ is defined by (2.22)*.*

Then for any Calderón-Zygmund operator T the operator T_v is bounded from $L^2(v)$ to $L^2(w)$.

2.3.4 General setup

Consider a measure space X with σ-finite measure μ let $\mathcal{L}_k = \{I_j^k\}_j$, $k \in \mathbb{Z}$ (or $k \in \mathbb{Z}_+$) be partitions of X into disjoint sets I_j^k, $0 < \mu(I_j^k) < \infty$.

We assume that the partition $\mathcal{L}_{k+1}$ is a refinement of $\mathcal{L}_k$.

Let $\mathfrak{A}$ be the σ-algebra generated by all the partitions $\mathcal{L}_k$. In what follows all functions on X we consider will be assumed to be $\mathfrak{A}$-measurable.

With respect to this σ-algebras we can define martingale averaging operators $\mathbb{E}_k$, and martingale difference operators $\Delta_k^n := -\mathbb{E}_k + \mathbb{E}_{k+n}$.

We adapt the following notation.

$\operatorname{ch} I$	The collection of children of $I \in \mathcal{L}$, i.e. if $I \in \mathcal{L}_n$ then $\operatorname{ch} I = \{J \in \mathcal{L}_{n+1} : J \subset I\}$.
$\operatorname{ch}_k I$	The collection of children of the order k of $I \in \mathcal{L}$; $\operatorname{ch}_0(I) = \{I\}$, $\operatorname{ch}_{k+1}(I) = \{\operatorname{ch}(J) : J \in \operatorname{ch}_k(I)\}$.
$\langle f \rangle_I$	The average of f over I, $\langle f \rangle_I = \mu(I)^{-1} \int_I f(x) d\mu(x)$;
E_I	The averaging operator, $E_I f := \langle f \rangle_I \mathbf{1}_I$; note that $E_k = \sum_{I \in \mathcal{L}_k} E_I$.
Δ_I	Martingale difference operator, $\Delta_I := -E_I + \sum_{J \in \operatorname{ch}(I)} E_J$; note that $\Delta_k = \sum_{I \in \mathcal{L}_k} \Delta_I$.
Δ_I^n	Martingale difference operator of order n, $$\Delta_I^n := -E_I + \sum_{J \in \operatorname{ch}_n(I)} E_J.$$

Since the measure μ is assumed to be fixed we sometimes will be using $|E|$ for $\mu(E)$ and dx for $d\mu(x)$. We also will be using L^2 for $L^2(\mu)$

The prototypical example is $X = \mathbb{R}$ or $\mathbb{R}^d$ with $\mathcal{L}$ being a dyadic lattice $\mathcal{D}$.

2.3.4.1 Haar shifts

We will use a slightly more general definition of a Haar shift.

Definition 18. A Haar shift $\mathbb{S}$ of complexity n is given by

$$\mathbb{S}f = \sum_{I\in\mathcal{D}} \mathbb{S}_I \Delta_I^n f,$$

where the operators $\mathbb{S}_I$ act on $\Delta_I^n L^2$ and can be represented as integral operators with kernels a_I, $\|a_I\|_\infty \le |I|^{-1}$. The latter means that for all $f, g \in \Delta_I^n L^2$

$$\langle \mathbb{S}_I f, g\rangle = \int_I \int_I a_I(x,y) f(y) g(x) dx dy.$$

This is a slightly more general definition than the one in [HPTV], but only the estimate $\|a_I\|_\infty \le |I|^{-1}$ is essential for our construction. Note also that according to the definition in [HPTV] the complexity of the corresponding shift is $n-1$, not n, which really does not matter; we just find our definition of complexity a bit more convenient.

The estimate $\|a_I\|_\infty \le |I|^{-1}$ means that the operators $\mathbb{S}_I$ are "$L^1 \times L^1$ normalized", meaning that

$$|\langle \mathbb{S}_I f, g\rangle| \le |I| \frac{\|f\|_1}{|I|} \frac{\|g\|_1}{|I|} \qquad \forall f, g \in \Delta_I^n L^2 \tag{2.28}$$

Haar shifts of complexity 1 are simply "$L^1 \times L^1$ normalized" martingale transforms; martingale transform here means in particular that the subspaces Δ_I are orthogonal, and $\mathbb{S}$ can be represented as an orthogonal sum of the operators $\mathbb{S}_I$.

A Haar shift of complexity $n \ge 2$ is not generally a martingale transform, meaning that the subspaces Δ_I^n generally intersect, so $\mathbb{S}$ does not split into direct sum of $\mathbb{S}_I$.

However, if one goes with step n, then the corresponding operator is a martingale transform, so a Haar shift of complexity n can be represented as a sum of n Haar shifts of complexity 1. Namely, for $k = 1, 2, \dots, n-1$ define

$$\mathcal{L}^k = \{I \colon I \in \mathcal{L}_{k+nj}, j \in \mathbb{Z}\},$$

and let

$$\mathbb{S}_k = \sum_{I \in \mathcal{L}^k} \mathbb{S}_I.$$

Then $\mathbb{S} = \sum_{k=0}^{n-1} \mathbb{S}_k$ and each $\mathbb{S}_k$ is a Haar shift of complexity 1 with respect to the lattice $\mathcal{L}^k$.

Remark 5. Therefore, uniform estimate for the Haar shifts of complexity 1 (i.e. for the "$L^1 \times L^1$ normalized" martingale transforms) gives the linear in complexity estimate for the general Haar shifts. Notice that the estimate does not depend on the number of children.

2.3.4.2 Paraproducts

Given the lattice $\mathcal{L}$ and a locally integrable function b, the paraproduct $\Pi = \Pi_b = \Pi_b(\mathcal{L})$ is defined as

$$\Pi f := \sum_{I \in \mathcal{L}} (E_I f)(\Delta_I b).$$

The necessary and sufficient condition for the paraproduct to be bounded is that

$$\sup_{J \in \mathcal{L}} |J|^{-1} \sum_{I \in \mathcal{L} : I \subset J} \|\Delta_I b\|_2^2 < \infty.$$

In the case of dyadic lattice in $\mathbb{R}^d$ or, more generally in the homogeneous situation, when

$$\inf_{J \in \mathcal{L}} \inf_{I \in \mathrm{ch}(J)} \frac{|I|}{|J|} > 0$$

this condition is equivalent to b belonging to the corresponding martingale BMO space $\text{BMO}_{\mathcal{L}}$

2.3.5 Reduction to the martingale case.

To reduce the problem to the martingale case we use the following result that can be found in [H] and [HPTV]:

Theorem 2.3.4. *Let T be a Calderón–Zygmund operator in $\mathbb{R}^d$. There exists a probability space $(\Omega, \mathbb{P})$ of dyadic lattices $\mathcal{D}_\omega$, such that*

$$T = C\left(\int_\Omega \sum_{n=1}^{\infty} 2^{-\varepsilon n}\, \mathbb{S}_n(\omega) d\mathbb{P}(\omega) + \int_\Omega (\Pi^1(\omega) + (\Pi^2(\omega))^*) d\mathbb{P}(\omega)\right),$$

where $\mathbb{S}_n(\omega)$ are Haar shifts of complexity n with respect to the lattice $\mathcal{D}_\omega$, $\Pi^{1,2}(\omega)$ are the paraproducts with respect to the lattice $\mathcal{D}_\omega$, $\|\Pi^{1,2}(\omega)\| \le 1$.

The constants C and ε depend on d, $\|T\|$ and Calderón–Zygmund parameters of the kernel of T.

Theorem 2.3.4 implies immediately that the main theorem (Theorem 2.3.3) follows from the theorem below.

Theorem 2.3.5. *Let the weights v, w satisfy the assumptions of Theorem 2.3.3. Then*

1. *For all Haar shifts $\mathbb{S}$ of order 1 the operators $\mathbb{S}(\cdot v)$ are uniformly bounded from $L^2(v)$ to $L^2(w)$, $\|\mathbb{S}(\cdot v)\|_{L^2(v)\to L^2(w)} \le C$, where C depends on Ψ_2, Ψ_2, the supremum in (2.27), but not on the lattice $\mathcal{L}$.*

2. *For all Haar shifts $\mathbb{S}_n$ the operators $\mathbb{S}_n(\cdot v)$ are uniformly bounded from $L^2(v)$ to $L^2(w)$ by Cn, where C .*

3. Let $\Pi = \Pi_b$ *be a paraproduct such that*

$$|J|^{-1} \sum_{I\in\mathcal{L}: I\subset J} \|\Delta_I b\|_\infty^2 |I| \le 1 \qquad \forall J \in \mathcal{L}. \tag{2.29}$$

Then the operator $\Pi(\cdot v)$ *is bounded from* $L^2(v)$ *to* $L^2(w)$ *by* C*, where again* C *depends on* Ψ_1*,* Ψ_2*, the supremum in* (2.27)*, but not on the lattice* $\mathcal{L}$*.*

Remark 6. For the homogeneous lattices, i.e. for lattices satisfying

$$\inf_{J\in\mathcal{L}} \inf_{I\in \mathrm{ch}(J)} \frac{|I|}{|J|} =: \delta > 0$$

all the normalized L^p norms $|I|^{-1/p}\|\Delta_I g\|_p$, $p \in [1,\infty]$ are equivalent in the sense of two sided estimates. So for such lattices condition (2.29) means that $\|\Pi\| \le C(\delta)$. So Theorem 2.3.5 gives the estimates that being fed to Theorem 2.3.4 imply Theorem 2.3.3.

Theorem 2.3.6. *Let* Ψ *be as above. Then for any weight* w *on* X *such that* $\mathbf{n}_\Psi(N_I^w) < \infty$ *for all* $I \in \mathcal{L}$

$$\sum_{I\in\mathcal{L}} \mathbf{n}_\Psi(N_I^w)^{-1} \left(|I|^{-1} \int_X |\Delta_I(f w^{1/2})| dx \right)^2 |I| \le C\|f\|_{L^2(dx)}^2 \tag{2.30}$$

for all $f \in L^2(dx)$*; here* $C = C(\Psi)$ *and in the summation we skip* I *on which* $w \equiv 0$*.*

Let us see that this theorem implies the condition 1 of Theorem 2.3.5. Assume, multiplying the weights by appropriate constants that the inequality

$$\mathbf{n}_{\Psi_1}(N_I^w)\mathbf{n}_{\Psi_2}(N_I^v) \le 1 \tag{2.31}$$

holds for all $I \in \mathcal{L}$. Then

$$
\begin{aligned}
|\langle \mathbb{S}(fw^{1/2}), gv^{1/2}\rangle| &\le \sum_{I\in\mathcal{L}} |\langle \mathbb{S}_I \Delta_I(fw^{1/2}), \Delta_I(gv^{1/2})\rangle| \\
&\le \sum_{I\in\mathcal{L}} |I|^{-1} \|\Delta_I(fw^{1/2})\|_1 \|\Delta_I(gv^{1/2})\|_1 \\
&\le \sum_{I\in\mathcal{L}} |I|^{-1} \frac{\|\Delta_I(fw^{1/2})\|_1 \|\Delta_I(gv^{1/2})\|_1}{\left(\mathbf{n}_{\Psi_1}(N_I^w)\mathbf{n}_{\Psi_2}(N_I^v)\right)^{1/2}} \\
&\le \frac{1}{2}\sum_{I\in\mathcal{L}} |I|^{-1} \frac{\|\Delta_I(fw^{1/2})\|_1^2}{\mathbf{n}_{\Psi_1}(N_I^w)} + \frac{1}{2}\sum_{I\in\mathcal{L}} |I|^{-1} \frac{\|\Delta_I(gv^{1/2})\|_1^2}{\mathbf{n}_{\Psi_2}(N_I^v)}\,.
\end{aligned}
$$

The second inequality here follows from "$L^1 \times L^1$ normalization" condition (2.28), the second one from (2.31) and the last one is just the trivial inequality $2xy \le x^2 + y^2$.

Applying Theorem 2.3.6 to each sum we get that

$$
|\langle \mathbb{S}(fw^{1/2}), gv^{1/2}\rangle| \le \frac{1}{2}\left(C(\Psi_1)\|f\|_2^2 + C(\Psi_2)\|g\|_2^2\right).
$$

Replacing $f \mapsto tf$, $g \mapsto t^{-1}g$, $t > 0$ we get

$$
|\langle \mathbb{S}(fw^{1/2}), gv^{1/2}\rangle| \le \frac{1}{2}\left(t^2 C(\Psi_1)\|f\|_2^2 + t^{-2} C(\Psi_2)\|g\|_2^2\right).
$$

Taking infimum over all $t > 0$ and recalling that $2ab = \inf_{t>0}(t^2 a + t^{-2} b)$ for $a, b \ge 0$ we obtain

$$
|\langle \mathbb{S}(fw^{1/2}), gv^{1/2}\rangle| \le (C(\Psi_1)C(\Psi_2))^{1/2}\|f\|_2\|g\|_2,
$$

which is exactly statement 1 of Theorem 2.3.5. □For the statement 3 of Theorem 2.3.5 we also need another embedding theorem.

Theorem 2.3.7. *Let Ψ be as above. Then for any normalized Carleson sequence $\{a_I\}_{I\in\mathcal{D}}$ ($a_I \geq 0$), i.e. for any sequence satisfying*

$$\sup_{I\in\mathcal{D}} |I|^{-1} \sum_{I'\in\mathcal{D}: I'\subset I} a_{I'}|I'| \leq 1$$

we get

$$\sum_{I\in\mathcal{D}} \frac{\langle fw^{1/2}\rangle_I^2}{\mathbf{n}_\Psi(N_I^w)} a_I |I| \leq C\|f\|^2_{L^2(dx)},$$

where again $C = C(\Psi)$.

Let us show that this theorem together with Theorem 2.3.6 implies statement 3 of Theorem 2.3.5. Let $a_I = \|\Delta_I b\|^2_\infty$.

Again, multiplying if necessary the weights v and w by appropriate constants we can assume (2.31). Then we can write

$$\begin{aligned}
|\langle \Pi_b(fw^{1/2}), gv^{1/2}\rangle| &\leq \sum_{I\in\mathcal{D}} |\langle fw^{1/2}\rangle_I| \cdot |\langle \Delta_I b, \Delta_I(gv^{1/2})\rangle| \\
&\leq \sum_{I\in\mathcal{D}} \frac{|\langle fw^{1/2}\rangle_I|(a_I)^{1/2}|I|^{1/2}}{\left(\mathbf{n}_{\Psi_1}(N_I^w)\right)^{1/2}} \cdot \frac{\|\Delta_I(gv^{1/2})\|_1}{\left(\mathbf{n}_{\Psi_2}(N_I^v)\right)^{1/2}|I|^{1/2}} \\
&\leq \left(\sum_{I\in\mathcal{D}} \frac{|\langle fw^{1/2}\rangle_I|^2 a_I}{\mathbf{n}_{\Psi_1}(N_I^w)}|I|\right)^{1/2} \left(\sum_{I\in\mathcal{D}} \frac{\|\Delta_I(gv^{1/2})\|_1^2}{\mathbf{n}_{\Psi_2}(N_I^v)|I|};\right)^{1/2}
\end{aligned}$$

the second inequality holds because of (2.31), and the last one is just the Cauchy–Schwarz inequality.

Estimating the sums in parentheses by Theorem 2.3.7 and 2.3.6 respectively we get statement 3 of Theorem 2.3.5. $\square$

2.3.6 Proof of (the Differential Embedding) Theorem 2.3.6: Bellman function and main differential inequality

Let $\varphi(s) := s\Psi(s)$. Multiplying Ψ by an appropriate constant we can assume without loss of generality that

$$\int_0^1 \frac{1}{\varphi(s)} ds = 1. \tag{2.32}$$

Define $m(s)$ on $[0,1]$ by $m(0) = m'(0) = 0$, $m''(s) = 1/\varphi(s)$. Identity (2.32) implies that m is well-defined and $0 \le m'(s) \le 1$, $0 \le m(s) \le s$. For a distribution function $N = N_I^w$ define

$$\mathbf{u}(N) = \int_0^\infty (2N(t) - m(N(t)))dt = 2\langle w \rangle_I - \int_0^\infty m(N(t))dt; \tag{2.33}$$

Note that the inequality $m(s) \le s$ implies that $\mathbf{u}(N_I^w) \ge \langle w \rangle_I$.

The functional $\mathbf{u}$ is defined on the convex set of distribution functions, i.e. on the set of decreasing functions $N : [0, \infty) \to [0, 1]$ such that $\int_0^\infty N(t)dt < \infty$.

In what follows we can consider only finitely supported functions N, and then use standard approximation reasoning. Consider two distribution functions N and N_1 and let $\Delta N = N_1 - N$. Denote also

$$\mathbf{w} := \int_0^\infty N(t)dt, \qquad \mathbf{w}_1 := \int_0^\infty N_1(t)dt,$$

and let

$$\Delta\mathbf{w} := \mathbf{w}_1 - \mathbf{w} = \int_0^\infty \Delta N(t)dt;$$

the motivation for this notation is that if N and N_1 are the distribution functions of the weights w and w_1, then the integrals are the averages on the corresponding weights. Denote also

$$\mathbf{w}_\Delta := \int_0^\infty |\Delta N(t)| dt; \tag{2.34}$$

clearly $|\Delta \mathbf{w}| \leq \mathbf{w}_\Delta$.

Let us compute derivatives of $\mathbf{u}$ in the direction of ΔN. The first derivative is given by

$$\mathbf{u}'_{\Delta N}(N) = \frac{d}{d\tau}\mathbf{u}(N + \tau\Delta N)\Big|_{\tau=0} = \int_0^\infty \big(2 - m'(N(t))\big)\, \Delta N(t) dt,$$

so, in particular

$$|\mathbf{u}'_{\Delta N}| \leq 2\mathbf{w}_\Delta.$$

Therefore we can write

$$\mathbf{u}'_{\Delta N} = \kappa \mathbf{w}_\Delta, \qquad \kappa = \kappa(\Delta N),\ |\kappa| \leq 2. \tag{2.35}$$

The second derivative in the direction $\Delta N = N_1 - N$ is given by

$$-\mathbf{u}''_{\Delta N}(N) = -\frac{d^2}{d\tau^2}\mathbf{u}(N + \tau\Delta N)\Big|_{\tau=0} = \int_0^\infty \varphi(N(t))^{-1}(\Delta N(t))^2\, dt$$

By Cauchy-Schwarz, the integral in the right side is at least

$$\Big[\int_0^\infty N(t)\Psi(N(t))\,dt\Big]^{-1}\Big[\int_0^\infty |\Delta N(t)|\,dt\Big]^2 \;= \mathbf{n}(N)^{-1}\Big[\int_0^\infty |\Delta N(t)|\,dt\Big]^2$$
$$= \mathbf{n}(N)^{-1}(\mathbf{w}_\Delta)^2,$$

so

$$-\mathbf{u}''_{\Delta N}(N) \geq \frac{(\mathbf{w}_\Delta)^2}{\mathbf{n}(N)} \tag{2.36}$$

For the scalar variable $f \in \mathbb{R}$ and the distribution function N define the Bellman function $\widetilde{\mathcal{B}}(f, N) = \mathcal{B}(\mathbf{f}, \mathbf{u}(N))$ where

$$\mathcal{B}(\mathbf{f}, \mathbf{u}) = \frac{\mathbf{f}^2}{\mathbf{u}}.$$

Computing second derivative of $\widetilde{\mathcal{B}}$ in the direction $\Delta = (\Delta\mathbf{f}, \Delta N)$ we get

$$\widetilde{\mathcal{B}}''_\Delta = \begin{pmatrix} \Delta\mathbf{f} \\ \mathbf{u}'_{\Delta N} \end{pmatrix}^T \begin{pmatrix} \mathcal{B}_{\mathbf{ff}} & \mathcal{B}_{\mathbf{fu}} \\ \mathcal{B}_{\mathbf{fu}} & \mathcal{B}_{\mathbf{uu}} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{f} \\ \mathbf{u}'_{\Delta N} \end{pmatrix} + \mathcal{B}_{\mathbf{u}}\mathbf{u}''_{\Delta N}$$

In the last formula the derivative of $\widetilde{\mathcal{B}}$ is evaluated at the point (f, N), and derivatives of $\mathcal{B}$ are evaluated at $(\mathbf{f}, \mathbf{u}(N))$.

The Hessian is easy to compute

$$\begin{pmatrix} \mathcal{B}_{\mathbf{ff}} & \mathcal{B}_{\mathbf{fu}} \\ \mathcal{B}_{\mathbf{fu}} & \mathcal{B}_{\mathbf{uu}} \end{pmatrix} = \begin{pmatrix} \frac{2}{\mathbf{u}} & -\frac{2\mathbf{f}}{\mathbf{u}^2} \\ -\frac{2\mathbf{f}}{\mathbf{u}^2} & \frac{2\mathbf{f}^2}{\mathbf{u}^3} \end{pmatrix}; \tag{2.37}$$

note that this matrix is positive semidefinite.

Since $\mathcal{B}_{\mathbf{u}} = -\mathbf{f}^2/\mathbf{u}^2$, we get using (2.36)

$$\mathcal{B}_{\mathbf{u}}\mathbf{u}''_{\Delta N} \geq \frac{\mathbf{f}^2}{\mathbf{u}^2\mathbf{n}}(\mathbf{w}_\Delta)^2.$$

Thus, gathering everything and using (2.35) we get

$$\widetilde{\mathcal{B}}''_\Delta \geq \begin{pmatrix} \Delta\mathbf{f} \\ \kappa\mathbf{w}_\Delta \end{pmatrix}^T \begin{pmatrix} \frac{2}{\mathbf{u}} & -\frac{2\mathbf{f}}{\mathbf{u}^2} \\ -\frac{2\mathbf{f}}{\mathbf{u}^2} & \frac{2\mathbf{f}^2}{\mathbf{u}^3}(1+\frac{\mathbf{u}}{2\kappa^2\mathbf{n}}) \end{pmatrix} \begin{pmatrix} \Delta\mathbf{f} \\ \kappa\mathbf{w}_\Delta \end{pmatrix} \tag{2.38}$$

The matrix here is obtained from the Hessian in (2.37) by multiplying the lower right entry by $1 + \frac{\mathbf{u}}{2\kappa^2\mathbf{n}} \geq 1$, so it has more positivity than the Hessian. In particular, if we divide the upper left entry of the matrix in (2.38) by the same quantity $1 + \frac{\mathbf{u}}{2\kappa^2\mathbf{n}}$, the matrix still be positive semidefinite. But our matrix in (2.38) has something bigger in the upper-left corner!

Therefore, since

$$1 - \Big(1 + \frac{\mathbf{u}}{2\kappa^2\mathbf{n}}\Big)^{-1} = \frac{\mathbf{u}}{2\kappa^2\mathbf{n} + \mathbf{u}}$$

we get that

$$\widetilde{\mathcal{B}}''_\Delta \geq \frac{2(\Delta\mathbf{f})^2}{2\kappa^2\mathbf{n} + \mathbf{u}} \geq \frac{2(\Delta\mathbf{f})^2}{2\cdot 2^2\mathbf{n} + \mathbf{u}} \geq c\frac{(\Delta\mathbf{f})^2}{\mathbf{n}}; \tag{2.39}$$

the last inequality holds for some $c > 0$ because $\mathbf{u} \leq 2\mathbf{w} \leq C\mathbf{n}$.

Let us explain it. In fact, we want

$$\int N_I(t)dt = \langle w\rangle_I \leq C\int N_I(t)\Psi(N_I(t))dt.$$

Clearly, it is enough to consider the set $B = \{t\colon \Psi(N_I(t)) \leq 1\}$. Since Ψ is decreasing, for

$t \in B$ we get that $N_I(t) \geq \Psi^{-1}(1)$. Since $s \mapsto s\Psi(s)$ is increasing, we get $N_I(t)\Psi(N_I(t)) \geq \Psi^{-1}(1) \geq \Psi^{-1}(1)N_I(t)$ (the last is because N_I is normalized). We are done.

Inequality (2.39) is exactly what we will use to obtain the Main inequality in difference form in the next section.

2.3.7 Main inequality in the finite difference form

2.3.7.1 Dyadic case

Lemma 2.3.8. *Let*

$$\mathbf{f} = \frac{\mathbf{f}_1 + \mathbf{f}_2}{2}, \qquad N(t) = \frac{N_1(t) + N_2(t)}{2}.$$

Then

$$\frac{1}{2}\Big(\mathcal{B}(\mathbf{f}_1, \mathbf{u}(N_1)) + \mathcal{B}(\mathbf{f}_2, \mathbf{u}(N_2))\Big) - \mathcal{B}(\mathbf{f}, \mathbf{u}(N)) \geq \frac{c}{4} \cdot \frac{(\mathbf{f}_1 - \mathbf{f})^2}{\mathbf{n}(N)}. \tag{2.40}$$

where c is the constant from (2.39). *(Note that $\mathbf{f}_1 - \mathbf{f} = \mathbf{f} - \mathbf{f}_2$, so we can replace $(\mathbf{f}_1 - \mathbf{f})^2$ in the right side by $(\mathbf{f}_2 - \mathbf{f})^2$)*

Proof. Notice that

$$\frac{s_1 + s_2}{2}\Psi\left(\frac{s_1 + s_2}{2}\right) \geq \frac{s_1 + s_2}{2}\Psi\left(s_1 + s_2\right) \geq \frac{1}{2}s_1\Psi(s_1); \tag{2.41}$$

here the first inequality holds because Ψ is decreasing and the second one because $s\Psi(s)$ is increasing. Of course, we can interchange s_1 and s_2 in the above inequality.

Let $\Delta\mathbf{f} := \mathbf{f}_1 - \mathbf{f}$, $\Delta N := N_1 - N$. Define

$$F(\tau) = \mathcal{B}(\mathbf{f} + \tau\Delta\mathbf{f}, \mathbf{u}(N + \tau\Delta N)) + \mathcal{B}(\mathbf{f} - \tau\Delta\mathbf{f}, \mathbf{u}(N - \tau\Delta N))$$

Taylor's formula together with the estimate (2.39) imply that

$$F(1) - F(0) \geq \frac{c}{2}(\Delta\mathbf{f})^2 \left(\frac{1}{\mathbf{n}(N + \tau\Delta N)} + \frac{1}{\mathbf{n}(N - \tau\Delta N)}\right) \tag{2.42}$$

for some $\tau \in (0, 1)$.

Estimate (2.41) implies that

$$\mathbf{n}(N) \geq \frac{1}{2}\mathbf{n}(N \pm \tau\Delta N),$$

so

$$\left(\frac{1}{\mathbf{n}(N + \tau\Delta N)} + \frac{1}{\mathbf{n}(N - \tau\Delta N)}\right) \geq \frac{1}{\mathbf{n}(N)}.$$

Then it follows from (2.42) that

$$F(1) - F(0) \geq \frac{c}{2} \cdot \frac{(\Delta\mathbf{f})^2}{\mathbf{n}(N)}.$$

Recalling the definition of F and dividing this inequality by 2 we get (2.40). □

2.3.7.2 General case

Let φ and $\widetilde{\mathcal{B}}$ be as above.

Lemma 2.3.9. *Let* $\mathbf{f}, \mathbf{f}_k \in \mathbb{R}$, $\alpha_k \in \mathbb{R}_+$ *and the distribution functions* N, N_k, $k = 1, 2, \ldots, n$

satisfy

$$\mathbf{f} = \sum_{k=1}^{n} \alpha_k \mathbf{f}_k, \qquad N = \sum_{k=1}^{n} \alpha_k N_k, \qquad \sum_{k=1}^{n} \alpha_k = 1.$$

Then

$$-\widetilde{\mathcal{B}}(\mathbf{f}, N) + \sum_{k=1}^{n} \alpha_k \widetilde{\mathcal{B}}(\mathbf{f}_k, N_k) \geq \frac{c}{16} \cdot \frac{1}{\mathbf{n}(N)} \left(\sum_{k=1}^{n} \alpha_k |\mathbf{f}_k - \mathbf{f}| \right)^2 \tag{2.43}$$

2.3.8 Proof of (the Embedding) Theorem 2.3.7.

2.3.8.1 An auxiliary function

Let Ψ be the function from Theorem 2.3.7. Define $\varphi(s) := s\Psi(s)$.

For the numbers $A \in [1, 2]$, $N \in \mathbb{R}_+$ define

$$T(A, N) := N \int_0^{N/A} \frac{1}{\varphi(s)} ds$$

Lemma 2.3.10. *The function T is convex and satisfies the differential inequality*

$$-\frac{\partial T}{\partial A} \geq \frac{1}{4} \cdot \frac{N^2}{\varphi(N)}.$$

Proof. Differentiating the integral we get

$$-\frac{\partial T}{\partial A} = \frac{N^2}{A^2 \varphi(N/A)} \geq \frac{1}{4} \cdot \frac{N^2}{\varphi(N)}. \tag{2.44}$$

since φ is increasing and $1 \leq A \leq 2$.

To prove the convexity notice that T is linear on the lines $N = kA$, so the Hessian d^2T

degenerates.

Differentiating (2.44) we get

$$\frac{\partial^2 T}{\partial A^2} = N^2 \frac{2A\varphi(N/A) - N\varphi'(N/A)}{(A^2\varphi(N/A))^2}$$

Note that the right side is positive if $s\varphi'(s) < 2\varphi(s)$ (because $\varphi(s) > 0$).

But for our function even a stronger inequality $s\varphi'(s) \le \varphi(s)$ is satisfied! Indeed, since $\varphi(s) = s\Psi(s)$ is increasing and Ψ is decreasing, then

$$0 \le (s\Psi(s))' = \Psi(s) + s\Psi'(s) \le \Psi(s)$$

(the second inequality holds because Ψ is decreasing). Multiplying this inequality by s we get $s\varphi'(s) \le \varphi(s)$.

Therefore, since $\varphi(s) > 0$, we conclude that $\frac{\partial^2 T}{\partial A^2} > 0$.

But the Hessian d^2T is singular, and it is an easy exercise in linear algebra to show that a singular Hermitian 2×2 matrix with a positive entry on the main diagonal is positive semidefinite. □

2.3.8.2 Bellman function and the main differential inequality.

Let now N be a distribution function, and let

$$\mathbf{T}(A, N) = \int_0^\infty T(A, N(t))\, dt.$$

As in Section 2.3.6 assume, multiplying Ψ by an appropriate constant, that

$$\int_0^1 \frac{1}{\varphi(s)} ds = 1.$$

Then $T(A, N(t)) \leq N(t)$, so

$$\mathbf{T}(A, N) \leq \int_0^\infty N(t)dt =: \mathbf{w} = \mathbf{w}(N).$$

For $\mathbf{f} \in \mathbb{R}$, $M \in [0,1]$ and for a distribution function N define the function $\widetilde{\mathcal{B}}(\mathbf{f}, N, M) := \mathcal{B}(\mathbf{f}, \mathbf{u}(M, N))$, where

$$\mathcal{B}(\mathbf{f}, \mathbf{u}) = \frac{\mathbf{f}^2}{\mathbf{u}}$$

and

$$\mathbf{u} = \mathbf{u}(M, N) = 2\int_0^\infty N(t)dt - \mathbf{T}(M+1, N)$$

$$=: 2\mathbf{w}(N) - \mathbf{T}(M+1, N).$$

Note that $2\mathbf{w}(N) \geq \mathbf{u}(M, N) \geq \mathbf{w}(N)$.

We claim that the function $\widetilde{\mathcal{B}}$ is convex. Indeed, fix a direction $\Delta := (\Delta\mathbf{f}, \Delta N, \Delta M)^T$ and compute the second derivative $\widetilde{\mathcal{B}}''_\Delta$ in this direction

$$\widetilde{\mathcal{B}}''_\Delta = \frac{d^2}{d\tau^2} \widetilde{\mathcal{B}}(\mathbf{f} + \tau\Delta\mathbf{f}, N + \tau\Delta N, M + \tau\Delta M) \Big|_{\tau=0}.$$

We get

$$\widetilde{\mathcal{B}}''_\Delta = \begin{pmatrix} \Delta\mathbf{f} \\ \mathbf{u}'_\Delta \end{pmatrix}^T \begin{pmatrix} \mathcal{B}_{\mathbf{ff}} & \mathcal{B}_{\mathbf{fu}} \\ \mathcal{B}_{\mathbf{fu}} & \mathcal{B}_{\mathbf{uu}} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{f} \\ \mathbf{u}'_\Delta \end{pmatrix} + \mathcal{B}_{\mathbf{u}}\mathbf{u}''_\Delta.$$

The Hessian

$$\begin{pmatrix} \mathcal{B}_{\mathbf{ff}} & \mathcal{B}_{\mathbf{fu}} \\ \mathcal{B}_{\mathbf{fu}} & \mathcal{B}_{\mathbf{uu}} \end{pmatrix} = \begin{pmatrix} \frac{2}{\mathbf{u}} & -\frac{2\mathbf{f}}{\mathbf{u}^2} \\ -\frac{2\mathbf{f}}{\mathbf{u}^2} & \frac{2\mathbf{f}^2}{\mathbf{u}^3} \end{pmatrix}$$

is clearly positive semidefinite, so the first term is nonnegative. For the second term notice that

$$\mathcal{B}_{\mathbf{u}} = -\frac{\mathbf{f}^2}{\mathbf{u}^2}, \qquad \mathbf{u}''_\Delta = -\mathbf{T}''_\Delta \le 0 \tag{2.45}$$

because T, and therefore $\mathbf{T}$ is convex. Thus $\widetilde{\mathcal{B}}''_\Delta \ge 0$, so $\widetilde{\mathcal{B}}$ is convex. Let us compute the partial derivative

$$-\frac{\partial\widetilde{\mathcal{B}}}{\partial M} = -\mathcal{B}_{\mathbf{u}}\frac{\partial\mathbf{u}}{\partial M} = \frac{\mathbf{f}^2}{\mathbf{u}^2}\cdot\left(-\frac{\partial\mathbf{T}}{\partial M}\right) \tag{2.46}$$

By Lemma 2.3.10

$$\begin{aligned} -\frac{\partial\mathbf{T}}{\partial M} &\ge \frac{1}{4}\cdot\int_0^\infty \frac{N(t)^2}{\varphi(N(t))}dt \\ &\ge \frac{1}{4}\left(\int_0^\infty N(t)dt\right)^2\left(\int_0^\infty \varphi(N(t))dt\right)^{-1} = \frac{1}{4}\cdot\frac{\mathbf{w}(N)^2}{\mathbf{n}(N)}; \end{aligned}$$

the second inequality here is just the Cauchy–Schwarz inequality. Combining with (2.46) and recalling that $\mathbf{u} \leq 2\mathbf{w}$ we get

$$-\frac{\partial \widetilde{\mathcal{B}}}{\partial M} \geq \frac{1}{16} \cdot \frac{\mathbf{f}^2}{\mathbf{n}} \tag{2.47}$$

This inequality (together with the convexity of $\widetilde{\mathcal{B}}$) is the main differential inequality for our function.

2.3.9 Finite difference form of the main inequality

Let $X = (\mathbf{f}, N, M)$, $X_k = (\mathbf{f}_k, N_k, M_k)$, ($\mathbf{f}, \mathbf{f}_k \in \mathbb{R}$, $M, M_k \in [0,1]$, N, N_k are the distribution functions) satisfy

$$\mathbf{f} = \sum_{k=1}^{n} \alpha_k \mathbf{f}_k, \qquad N = \sum_{k=1}^{n} \alpha_k N_k, \qquad M = a + \sum_{k=1}^{n} \alpha_k M_k, \ \ a \geq 0,$$

where

$$\sum_{k=1}^{n} \alpha_k = 1, \qquad \alpha_k \geq 0.$$

Then

$$-\widetilde{\mathcal{B}}(X) + \sum_{k=1}^{n} \alpha_k \widetilde{\mathcal{B}}(X_k) \geq \frac{1}{16} \cdot \frac{a\mathbf{f}^2}{\mathbf{n}} \tag{2.48}$$

where $\mathbf{n} = \mathbf{n}(N)$.

Indeed, for $M_0 := \sum_{k=1}^{n} \alpha_k M_k$ the main inequality (2.47) implies

$$\widetilde{\mathcal{B}}(\mathbf{f}, N, M_0) - \widetilde{\mathcal{B}}(\mathbf{f}, N, M) \geq \frac{1}{16} \cdot \frac{a\mathbf{f}^2}{\mathbf{n}}.$$

The convexity of $\widetilde{\mathcal{B}}$ implies that

$$\widetilde{\mathcal{B}}(\mathbf{f}, N, M_0) \leq \sum_{k=1}^{n} \alpha_k \widetilde{\mathcal{B}}(X_k)$$

which together with the previous inequality gives us (2.48).

2.3.9.1 From main inequality (2.48) to Theorem 2.3.7.

The reasoning here is almost verbatim the same as in Section 2.3.10.

For a cube $I \in \mathcal{L}$ let us denote $\mathbf{f}_I = \langle f w^{1/2} \rangle_I$, $N_I = N_I^w$, $M_I = |I|^{-1} \sum_{I' \subset I} a_{I'}$, $\mathbf{w}_I = \langle w \rangle_I$, $\mathbf{u}_I = \mathbf{u}(M_I, N_I)$.

Fix $I^0 \in \mathcal{L}$, and let I_k be its children. Applying the inequality (2.48) with $\alpha_k = |I_k|/|I^0|$, $\mathbf{f}_k = \mathbf{f}_{I_k}$, $N_k = N^w_{I_k}$, $M_k = M_{I_k}$ we get that

$$\frac{1}{16} \cdot \frac{a_{I^0} \mathbf{f}^2_{I^0}}{\mathbf{n}(N^w_{I^0})} |I^0| \leq -|I^0| \widetilde{\mathcal{B}}(X_{I^0}) + \sum_{I \in \mathrm{ch}(I^0)} |I| \widetilde{\mathcal{B}}(X_I)$$

Writing the corresponding estimates for the children of I^0, then for their children, we get after going n generations down and using the telescoping sum in the right side

$$\begin{aligned} \frac{1}{16} \sum_{\substack{I \in \mathrm{ch}_k(I^0) \\ 0 \leq k < n}} \frac{a_I \mathbf{f}^2_I}{\mathbf{n}(N^w_I)} |I| &\leq -|I^0| \widetilde{\mathcal{B}}(X_{I^0}) + \sum_{I \in \mathrm{ch}_n(I^0)} |I| \widetilde{\mathcal{B}}(X_I) \\ &\leq \sum_{I \in \mathrm{ch}_n(I^0)} |I| \widetilde{\mathcal{B}}(X_I); \end{aligned}$$

the last inequality holds because $\widetilde{\mathcal{B}} \geq 0$.

Since

$$\widetilde{\mathcal{B}}(X_I) \le \mathbf{f}_I^2/\mathbf{u}_I \le \mathbf{f}_I^2/\mathbf{w}_I$$

(the last inequality holds because $\mathbf{u} \ge \mathbf{w}$) and by Cauchy–Schwarz

$$|\langle fw^{1/2}\rangle_I|^2 \le \langle |f|^2\rangle_I \langle w\rangle_I,$$

we conclude, exactly as in Section 2.3.10 that

$$|I|\widetilde{\mathcal{B}}(X_I) \le |I|\langle |f|^2\rangle_I = \int_I |f|^2 d\mu,$$

so

$$\frac{1}{16} \sum_{\substack{I\in \mathrm{ch}_k(I^0)\\ 0\le k<n}} \frac{a_I \mathbf{f}_I^2}{\mathbf{n}(N_I^w)}|I| \le \int_{I^0} |f|^2 d\mu.$$

Conclusion of the proof is exactly as in Section 2.3.10: we first let $n \to \infty$, and then taking the sum over $I^0 \in \mathcal{L}_{-m}$ and letting $m \to \infty$ get the desired estimate. □

Proof. The reasoning below is a "baby version" of the reasoning used to prove the main estimate (Lemma 6.1) in [T].

For a weight $\alpha = \{\alpha_k\}_{k=1}^n$, $\alpha_k \ge 0$, let $\ell^p(\alpha)$ be the weighted (finite-dimensional) ℓ^p spaces, $\|x\|_{\ell^p(\alpha)}^p = \sum_{k=1}^n \alpha_k |x_k|^p$ ($\ell^\infty(\alpha)$ is just the usual finite-dimensional ℓ^∞).

Let $\langle\,\cdot\,,\,\cdot\,\rangle_\alpha$ be the standard duality $\langle x, y\rangle_\alpha = \sum_{k=1}^n \alpha_k x_k y_k$.

Define $\mathbf{e} \in \ell^p(\alpha)$, $\mathbf{e} = (1, 1, \dots, 1)$.

Consider the quotient space $\mathcal{X} = \ell^1(\alpha)/\operatorname{span}\{\mathbf{e}\}$. For $x \in \ell^1(\alpha)$ let

$$x^0 := x - \|\mathbf{e}\|_{\ell^1(\alpha)}^{-1} \langle x, \mathbf{e}\rangle_\alpha \mathbf{e},$$

so $\sum_{k=1}^n \alpha_k x_k^0 = 0$. Then

$$\|x\|_{\mathcal{X}} \le \|x^0\|_{\ell^1(\alpha)} \le 2\|x\|_{\mathcal{X}}. \tag{2.49}$$

Indeed, the first inequality is trivial (follows from the definition of the norm in the quotient space). As for the second one, $|\langle x, \mathbf{e}\rangle_\alpha| \le \|x\|_{\ell^1(\alpha)}$, so it follows from the triangle inequality that

$$\|x^0\|_{\ell^1(\alpha)} \le \|x\|_{\ell^1(\alpha)} + \|\mathbf{e}\|_{\ell^1(\alpha)}^{-1} |\langle x, \mathbf{e}\rangle| \cdot \|\mathbf{e}\|_{\ell^1(\alpha)} \le 2\|x\|_{\ell^1(\alpha)}.$$

This inequality remains true if one replaces x by $x - \lambda\mathbf{e}$, $\lambda \in \mathbb{R}$, so the second inequality in (2.49) is proved.

The dual space $\mathcal{X}^*$ can be identified with s subspace of $\ell^\infty = \ell^\infty(\alpha)$ consisting of $x^* \in \ell^\infty(\alpha)$ such that $\langle \mathbf{e}, x^*\rangle_\alpha = 0$ (with the usual ℓ^∞-norm).

So, for the vector $x = (x_1, x_2, \dots x_n)$, $x_k = \mathbf{f}_k - \mathbf{f}$ (notice that $\langle x, \mathbf{e}\rangle_\alpha = 0$ there is $\beta = \{\beta_k\}_{k=1}^n$, $|\beta_k| \le 1$ such that $\sum_{k=1}^n \alpha_k \beta_k = 0$ and

$$\sum_{k=1}^n \alpha_k \beta_k (\mathbf{f}_k - \mathbf{f}) = \|x\|_{\mathcal{X}} \ge \frac{1}{2}\|x\|_{\ell^1(\alpha)} = \frac{1}{2}\sum_{k=1}^n \alpha_k |\mathbf{f}_k - \mathbf{f}|.$$

Define $\mathbf{f}^+$, $\mathbf{f}^-$, N^+, N^- by

$$\mathbf{f}^\pm = \sum_{k=1}^n \alpha_k (1 \pm \beta_k)\mathbf{f}_k, \qquad N^\pm := \sum_{k=1}^n \alpha_k (1 \pm \beta_k) N_k.$$

By Lemma 2.3.8

$$\frac{1}{2}\Big(\widetilde{\mathcal{B}}(\mathbf{f}^+, N^+)) + \widetilde{\mathcal{B}}(\mathbf{f}^-, N^-))\Big) - \widetilde{\mathcal{B}}(\mathbf{f}, N) \geq \frac{c}{4} \cdot \frac{(\mathbf{f}^+ - \mathbf{f})^2}{\mathbf{n}(N)} \tag{2.50}$$

We know that

$$\mathbf{f}^+ - \mathbf{f} = \sum_{k=1}^{n} \alpha_k \beta_k \mathbf{f}_k = \sum_{k=1}^{n} \alpha_k \beta_k (\mathbf{f}_k - \mathbf{f}) \geq \frac{1}{2} \sum_{k=1}^{n} \alpha_k |\mathbf{f}_k - \mathbf{f}|$$

(the second equality holds because $\sum_{k=1}^{n} \alpha_k \beta_k = 0$), so the right side of (2.50) is estimated below by

$$\frac{c}{16} \cdot \frac{1}{\mathbf{n}(N)} \left(\sum_{k=1}^{n} \alpha_k |\mathbf{f}_k - \mathbf{f}| \right)^2$$

Since the function $\widetilde{\mathcal{B}}$ is convex

$$\widetilde{\mathcal{B}}(\mathbf{f}^+, N^+) \leq \sum_{k=1}^{n} \alpha_k (1 + \beta_k) \widetilde{\mathcal{B}}(\mathbf{f}_k, N_k),$$
$$\widetilde{\mathcal{B}}(\mathbf{f}^-, N^-) \leq \sum_{k=1}^{n} \alpha_k (1 - \beta_k) \widetilde{\mathcal{B}}(\mathbf{f}_k, N_k)$$

and adding these inequalities we can estimate above the left side of (2.50) by

$$-\widetilde{\mathcal{B}}(\mathbf{f}, N) + \sum_{k=1}^{n} \alpha_k \widetilde{\mathcal{B}}(\mathbf{f}_k, N_k).$$

□

2.3.10 From main inequality (2.43) to Theorem 2.3.6.

Fix an interval I^0 and let I_k be its children. Applying Lemma 2.3.9 with $\mathbf{f}_k = \langle f w^{1/2}\rangle_{I_k}$, $N_k = N^w_{I_k}$ and $\alpha_k = |I_k|/|I^0|$ we get denoting $\tilde f := f w^{1/2}$

$$\frac{c}{16}\cdot\frac{\|\Delta_{I^0}\tilde f\|_1^2}{\mathbf{n}(N^w_{I^0})\cdot|I^0|} \le -|I^0|\widetilde{\mathcal{B}}(\langle\tilde f\rangle_{I^0}, N^w_{I^0}) + \sum_{I\in\mathrm{ch}(I^0)} |I|\cdot\widetilde{\mathcal{B}}(\langle\tilde f\rangle_I, N^w_I)$$

Applying this formula to all children of I^0, then to their children and adding up the inequalities we get after going n generations down that

$$\frac{c}{16}\sum_{\substack{I\in\mathrm{ch}_k(I^0)\\0\le k<n}}\frac{\|\Delta_I\tilde f\|_1^2}{\mathbf{n}(N^w_I)\cdot|I|} \le -|I^0|\widetilde{\mathcal{B}}(\langle\tilde f\rangle_{I^0}, N^w_{I^0}) + \sum_{I\in\mathrm{ch}_n(I^0)}|I|\cdot\widetilde{\mathcal{B}}(\langle\tilde f\rangle_I, N^w_I)$$

$$\le \sum_{I\in\mathrm{ch}_n(I^0)}|I|\cdot\widetilde{\mathcal{B}}(\langle\tilde f\rangle_I, N^w_I).$$

We know that $\widetilde{\mathcal{B}}(\mathbf{f}, N) \le C\frac{\mathbf{f}^2}{\mathbf{u}(N)}$, and since (see (2.33)) $\mathbf{u}(N^w_I) \ge \langle w\rangle_I$ we conclude using the Cauchy–Schwarz estimate $|\langle f w^{1/2}\rangle_I|^2 \le \langle |f|^2\rangle_I\langle w\rangle_I$ that

$$|I|\cdot\widetilde{\mathcal{B}}(\langle\tilde f\rangle_I, N^w_I) \le C|I|\frac{\langle f w^{1/2}\rangle_I^2}{\langle w\rangle_I} = C|I|\langle|f|^2\rangle_I = C\int_I |f|^2 d\mu.$$

Therefore, estimating the right side we get

$$\frac{c}{16}\sum_{\substack{I\in\mathrm{ch}_k(I^0)\\0\le k<n}}\frac{\|\Delta_I\tilde f\|_1^2}{\mathbf{n}(N^w_I)\cdot|I|} \le C\int_{I^0}|f|^2 d\mu.$$

Since the right side does not depend on n we can make $n \to \infty$, and have the sum in the left side over all $I \in \mathcal{L}$, $I \subset I^0$.

Taking the sum over all $I^0 \in \mathcal{L}_{-m}$ and letting $m \to \infty$ we get conclusion of the theorem.

$\square$

2.4 The Bellman function approach to the separated bump conjecture

2.4.1 A quick reminder

We again work with Young functions Φ that satisfy

$$\Phi \text{ is convex increasing function such that } \int_1^\infty \frac{dt}{\Phi(t)} < \infty\,. \tag{2.51}$$

As we have seen, in the separated bump conjecture we work with the following quite natural one-sided bump assumption:

There exists a constant C, such that for any interval I the following holds:

$$\begin{aligned} &\|u\|_{L^\Phi_I} \cdot \|v\|_{L^1(I, \frac{dx}{|I|})} \leqslant C, \\ &\text{and} \\ &\|u\|_{L^1(I, \frac{dx}{|I|})} \cdot \|v\|_{L^\Phi_I} \leqslant C. \end{aligned} \tag{2.52}$$

And now one-sided bump conjecture is the following statement: suppose (2.52) holds for all intervals (cubes), and suppose Φ satisfies integrability condition (2.51), then any

Calderón–Zygmund operator is bounded from $L^2(u)$ into $L^2(v)$ in the sense

$$\forall f \in C_0^\infty \quad \|T(fu)\|_{L^2(v)} \leqslant C\|f\|_{L^2(u)}. \tag{2.53}$$

2.4.2 A construction of function Ψ

To formulate the main result we use a certain language.

For that we need the following construction. Define a function Ψ in the following parametric way:

$$\begin{cases} s = \frac{1}{\Phi(t)\Phi'(t)} \\ \Psi(s) := \Phi'(t). \end{cases}$$

Of course, we define Ψ in this way near $s = 0$.

We give the following definition.

Definition 19. A function Φ is called *regular bump*, if for any function u there holds

$$\|u\|_{L_I^\Phi} \geqslant C \int N_I(t)\Psi(N_I(t))dt.$$

Remark 7. An example of regular bump is the following: $\Phi(t) = t\rho(t)$, and

$$t\frac{\rho'(t)}{\rho(t)} \log \rho(t) \to 0, \text{ as } t \to \infty.$$

The important result is the following.

Lemma 2.4.1. *The function $s \mapsto \Psi(s)$ is decreasing; the function $s \mapsto s\Psi(s)$ is increasing; the function $\frac{1}{s\Psi(s)}$ is integrable near 0. Moreover, the following inequality is true with a* *uniform constant C (which may depend only on Φ):*

$$C\|u\|_{L_I^\Phi} \geqslant \int N_I(t)\Psi(N_I(t))dt,$$

where

$$N_I(t) = \frac{1}{|I|}|\{x \in I : u(x) \geqslant t\}|.$$

Further, for "regular" functions Φ we have that

$$\|u\|_{L_I^\Phi} \sim \int N_I(t)\Psi(N_I(t))dt.$$

2.4.3 The main results. Boundedness and weak boundedness.

Given a function Φ, satisfying (2.51), build the corresponding function Ψ as in Section 2.4.2. We prove the following theorems. Regularity conditions are not very important, but the last condition in the statement of the theorem is actually an important restriction. This is the restriction one would wish to get rid of. Or to prove that it is actually needed. Lately we believe that one cannot get rid of it. We give a non-standard definition.

Definition 20. A function f is "weakly concave" on its domain, if for any numbers $x_1, \ldots, x_n$ and $\lambda_1, \ldots, \lambda_n$, such that $0 \leqslant \lambda_j \leqslant 1$, and $\sum \lambda_j = 1$, the following inequality holds:

$$f(\sum \lambda_j x_j) \geqslant C \sum \lambda_j f(x_j),$$

where the constant C does not depend on n.

Theorem 2.4.2. *Suppose there exists a function Φ_0 with corresponding Ψ_0, such that:*

- Φ_0 *satisfies* (2.51);
- Φ *and* Φ_0 *are regular bumps;*
- *There is a function* ε*, such that* $\Psi_0(s) \leqslant C\Psi(s)\varepsilon(\Psi(s))$*;*
- *The function* $t \mapsto t\varepsilon(t)$ *is weakly concave, in the sense of the Definition 20;*
- *The function* $t \mapsto t\varepsilon(t)$ *is strictly increasing near* ∞*;*
- *The function* $t \mapsto t\varepsilon(t)$ *is concave near* ∞*;*
- *The function* $t \mapsto \frac{\varepsilon(t)}{t}$ *is integrable at* ∞*.*

Suppose that there exists a constant C*, such that a one-sided bump condition* (2.52) *holds. Then any Calderón–Zygmund operator is bounded from* $L^2(u)$ *into* $L^2(v)$ *in the sense of* (2.53).

Theorem 2.4.3. *Suppose the function* Φ *satisfies all conditions from the theorem above. Suppose that there exists a constant* C*, such that*

$$\|u\|_{L^1(I,\frac{dx}{|I|})} \cdot \|v\|_{L^\Phi_I} \leqslant C.$$

Then any Calderon-Zygmund operator is weakly bounded from $L^2(u)$ *into* $L^{2,\infty}(v)$*, i.e. there exists a constant* C*, such that for any function* $f \in C_0^\infty$ *there holds*

$$\|T(fu)\|_{L^{2,\infty}(v)} \leqslant C\|f\|_{L^2(u)}. \tag{2.54}$$

2.4.4 Examples of Φ satisfying the restrictions of the main results: the cases from [CURV]

The biggest difference of the above results with those of [CURV] is that here we gave the integral condition on the corresponding bump function Φ. To compare with [CURV] we notice that in [CURV] theorems above were proved in two cases:

1. $\Phi(t) = t\log^{1+\sigma}(t)$;

2. $\Phi(t) = t\log(t)\log\log^{1+\sigma}(t)$, for sufficiently big σ.

We show that these results are covered by our theorems.

First, suppose $\Phi(t) = t\log^{1+\sigma}(t)$. Then $\Psi(s) \asymp \log^{1+\sigma}(\frac{1}{s})$. We put $\Phi_0(s) = t\log^{1+\frac{\sigma}{2}}(t)$, and then $\varepsilon(t) = t^{-\frac{\sigma}{2(1+\sigma)}}$. Then, clearly, all properties of ε from our theorem are satisfied.

Next, suppose $\Phi(t) = t\log(t)\log\log^{1+\sigma}(t)$. Then $\Psi(s) \asymp \log(\frac{1}{s})\log\log^{1+\sigma}(\frac{1}{s})$. We put $\Phi_0(t) = t\log(t)\log\log^{1+\delta\sigma}(t)$, $\delta < 1$ which gives $\varepsilon(t) = \log^{-(1-\delta)\sigma}(t)$. Then, the integral $\int^{\infty}\frac{\varepsilon(t)}{t}\,dt$ converges if $\sigma > 1$, and we choose δ to be very small.

Moreover, examining the proof of Theorem 5.1 from [CURV], we get the result from our paper but with a condition

$$\text{The function } t \mapsto \frac{\sqrt{\varepsilon(t)}}{t} \text{ is integrable at } \infty.$$

We notice that for regular functions we have $\varepsilon(t) \to 0$ when $t \to \infty$, and so $\varepsilon(t) < \sqrt{\varepsilon(t)}$. Thus, our results work for more function ε and, thus, bumps Φ.

2.4.5 Self improvements of Orlicz norms.

In this section we prove a technical result, which has the following "hand-waving" explanation: suppose we take a function Φ and a smaller function Φ_0. We explain how small can be the quotient $\frac{\|u\|_{L_I^{\Phi_0}}}{\|u\|_{L_I^{\Phi}}}$ in terms of smallness of $\frac{\Phi_0}{\Phi}$. In what follows we consider only "regular bumps" functions in the sense of the Definition 19.

Suppose we have two functions Φ and Φ_0, and we have built functions Ψ and Ψ_0. We suppose that

$$\Psi_0(s) \leqslant C\Psi(s)\varepsilon(\Psi(s)).$$

The following theorem holds.

Theorem 2.4.4. *Let I be an arbitrary interval (cube). If a function $t \mapsto t\varepsilon(t)$ is weakly concave, then*

$$\|u\|_{L_I^{\Phi_0}} \leqslant C\|u\|_{L_I^{\Phi}}\, \varepsilon\Big(\frac{\|u\|_{L_I^{\Phi}}}{\langle u\rangle_I}\Big).$$

To do that we need the following easy lemma:

Lemma 2.4.5. *For weakly concave functions the Jensen inequality holds with a constant:*

$$\int f(g(t))d\mu(t) \leqslant Cf(\int g(t)d\mu(t)).$$

Proof. This is true since if g is a step function, then this is just a definition. Then we pass to the limit. Here we essentially used that we can take a convex combination of n points, and the constant in the definition above does not depend on n. □

Proof of the Theorem. In the proof we omit the index I. Since for regular bumps we know that

$$\|u\|_{L^\Phi} \sim \int \Psi(N(t))N(t)dt,$$

we simply need to prove that

$$\int \Psi_0(N(t))N(t)dt \leqslant C \int \Psi(N(t))N(t)dt \, \varepsilon\Big(\frac{\int \Psi(N(t))N(t)dt}{\int N(t)dt}\Big)$$

Our first step is the obvious estimate of the left-hand side:

$$\int \Psi_0(N(t))N(t)dt \leqslant C \int \Psi(N(t))\varepsilon(\Psi(N(t))N(t)dt \, .$$

Denote $a(t) = t\varepsilon(t)$. Then we need to prove that

$$\int a(\Psi(N(t))N(t)dt \leqslant C \int N(t)dt \; a\Big(\frac{\int \Psi(N(t))N(t)dt}{\int N(t)dt}\Big) \, .$$

We denote

$$d\mu = \frac{N(t)}{\int N(t)dt}dt,$$

it is a probability measure. Moreover, by assumption, $t \mapsto a(t)$ is concave. Therefore, by Jensen's inequality (from the Lemma),

$$\int a(f(t))d\mu(t) \leqslant Ca\Big(\int f(t)d\mu(t)\Big) \, .$$

Take $f(t) = \Psi(N(t))$, and the result follows. □

2.4.6 Examples

2.4.6.1 Example 1: log-bumps

First, if $\Phi(t) = t\log^{1+\sigma}(t)$, then $\Psi(s) = \log^{1+\sigma}(1/s)$, and

$$\frac{\Psi_0(s)}{\Psi(s)} = \log^{-\frac{\sigma}{2}}(1/s) = \Psi^{-\frac{\sigma}{2(1+\sigma)}}.$$

Thus, $\varepsilon(t) = t^{-\frac{\sigma}{2(1+\sigma)}}$, and everything is fine.

2.4.6.2 Example 2: log log-bumps

Next example is with double logs. In fact, when

$$\Psi(s) = log(1/s)(\log\log(1/s))^{1+\sigma}, \quad \Psi_0(s) = log(1/s)(\log\log(1/s))^{1+\sigma/2}$$

then

$$\frac{\Psi_0(s)}{\Psi(s)} = \log\log^{-\sigma/2}(1/s) \sim (\log(\Psi(s)))^{-\sigma/2}.$$

Thus, $\varepsilon(t) = (\log t)^{-\frac{\sigma}{2}}$. Everything would be also fine, except for one little thing: the function $t \mapsto t\varepsilon(t)$ is concave on infinity, but not near 1. However, $t \mapsto t\varepsilon(t)$ is weakly concave on $[2,\infty)$, and this is enough for our goals as without loss of generality, $\Psi(s) \geqslant 2$.

So let us prove that $a(t) = t\varepsilon(t)$ is weakly concave on $[2,\infty)$.

Let $\varkappa := \frac{\sigma}{2}$. The function a has a local minimum at $e^{\varkappa}$ and its concavity changes at $e^{\varkappa+1}$. We now take x_j, λ_j and $x = \sum \lambda_j x_j$. We first notice that if $x > e^{\varkappa+1}$, the we are done, because then $(x, \sum \lambda_j a(x_j))$ lies under the graph of a.

If $2 \leq x < e^{\varkappa+1}$, then $a(x) > \min_{[2,e^{\varkappa+1}]} a = c(\varkappa)$. Moreover, if ℓ is a line tangent to graph of a, starting at $(2, a(2))$, and ℓ "kisses" the graph at a point $(r, a(r))$, then $\sum \lambda_j a(x_j) \leqslant a(r) = c_1(\varkappa)$. This follows from the picture: a convex combination of $a(x_j)$ can not be higher than this line.

Therefore,

$$a(\sum \lambda_j x_j) \geqslant c(\varkappa) \geqslant C c_1(\varkappa) \geqslant \sum \lambda_j a(x_j)\,.$$

This finishes our proof.

2.4.7 Proof of the main result: notation and the first reduction.

We fix a dyadic grid $\mathcal{D}$. To prove our main results it is enough to show that the following implication holds:

$$\text{if for all } I \ \|u\|_{L^{\Phi}_I} \cdot \|v\|_{L^1(I,\frac{dx}{|I|})} \leqslant B_{u,v} \text{ then } \|\chi_J T_{\mathcal{D},\{a_I\}}(u\chi_J)\|^2_{L^2(v)} \leqslant C\, u(J),$$

where C does not depend neither on the grid, nor on the sequense $\{a_I\}$. It can, of course, depend on $B_{u,v}$. This will prove the weak bound $T : L^2(v) \to L^{2,\infty}(u)$. For simplicity, we denote $T_a = T_{\mathcal{D},\{a_I\}}$. It is an easy calculation that, under the joint A_2 condition (which is definitely satisfied under the bump condition), it is enough to get an estimate of the following form:

$$\frac{1}{|J|} \sum_{J \subset I} a_I \cdot \langle u \rangle_I \cdot \frac{1}{|I|} \sum_{K \subset I} a_K \langle u \rangle_K \langle v \rangle_K |K| \cdot |I| \leqslant C\, u(J)\,. \tag{2.55}$$

Remark 8. By the rescaling argument it is clear that we can assume $B_{u,v}$ as small as we need (where "smallness", of course, depends only on the function Φ). We need this remark, since all behaviors of our function ε are studied near 0.

Remark 9. Everything is reduced to (2.55). We concentrate on proving (2.55). Clearly, by scale invariance, it looks very tempting to make (2.55) a Bellman function statement. This will be exactly our plan from now on.

2.4.8 Bellman proof of (2.55): introducing the "main inequality"

We start this Section with the following notation. We fix two weights u and v, and a Carleson sequense $\{a_I\}$. We denote

$$u_I = \langle u \rangle_I, \quad v_I = \langle v \rangle_I; \quad N_I(t) = \frac{1}{|I|}|\{x\colon u(x) \geqslant t\}|;$$

$$A_I = \frac{1}{|I|}\sum_{J\subset I} a_J |J|;$$

$$L_I = \frac{1}{|I|}\sum_{J\subset I} a_J \langle u \rangle_J \langle v \rangle_J |J|.$$

We proceed with two theorems that prove our main result. Everywhere in the future we use that $\langle u \rangle_I \langle v \rangle_I = u_I v_I \leqslant \delta < 1$ for any I. We can do it due to simple rescaling.

Theorem 2.4.6. *Suppose that*

$$\frac{\Psi_0}{\Psi} \leqslant \varepsilon(\Psi),$$

where ε satisfies properties of Theorem 2.9, from which the main one is

$$\int^{\infty} \frac{\varepsilon(t)}{t}\, dt < \infty\,. \tag{2.56}$$

Let δ be small enough, and

$$\Omega_1 = \{(N, A)\colon 0 \leqslant N \leqslant 1;\ 0 \leqslant A \leqslant 1\}$$

and for some constant P

$$\Omega_2 = \{(u,v,L,A)\colon 0 \leqslant A \leqslant 1;\ u,v,L \geqslant 0;\ uv \leqslant \delta;\ L \leqslant P\cdot\sqrt{uv}\}.$$

Suppose we have found a function B_1, defined on Ω_1, and a function B_2, defined on Ω_2, such that:

$$0 \leqslant B_1 \leqslant N; \tag{2.57}$$

$$(B_1)'_A \geqslant 10\frac{N}{\Psi_0(N)}; \tag{2.58}$$

$$-d^2 B_1 \geqslant 0; \tag{2.59}$$

$$0 \leqslant B_2 \leqslant u; \tag{2.60}$$

$$(B_2)'_A \geqslant 0 \tag{2.61}$$

$$(B_2)'_A \geqslant c\cdot u\cdot L,\ \textit{when}\ P\cdot\sqrt{uv} \geq L \geqslant \frac{uv}{\varepsilon(\frac{1}{uv})}; \tag{2.62}$$

$$uv(B_2)'_L \geqslant -\delta_1 uL,\ \textit{for sufficiently small}\ \delta_1\ \textit{in the whole of}\ \Omega_2; \tag{2.63}$$

$$-d^2 B_2 \geqslant 0. \tag{2.64}$$

Then for the function of an interval $\mathcal{B}(I) := B_2(u_I, v_I, L_I, A_I) + \int\limits_0^\infty B_1(N_I(t), A_I)dt$ the following holds:

$$0 \leqslant \mathcal{B}(I) \leqslant 2u_I \tag{2.65}$$

$$\mathcal{B}(I) - \frac{\mathcal{B}(I_+)+\mathcal{B}(I_-)}{2} \geqslant C\, a_I\cdot u_I\cdot L_I. \tag{2.66}$$

Next, we state

Theorem 2.4.7. *If such two functions B_1 and B_2 exist, then* (2.55) *holds, namely*

$$\frac{1}{|I|}\sum_{J\subset I} a_I \cdot \langle u\rangle_J \cdot \frac{1}{|J|}\sum_{K\subset J} a_K \langle u\rangle_K \langle v\rangle_K |K| \cdot |I| \leqslant R^2 \int_I u.$$

Proof of the Theorem 2.4.7. This is a standard Green's formula applied to function $\mathcal{B}(I)$ on the tree of dyadic intervals. Let us explain the details.

Since the function $\mathcal{B}$ is non-negative, we have that

$$2|I|u_I \geqslant |I|\mathcal{B}(I) \geqslant |I|\mathcal{B}(I) - \sum_{k=1}^{2^n} |I_{n,k}|\mathcal{B}(I_{n,k}).$$

Here n is fixed, and $I_{n,k}$ are $n-th$ generation descendants of I. Clearly, all $|I_{n,k}|$ are equal to 2^{-n}.

Let us denote $\Delta(J) = |J|\mathcal{B}(J) - |J_+|\mathcal{B}(J_+) - |J_-|\mathcal{B}(J_-)$, where $J_\pm$ are children of J. By the property (2.66) we know that $\Delta(J) \geqslant C|J|\, a_J u_J L_J$. By the telescopic cancellation, we get that

$$|I|\mathcal{B}(I) - \sum_{k=1}^{2^n} |I_{n,k}|\mathcal{B}(I_{n,k}) = \sum_{m=0}^{n-1}\sum_{k=1}^{2^m} \Delta(I_{m,k}).$$

Combining our estimates, we get

$$2|I|u_I \geqslant C\sum_{m=0}^{n-1}\sum_{k=1}^{2^m} |I_{m,k}|a_{I_{m,k}}u_{I_{m,k}}L_{I_{m,k}} = C \sum_{J\subset I, |J|\geqslant 2^{-n}|I|} |J|a_J u_J L_J.$$

This is true for every n, with the constant C independent of n. Thus,

$$u_I \geqslant C\frac{1}{|I|}\sum_{J\subset I} a_J u_J L_J |J|.$$

The result follows from the definition of L_J. □

In the future we use the following variant of Sylvester criterion of positivity of matrix.

Lemma 2.4.8. *Let $M = (m_{ij})_{i,j=1}^3$ be a 3×3 real symmetric matrix such that $m_{11} < 0$, $m_{11}m_{22} - m_{12}m_{21} > 0$, and $\det M = 0$. Then M is nonpositive definite.*

Proof. Let E be a matrix with all entries being 0 except for $e_{33} = 1$. Consider $t > 0$ and $A := A(t) := M + tE$. It is easy to see that $a_{11} < 0$, $a_{11}a_{22} - a_{12}a_{21} > 0$, and $\det A = t \cdot (m_{11}m_{22} - m_{12}m_{21}) > 0$ when $t > 0$. By Sylvester criterion, matrices $A(t)$, $t > 0$, are all negatively definite. Therefore, tending t to $0+$, we obtain, that M is nonpositive definite. □

We need the following lemma, which is in spirit of [VaVo].

Lemma 2.4.9. *Let L_I be given by*

$$L_I = \frac{1}{|I|} \sum_{J \subset I} a_J \langle u \rangle_J \langle v \rangle_J |J|.$$

Let A_I given by $A_I = \frac{1}{|I|} \sum\limits_{J \subset I} a_J |J|$. Suppose that it is bounded by 1 for any dyadic I (Carleson condition). If for any dyadic interval I we have that $\langle u \rangle_I \langle v \rangle_I \leqslant 1$, then it holds that for any dyadic interval I we have $L_I \leqslant P\sqrt{\langle u \rangle_I \langle v \rangle_I}$.

Proof. It is true since the function $T(u, v, A) = 100\sqrt{uv} - \frac{uv}{A+1}$ is concave enough in the domain $G := \{0 \leqslant A \leqslant 1,\ uv < 1,\ u, v \geqslant 0\}$. One can adapt the proof from [VaVo].

First, we need to check that the function $T(x, y, A)$ is concave in G. Clearly, $T''_{A,A} < 0$. Next,

$$\det \begin{pmatrix} T''_{A,A} & T''_{A,v} \\ T''_{A,v} & T''_{v,v} \end{pmatrix} = \frac{x}{y(A+1)^4} \cdot (50(A+1)\sqrt{xy} - xy) > 0\,. \tag{2.67}$$

This expression is non-negative, because $A+1 \geqslant 1$, and $\sqrt{uv} \leqslant 1$. Finally,

$$\det \begin{pmatrix} T''_{A,A} & T''_{A,v} & T''_{A,u} \\ T''_{A,v} & T''_{v,v} & T''_{v,u} \\ T''_{A,u} & T''_{v,u} & T''_{u,u} \end{pmatrix} = 0.$$

Therefore, by Lemma 2.4.8 we conclude that $T(u,v,A)$ is a concave function.

Next,

$$T'_A = \frac{uv}{(A+1)^2} \geqslant \frac{1}{4}uv.$$

Thus, if we fix three points (u,v,A), $(u_\pm, v_\pm, A_\pm)$, such that $u = \frac{u_+ + u_-}{2}$, $v = \frac{v_+ + v_-}{2}$, and $A = \frac{A_+ + A_-}{2} + a$, we get by the Taylor formula:

$$T(u,v,A) - \frac{T(u_+,v_+,A_+) + T(u_-,v_-,A_-)}{2} \geqslant aT'_A(u,v,A) \geqslant C\, a \cdot uv.$$

This requires the explanation. The Taylor formula we used has a remainder with the second derivative at the intermediate point $P_\pm$ on segments $S_+ := [(u, v, \frac{A_- + A_+}{2}), (u,v,A_+)]$, $S_- := [(u,v,\frac{A_- + A_+}{2}), (u,v,A_-)]$. One of this segments definitely lies inside domain G, where T is concave, and this remainder will have the right sign. However the second segment can easily stick out of domain G, because G itself is not convex. But notice that if, for example, S_+ is not inside G, still $(x,y,B) \in S_+$ implies that one of the coordinates, say x, must be smaller than u. Then y can be bigger than v, but not much. In fact,

$$v_+ - v = v - v_- \;\Rightarrow\; v_+ \leq 2v - v_- \leq 2v\,.$$

Therefore, $y \leq v_+ \leq 2v$. Then we have that $xy \leq 2uv \leq 2$. Let us consider $\widetilde{G} := \{(x,y,A) :$

$0 \le A \le 1,\ x, y \ge 0,\ 0 \le xy \le 2\}$. Now come back to the proof that T is concave in G. In (2.67) we used that if $(x, y, A) \in G$, then $xy \le 1$ and the corresponding determinant is non-negative. But the same non-negativity in (2.67) holds under slightly relaxed assumption $(x, y, A) \in \widetilde{G}$.

We notice that our $u_I = \langle u \rangle_I$, $v_I = \langle v \rangle_I$, and $A_I = \frac{1}{|I|} \sum_{J \subset I} a_I |I|$ have the dynamics above. The rest of the proof reads exactly as the proof of the Theorem 2.4.7. □

Proof of the Theorem 2.4.6. We start with the following corollary from the Taylor expansion. Suppose we have three tuples (N, A), $(N_\pm, A_\pm)$, such that:

$$N = \frac{N_+ + N_-}{2}; \quad A = \frac{A_+ + A_-}{2} + m.$$

Moreover, suppose there are (u, v, L), $(u_\pm, v_\pm, L_\pm)$, such that

$$u = \frac{u_+ + u_-}{2}; \quad v = \frac{v_+ + v_-}{2}; \quad L = \frac{L_+ + L_-}{2} + m \cdot uv.$$

Then, since $d^2 B_1 \leqslant 0$, we write

$$B_1(N_+, A_+) \leqslant B_1(N, A) + (B_1)'_N(N, A)(N_+ - N) + (B_1)'_A(N, A)(A_+ - A).$$

Thus,

$$B_1(N, A) - \frac{B_1(N_+, A_+) + B_1(N_-, A_-)}{2} \geqslant (B_1)'_A(N, A) \cdot (A - \frac{A_+ + A_-}{2}) = m \cdot (B_1)'_A(N, A) \geqslant m \frac{N}{\Psi_0(N)}.$$

Similarly,

$$B_2(u,v,L,A) - \frac{B_2(u_+,v_+,L_+,A_+) + B_2(u_-,v_-,L_-,A_-)}{2} \geqslant m \cdot ((B_2)'_A(u,v,L,A) + uv(B_2)'_L)$$

First, suppose that $L_I \leqslant \frac{u_I v_I}{\varepsilon(\frac{1}{u_I v_I})}$. Then, using $m = a_I$ we get

$$\begin{aligned}
\mathcal{B}(I) - \frac{\mathcal{B}(I_+) + \mathcal{B}(I_-)}{2} \geqslant \\
\geqslant \int \left(B_1(N_I(t), A_I) - \frac{B_1(N_{I_+}(t), A_{I_+}) + B_1(N_{I_-}(t), A_{I_-})}{2} \right) dt + \\
+ \left(B_2(u_I, v_I, L_I, A_I) - \frac{B_2(u_{I_+}, v_{I_+}, L_{I_+}, A_{I_+}) + B_2(u_{I_-}, v_{I_-}, L_{I_-}, A_{I_-})}{2} \right) \\
\geqslant a_I \left((B_2)'_A(u_I, v_I, L_I, A_I) + u_I v_I (B_2)'_L(u_I, v_I, L_I, A_I) \right) + a_I \left(\int (B_1)'_A(N_I(t), A_I) dt \right) \geqslant \\
a_I \left(\int \frac{N_I(t)}{\Psi_0(N_I(t))} dt - \delta_1 u_I L_I \right).
\end{aligned} \tag{2.68}$$

The last inequality is true, since $(B_2)'_A \geqslant 0$ and $uv(B_2)'_L \geqslant -\delta_1 uL$ on the domain of B_2. We use Hölder's inequality (and that $\int N_I(t)dt = u_I$) to get:

$$\int \frac{N_I(t)}{\Psi_0(N_I(t))} dt \geqslant \frac{u_I^2}{\int N_I(t)\Psi_0(N_I(t))dt} \geqslant C \frac{u_I^2}{\int N_I(t)\Psi(N_I(t))dt \, \varepsilon\left(\frac{\int N_I(t)\Psi(N_I(t))dt}{u_I} \right)}. \tag{2.69}$$

Last inequality is Theorem 2.4.4. Therefore, we get that

$$\int \frac{N_I(t)}{\Psi_0(N_I(t))}dt \;\geqslant\; u_I \cdot \frac{u_I}{\|u\|_{L_I^\Phi}} \cdot \frac{1}{\varepsilon\left(\frac{\|u\|_{L_I^\Phi}}{u_I}\right)} \;=\; u_I \frac{u_I v_I}{\|u\|_{L_I^\Phi} v_I} \cdot \frac{1}{\varepsilon\left(\frac{\|u\|_{L_I^\Phi} v_I}{u_I v_I}\right)}. \quad (2.70)$$

We are going to use the one-sided bump condition $\|u\|_{L_I^\Phi} v_I \leqslant B_{u,v} \leqslant 1$. Thus,

$$u_I v_I \leqslant \frac{u_I v_I}{\|u\|_{L_I^\Phi} v_I}.$$

Since the function $x \mapsto \frac{x}{\varepsilon(\frac{1}{x})}$ is increasing near 0 (on $[0, c_\varepsilon]$) and bounded from below between c_ε and 1, we get

$$\frac{u_I v_I}{v_I \|u\|_{L_I^\Phi}} \cdot \frac{1}{\varepsilon\left(\frac{v_I \|u\|_{L_I^\Phi}}{u_I v_I}\right)} \geqslant C \cdot u_I v_I \frac{1}{\varepsilon(\frac{1}{u_I v_I})}.$$

Therefore,

$$\int \frac{N_I(t)}{\Psi_0(N_I(t))}dt \geqslant C u_I \frac{u_I v_I}{\varepsilon(\frac{1}{u_I v_I})} \geqslant C u_I L_I.$$

The last inequality follows from our assumption that $L_I \leqslant \frac{u_I v_I}{\varepsilon(\frac{1}{u_I v_I})}$. Putting everything together, we get

$$\mathcal{B}(I) - \frac{\mathcal{B}(I_+) + \mathcal{B}(I_-)}{2} \geqslant a_I u_I L_I (C - \delta_1) \geqslant C_1 \cdot a_I u_I L_I.$$

We proceed to the case $L_I \geqslant \frac{u_I v_I}{\varepsilon(\frac{1}{u_I v_I})}$. Then we write

$$\mathcal{B}(I) - \frac{\mathcal{B}(I_+) + \mathcal{B}(I_-)}{2} \geqslant B_2(u_I, v_I, L_I, A_I) - \frac{B_2(u_{I_+}, v_{I_+}, L_{I_+}, A_{I_+}) + B_2(u_{I_-}, v_{I_-}, L_{I_-}, A_{I_-})}{2}.$$

This is obviously true, since $(B_1)'_A \geqslant 0$ everywhere and B_1 is a concave function. Next, we use

$$B_2(u_I, v_I, L_I, A_I) - \frac{B_2(u_{I_+}, v_{I_+}, L_{I_+}, A_{I_+}) + B_2(u_{I_-}, v_{I_-}, L_{I_-}, A_{I_-})}{2} \geqslant a_I \left((B_2)'_A + uv(B_2)'_L\right) \geqslant ca_I \cdot u_I L_I, \quad (2.71)$$

by the property of B_2. Therefore, we are done. □

2.4.9 Fourth step: building the function B_2

In order to finish the proof, we need to build functions B_1 and B_2. In this section we will present the function B_2. Denote

$$\varphi(x) = \frac{x}{\varepsilon(\frac{1}{x})}.$$

This function is increasing (by regularity assumptions on ε in Theorem 2.9), therefore, there exists φ^{-1}. We introduce

$$B_2(u, v, L, A) = Cu - \frac{L^2}{v} \int\limits_{\frac{A+1}{L}}^{\infty} \varphi^{-1}\Big(\frac{1}{x}\Big)\, dx.$$

Let us explain why the integral is convergent. In fact, using change of variables, we get

$$\int_{1}^{\infty} \varphi^{-1}\Big(\frac{1}{x}\Big)\, dx = \int_{0}^{\varphi^{-1}(1)} \frac{\varepsilon(\frac{1}{t}) - t\frac{d}{dt}(\varepsilon(\frac{1}{t}))}{t}\, dt,$$

which converges at 0 by assumption (2.56).

Therefore, since $L_I \leqslant C\sqrt{u_I v_I}$, we get

$$0 \leqslant B(u_I, v_I, L_I, A_I) \leqslant C u_I.$$

Next,

$$(B_2)'_A + uv(B_2)'_L = \frac{L}{v}\varphi^{-1}\Big(\frac{L}{A+1}\Big) - u(A+1)\varphi^{-1}\Big(\frac{L}{A+1}\Big) - 2uL\int_{\frac{A+1}{L}}^{\infty} \varphi^{-1}\Big(\frac{1}{x}\Big)\, dx =$$

$$uL \cdot \left(\frac{1}{uv}\varphi^{-1}\Big(\frac{L}{A+1}\Big) - \frac{A+1}{L}\varphi^{-1}\Big(\frac{L}{A+1}\Big) - 2\int_{\frac{A+1}{L}}^{\infty} \varphi^{-1}\Big(\frac{1}{x}\Big)\, dx \right) \quad (2.72)$$

We use that $L \geqslant \frac{uv}{\varepsilon\left(\frac{1}{uv}\right)} = \varphi(uv)$. Then $\varphi^{-1}(L) \geqslant uv$, and, since $A+1 \sim 1$, we get

$$\frac{1}{uv}\varphi^{-1}\Big(\frac{L}{A+1}\Big) \geqslant C_1.$$

Moreover, since $uv \leqslant \delta$ is a small number, we get that L is small enough for the integral $\int_{\frac{A+1}{L}}^{\infty} \varphi^{-1}(\frac{1}{x})dx$ to be less than a small number c_2. Finally, let us compare $\frac{A+1}{L}\varphi^{-1}(\frac{L}{A+1})$

with a small number c_3. Since L is small, we can write

$$\varepsilon\Big(\frac{1}{c_3 L}\Big) \leqslant c_3.$$

We do it, since c_3 is fixed from the beginning (say, $c_3 = \frac{1}{10}$). Thus,

$$L \leqslant \varphi(c_3 L).$$

This implies

$$\varphi^{-1}(L) \leqslant c_3 L,$$

thus

$$\frac{1}{L}\varphi^{-1}(L) \leqslant c_3.$$

Since $A+1 \sim 1$, we get the desired. Therefore, if $L \geqslant \frac{uv}{\varepsilon(\frac{1}{uv})} = \varphi(uv)$ then $(B_2)'_A + uv(B_2)'_L \geqslant cuL$.

Moreover, in the whole domain of B_2 we get, since $(B_2)'_A \geqslant 0$,

$$(B_2)'_A + uv(B_2)'_L \geqslant uv(B_2)'_L \geqslant -(c_2 + c_3)uL$$

with small $c_2 + c_3$. This is a penultimate inequality in the statement of Theorem 2.4.6.

Now we shall prove the concavity of B_2. For this it is enough to prove the concavity of the function of three variables: $B(v, L, A) := B_2(u, v, L, A) - Cu$. Clearly, $(B)''_{vv} < 0$, which is obvious. Also, it is a calculation that

$$\det\begin{pmatrix} (B)''_{vv} & (B)''_{vA} & (B)''_{vL} \\ (B)''_{vA} & (B)''_{AA} & (B)''_{AL} \\ (B)''_{vL} & (B)''_{AL} & (B)''_{LL} \end{pmatrix} = 0.$$

Thus, we need to consider the matrix

$$\begin{pmatrix} (B)''_{vv} & (B)''_{vA} \\ (B)''_{vA} & (B)''_{AA} \end{pmatrix}$$

and to prove that its determinant is positive. We denote $f(t) = \varphi^{-1}(t)$, to simplify the next formula. The calculation shows that the determinant above is equal to

$$g\Big(\frac{L}{A+1}\Big) := -f\Big(\frac{L}{A+1}\Big)^2 + 2\Big(\frac{L}{A+1}\Big)^2 \cdot f'\Big(\frac{L}{A+1}\Big) \int\limits_{\frac{A+1}{L}}^{\infty} f\Big(\frac{1}{x}\Big)\, dx.$$

We need to prove that g is positive near 0. First, $g(0) = 0$. Next,

$$g'(s) = -2f(s)f'(s) + 4sf'(s) \int\limits_{\frac{1}{s}}^{\infty} f\Big(\frac{1}{x}\Big)\, dx + 2s^2 f''(s) \int\limits_{\frac{1}{s}}^{\infty} f\Big(\frac{1}{x}\Big)\, dx + 2f'(s)f(s) =$$
$$4sf'(s) \int\limits_{\frac{1}{s}}^{\infty} f\Big(\frac{1}{x}\Big)\, dx + 2s^2 f''(s) \int\limits_{\frac{1}{s}}^{\infty} f\Big(\frac{1}{x}\Big)\, dx. \quad (2.73)$$

We notice that f' is positive, since φ^{-1} is increasing near 0. Moreover, by the fact that φ is strictly monotonous, and by concavity of $t\varepsilon(t)$ (see Theorem 2.4.2), we get that φ is *strictly* convex, hence φ^{-1} is strictly convex near 0 as well. That is, f'' is also positive.

Therefore, $g'(s) > 0$, and so $g(s) > g(0) = 0$. The application of Lemma 2.4.8 finishes the proof of concavity of B (and therefore of the concavity of B_2). We are done.

Remark 10. We can always think that the bump constant $B_{u,v} \leqslant C_\varepsilon$, where C_ε is such that $L_I \leqslant c_\varepsilon$. Then we can use the monotonicity and concavity of the function φ near 0.

2.4.10 Fifth step: building the function B_1

We present the function from 2.3.8.1.

$$B_1(N, A) = CN - N \int_0^{\frac{N}{A}} \frac{ds}{s\Psi_0(s)}$$

Chapter 3

One weight estimate for the limiting case: the A_1 conjecture

3.1 The main result

We are on $I_0 := [0,1]$. As always $\mathcal{D}$ denote the dyadic lattice. In this chapter we use the usual Haar system $\{h_I\}$:

$$h_I(x) := \begin{cases} \frac{1}{\sqrt{|I|}}\,, \; x \in I_+ \\ -\frac{1}{\sqrt{|I|}}\,, \; x \in I_- \end{cases}$$

The weighted weak norm of an operator T is defined by

$$\|T\|_{L^{1,\infty}(w)} = \sup_{t>0, \|f\|_{L^1(w)}=1} t \cdot w\{x \colon |Tf(x)| \geqslant t\}$$

We consider the operator

$$T_\varepsilon \colon \varphi \to \sum_{I \subseteq I_0, I \in \mathcal{D}} \varepsilon_I (\varphi, h_I) h_I\,,$$

where $\varepsilon_I = \pm 1$. Notice that the sum does not contain the constant term.

Our main theorem is the following.

Theorem 3.1.1. *For any $p < \frac{1}{5}$ and for any large Q there exists a weight w, such that*

$[w]_1 = Q,$ *and*

$$\sup_{\varepsilon=\{\varepsilon_I\}} \|T_\varepsilon\|_{L^{1,\infty}(w)} \geqslant Q \log^p Q.$$

3.2 The Bellman approach

Put

$$F = \langle |f| w\rangle_I, f = \langle f, \rangle_I, \lambda = \lambda, w = \langle w\rangle_I, m = \inf_I w\,.$$

We are in the domain

$$\Omega := \{(F, w, m, f, \lambda) : F \geq |f|\, m,\;\; m \leq w \leq Q\, m\}\,. \tag{3.1}$$

Introduce

$$\mathcal{B}(F, w, m, f, \lambda) := \sup \frac{1}{|I|} w\{x \in I : \sum_{J\subseteq I, J\in D} \varepsilon_J(\varphi, h_J) h_J(x) > \lambda\}\,, \tag{3.2}$$

where the *sup* is taken over all $\varepsilon_J, |\varepsilon_J| \leq 1, J \in D$, $J \subseteq I$, and over all $f \in L^1(I, wdx)$ such that $F := \langle |f|\, w\rangle_I\,,\ f := \langle f\rangle_I$, $w = \langle w\rangle_I, m \leq \inf_I w$, and w are dyadic A_1 weights, such that $\forall I \in D\ \langle w\rangle_I \leq Q \inf_I w$, and Q being the best such constant. In other words $Q := [w]^{dyadic}_{A_1}$.

3.2.1 Homogeneity

By definition, it is clear that

$$s\mathcal{B}(F/s, w/s, m/s, f, \lambda) = \mathcal{B}(F, w, m, f, \lambda)\,,$$

$$\mathcal{B}(tF, w, m, tf, t\lambda) = \mathcal{B}(F, w, m, f, \lambda)\,.$$

Choosing $s = m$ and $t = \lambda^{-1}$, we can see that

$$\mathcal{B}(F, w, m, f, \lambda) = mB(\frac{F}{m\lambda}, \frac{w}{m}, \frac{f}{\lambda}) \tag{3.3}$$

for a certain function B. Introducing new variables $\alpha = \frac{F}{m\lambda}, \beta = \frac{w}{m}, \gamma = \frac{f}{\lambda}$ we write that B is defined in

$$G := \{(\alpha, \beta, \gamma) : |\gamma| \leq \alpha, 1 \leq \beta \leq Q\}\,. \tag{3.4}$$

3.2.2 The main inequality

Theorem 3.2.1. *Let* $P, P_+, P_- \in \Omega, P = (F, w, \min(m_+, m_-), f, \lambda),\ P_+ = (F + \alpha, w + \gamma, m_+, f + \beta, \lambda + \beta),\ P_- = (F - \alpha, w - \gamma, m_-, f - \beta, \lambda - \beta)$. *Then*

$$\mathcal{B}(P) - \frac{1}{2}(\mathcal{B}(P_+) + \mathcal{B}(P_-)) \geq 0\,. \tag{3.5}$$

At the same time, if $P, P_+, P_- \in \Omega, P = (F, w, \min(m_+, m_-), f, \lambda),\ P_+ = (F + \alpha, w + \gamma, m_+, f + \beta, \lambda - \beta),\ P_- = (F - \alpha, w + \gamma, m_+, f - \beta, \lambda + \beta)$. *Then*

$$\mathcal{B}(P) - \frac{1}{2}(\mathcal{B}(P_+) + \mathcal{B}(P_-)) \geq 0\,. \tag{3.6}$$

In particular, with fixed m*, and with all points being inside* Ω *we get*

$$\mathcal{B}(F, w, m, f, \lambda) - \frac{1}{4}(\mathcal{B}(F - dF, w - dw, m, f - d\lambda, \lambda - d\lambda) + \mathcal{B}(F - dF, w - dw, m, f + d\lambda, \lambda - d\lambda) +$$

$$\mathcal{B}(F + dF, w + dw, m, f - d\lambda, \lambda + d\lambda) + \mathcal{B}(F + dF, w + dw, m, f + d\lambda, \lambda + d\lambda)) \geq 0\,. \tag{3.7}$$

Remark.1) Differential notations $dF, dw, d\lambda$ just mean small numbers. 2) In (3.7) we loose a bit of information (in comparison to (3.5),(3.6)), but this is exactly (3.7) that we are going to use in the future.

Proof. Fix $P, P_+, P_- \in \Omega$. Let φ_+, φ_-, w_+, w_- be functions and weights giving the supremum in $B(P_+), B(P_-)$ respectively up to a small number $\eta > 0$. Using the fact that $\mathcal{B}$ does not depend on I, we think that φ_+, w_+ is on I_+ and φ_-, w_- is on I_-. Consider

$$\varphi(x) := \begin{cases} \varphi_+(x)\,,\ x \in I_+ \\ \varphi_-(x)\,,\ x \in I_- \end{cases}$$

$$\omega(x) := \begin{cases} w_+(x)\,,\ x \in I_+ \\ w_-(x)\,,\ x \in I_- \end{cases}$$

Notice that then

$$(\varphi, h_I) \cdot \frac{1}{\sqrt{|I|}} = \beta\,. \tag{3.8}$$

Then it is easy to see that

$$\langle |\varphi|\omega\rangle_I = F = P_1,\ \ \langle\varphi\rangle_I = f = P_4\,. \tag{3.9}$$

Notice that for $x \in I_+$ using (3.8), we get if $\varepsilon_I = -1$

$$\frac{1}{|I|} w_+ \{x \in I_+ : \sum_{J \subseteq I_+, J \in D} \varepsilon_J (\varphi, h_J) h_J(x) > \lambda\} =$$

$$\frac{1}{|I|} w_+ \{x \in I_+ : \sum_{J \subseteq I_+, J \in D} \varepsilon_J (\varphi, h_J) h_J(x) > \lambda + \beta\}$$

$$= \frac{1}{2|I_+|} w_+ \{x \in I_+ : \sum_{J \subseteq I_+, J \in D} \varepsilon_J (\varphi_+, h_J) h_J(x) > P_{+,3}\} \geq \frac{1}{2} B(P_+) - \eta \, .$$

Similarly, for $x \in I_-$ using (3.8), we get if $\varepsilon_I = -1$

$$\frac{1}{|I|} w_- \{x \in I_- : \sum_{J \subseteq I, J \in D} \varepsilon_J (\varphi, h_J) h_J(x) > \lambda\} =$$

$$\frac{1}{|I|} w_- \{x \in I_- : \sum_{J \subseteq I_-, J \in D} \varepsilon_J (\varphi, h_J) h_J(x) > \lambda - \beta\}$$

$$= \frac{1}{2|I_-|} w_- \{x \in I_- : \sum_{J \subseteq I_-, J \in D} \varepsilon_J (\varphi_-, h_J) h_J(x) > P_{-,3}\} \geq \frac{1}{2} B(P_-) - \eta \, .$$

Combining the two left hand sides we obtain for $\varepsilon_I = -1$

$$\frac{1}{|I|} \omega \{x \in I_+ : \sum_{J \subseteq I, J \in D} \varepsilon_J (\varphi, h_J) h_J(x) > \lambda\} \geq \frac{1}{2} (B(P_+) + B(P_-)) - 2\eta \, .$$

Let us use now the simple information (3.9): if we take the supremum in the left hand side over all functions φ, such that $\langle |\varphi| \, w \rangle_I = F, \langle \varphi \rangle_I = f, \langle \omega \rangle = w$, and weights ω: $\langle \omega \rangle = w$, in dyadic A_1 with A_1-norm at most Q, and supremum over all $\varepsilon_J = \pm 1$ (only $\varepsilon_I = -1$ stays fixed), we get a quantity smaller or equal than the one, where we have the supremum over all functions φ, such that $\langle |\varphi| \, \omega \rangle = F, \langle \varphi \rangle_I = f, \langle \omega \rangle = w$, and weights ω: $\langle \omega \rangle = w$, in dyadic A_1 with A_1-norm at most Q, and an unrestricted supremum over all $\varepsilon_J = \pm 1$ including

$\varepsilon_I = \pm 1$. The latter quantity is of course $\mathcal{B}(F, w, m, f, \lambda)$. So we proved (3.5).

To prove (3.6) we repeat verbatim the same reasoning, only keeping now $\varepsilon_I = 1$. We are done.

□

Remark. This theorem is a sort of "fancy" concavity property, the attentive reader would see that (3.5), (3.6) represent bi-concavity not unlike demonstrated by the celebrated Burkholder's function. We will use the consequence of bi-concavity encompassed by (3.7). There is still another concavity if we allow to have $|\varepsilon_J| \leq 1$.

Theorem 3.2.2. *In the definition of* $\mathcal{B}$ *we allow now to take supremum over all* $|\varepsilon_j| \leq 1$. *Let* $P, P_+, P_- \in \Omega, P = (F, w, m, f, \lambda)$, $P_+ = (F + \alpha, w + \gamma, m, f + \beta, \lambda)$, $P_- = (F - \alpha, w - \gamma, m, f - \beta, \lambda)$. *Then*

$$\mathcal{B}(P) - \frac{1}{2}(\mathcal{B}(P_+) + \mathcal{B}(P_-)) \geq 0\,. \tag{3.10}$$

Proof. We repeat the proof of (3.5) but with $\varepsilon_I = 0$.

□

Theorem 3.2.3. *For fixed* F, w, f, λ *function* $\mathcal{B}$ *is decreasing in* m.

Proof. Let $m = \min(m_-, m_+) = m_-$. And let $m_+ > m$. Then (3.5) becomes

$$\mathcal{B}(F, w, m, f, \lambda) - \mathcal{B}(F, w, m_+, f, \lambda) \geq 0\,.$$

This is what we want. □

3.3 The unweighted estimate: the exact Bellman function

We first deal with the case when there is no weight, i.e. with the case when $w = 1$ a.e. We notice that this is the boundary of our domain Ω: $w = m$.

Introduce a function

$$\mathcal{B}_0(\lambda, f, F) = \sup \left| \left\{ x\colon \sum_{I \subset I_0,\, I \in D} \varepsilon_I (\varphi, h_I) h_I(x) \geqslant \lambda \right\} \right|,$$

where the supremum is taken over all families $\{\varepsilon_I\}$ such that $|\varepsilon_I| = 1$, and all functions φ with $\langle |\varphi| \rangle_{I_0} = F$, $\langle \varphi \rangle_{I_0} = f$.

Let $\Omega_0 = \{(\lambda, f, F)\colon F \geqslant |f|\}$ be the domain of $\mathcal{B}_0$.

Denote

$$B_0(\lambda, f, F) = \begin{cases} 1, & \lambda \leqslant F \\ 1 - \frac{(\lambda - F)^2}{\lambda^2 - f^2}, & \lambda \geqslant F, \end{cases} \qquad (F, f, \lambda) \in \Omega_0.$$

Our main theorem is the following.

Theorem 3.3.1. *For any* $(\lambda, f, F) \in \Omega_0$ *it holds that* $\mathcal{B}_0(F, f, \lambda) = B_0(F, f, \lambda)$.

Firstly, it will be more convenient to work with a slightly modified function. We need a definition.

Definition 21. A function ψ is called a martingale transform of a function φ, if for some family $\{\varepsilon_I\}$, with $|\varepsilon_I| = 1$,

$$\psi(x) = \langle \psi \rangle_{I_0} + \sum_{I \subset I_0,\, I \in D} \varepsilon_I (\varphi, h_I) h_I(x), \quad x \in I_0.$$

Denote

$$\mathcal{B}(g,f,F) = \sup |\{x\colon \psi(x) \geqslant 0\}|,$$

where the supremum is taken over all functions φ with $\langle|\varphi|\rangle_{I_0} = F$, $\langle\varphi\rangle_{I_0} = f$, and all martingale transforms ψ of φ with $\langle\psi\rangle_{I_0} = g$. It is easy to see that

$$\mathcal{B}_0(\lambda,f,F) = \mathcal{B}(-g,f,F).$$

Denote $\Omega = \{(g,f,F)\colon F \geqslant |f|\}$ and

$$B(g,f,F) = \begin{cases} 1, & -g \leqslant F \\ 1 - \frac{(g+F)^2}{g^2-f^2}, & -g \geqslant F, \end{cases} \qquad (g,f,F) \in \Omega.$$

Then our main theorem is equivalent to the following one.

Theorem 3.3.2. *For any $(g,f,F) \in \Omega$ it holds $\mathcal{B}(g,f,F) = B(g,f,F)$.*

Corollary 3.3.3. *For any function $\varphi \in L^1$, any number $\lambda \geqslant 0$ and any family $\{\varepsilon_I\}$ with $|\varepsilon_I| = 1$ it holds*

$$\left|\left\{x\colon \sum_{I\subset I_0,\, I\in D} \varepsilon_I(\varphi,h_I)h_I(x) \geqslant \lambda\right\}\right| \leqslant 2\frac{\|\varphi\|_1}{\lambda}$$

Proof. It is easy to verify that

$$\sup\left(\mathcal{B}_0(\lambda,f,F)\cdot\frac{\lambda}{F}\right) = 2.$$

Thus,

$$\left|\left\{x\colon \sum_{I\subset I_0,\, I\in D} \varepsilon_I(\varphi,h_I)h_I(x) \geqslant \lambda\right\}\right| \leqslant 2\frac{F}{\lambda} = 2\frac{\|\varphi\|_1}{\lambda}.$$

□

Corollary 3.3.4. *For any function $\varphi \in L^1$, any number $\lambda \geqslant 0$ and any family $\{\varepsilon_I\}$ with $|\varepsilon_I| = 1$ it holds*

$$\left|\left\{x\colon \sum_{I\subset I_0,\, I\in D} \varepsilon_I(\varphi, h_I)h_I(x) \geqslant \lambda\right\}\right| \leqslant 4\frac{\|\varphi\|_1}{\lambda}$$

Proof.

$$\left|\left\{x\colon \sum_{I\subset I_0,\, I\in D} |\varepsilon_I(\varphi, h_I)h_I(x)| \geqslant \lambda\right\}\right| = \left|\left\{x\colon \sum_{I\subset I_0,\, I\in D} \varepsilon_I(\varphi, h_I)h_I(x) \geqslant \lambda\right\}\right| + \left|\left\{x\colon \sum_{I\subset I_0,\, I\in D} \varepsilon_I(-\varphi, h_I)h_I(x) \geqslant \lambda\right\}\right| \leqslant 4\frac{\|\varphi\|_1}{\lambda} \quad (3.11)$$

□

We start to prove our main theorem.

3.3.1 $B \geqslant \mathcal{B}$

We need a technical lemma.

Lemma 3.3.5. *Let $x^\pm$ be two points in Ω such tat $|f^+ - f^-| = |g^+ - g^-|$ and $x = \frac{1}{2}(x^+ + x^-)$. Then*

$$B(x) - \frac{B(x^+) + B(x^-)}{2} \geq 0\,. \quad (3.12)$$

Given the lemma, we prove the following theorem.

Theorem 3.3.6. *For any point $x \in \Omega$ it holds $B(x) \geqslant \mathcal{B}(x)$.*

Proof. Let us fix a point $x \in \Omega$ and a pair of admissible functions φ, ψ on I_0 corresponding to x. For any $I \in D$ by the symbol x^I we denote the point $(\langle\psi\rangle_I, \langle\varphi\rangle_I, \langle|\varphi|\rangle_I,)$. We notice that since ψ is a martingale transform of φ, we always have

$$|f^{I^+} - f^{I^-}| = |g^{I^+} - g^{I^-}|,$$

and

$$x^I = \frac{x^{I^+} + x^{I^-}}{2}.$$

Using consequently main inequality for the function B we can write down the following chain of inequalities

$$B(x) \geq \frac{1}{2}\big(B(x^{I_0^+}) + B(x^{I_0^-})\big) \geq \sum_{I \in D,\, |I| = 2^{-n}} \frac{1}{|I|} B(x^I) = \int_0^1 B(x^{(n)}(t))dt\,,$$

where $x^{(n)}(t) = x^I$, if $t \in I$, $|I| = 2^{-n}$.

Note that $x^{(n)}(t) \to (\psi(t), \varphi(t), |\varphi(t)|)$ almost everywhere (at any Lebesgue point t), and therefore, since B is continuous and bounded, we can pass to the limit in the integral. So, we come to the inequality

$$B(x) \geq \int_0^1 B(\psi(t), \varphi(t), |\varphi(t)|)dt \geq \int_{\{t\colon \psi(t) \geq 0\}} = \big|\{t \in I_0 \colon \psi(t) \geq 0\}\big| \tag{3.13}$$

where we have used the property $B(g, f, |f|) = 1$ for $g \geq 0$. Now, taking supremum in (3.13) over all admissible pairs φ, ψ, we get the required estimate $B(x) \geq \mathcal{B}(x)$. □

3.3.2 $B(g, f, F) \leqslant \mathcal{B}(g, f, F)$

This section is devoted to the following theorem.

Theorem 3.3.7. *For any point $x \in \Omega$ it holds $B(x) \leqslant \mathcal{B}(x)$.*

To prove the theorem we need to present two sequences of functions $\{\varphi_n\}$, $\{\psi_n\}$, such that

- For every n the function ψ_n is a martingale transform of φ_n;
- For every n: $\langle|\varphi_n|\rangle_{I_0} = F$, $\langle\varphi_n\rangle_{I_0} = f$, $\langle\psi_n\rangle_{I_0} = g$;
- It holds that $B(g, f, F) = \lim\limits_{n\to\infty} |\{x\colon \psi_n(x) \geqslant 0\}|$.

We need the following definition.

Definition 22. We call a pair (φ, ψ) admissible for the point (g, f, F) if ψ is a martingale transform of φ, and $\langle|\varphi|\rangle_{I_0} = F$, $\langle\varphi\rangle_{I_0} = f$, $\langle\psi\rangle_{I_0} = g$.

Definition 23. We call a pair (φ, ψ) an ε-extremizer for a point (g, f, F), if this pair is admissible for this point and $|\{x\colon \psi(x) \geqslant 0\}| \geqslant B(g, f, F) - \varepsilon$.

The following lemma is almost obvious.

Lemma 3.3.8. 1. *For a positive number s: $B(sg, sf, sF) = B(g, f, F)$. Moreover, if a pair (φ, ψ) is admissible for a point (g, f, F) then $(s\varphi, s\psi)$ is admissible for (sg, sf, sF). If a pair (φ, ψ) is an ε-extremizer for a point (g, f, F) then $(s\varphi, s\psi)$ is an ε-extremizer for (sg, sf, sF).*

2. *$B(g, f, F) = B(g, -f, F)$. Moreover, if a pair (φ, ψ) is admissible for a point (g, f, F) then $(-\varphi, \psi)$ is admissible for $(g, -f, F)$. If a pair (φ, ψ) is an ε-extremizer for a point (g, f, F) then $(-\varphi, \psi)$ is an ε-extremizer for $(g, -f, F)$.*

The next lemma is a key to our "splitting" technique.

Lemma 3.3.9. *Suppose two pairs $(\varphi_\pm, \psi_\pm)$ are admissible for points $(g^\pm, f^\pm, F^\pm)$ correspondingly. Suppose also that*

$$F = \frac{F^+ + F^-}{2}, \quad f = \frac{f^+ + f^-}{2}, \quad , g = \frac{g^+ + g^-}{2}, \quad |f^+ - f^-| = |g^+ - g^-|.$$

Then a pair (φ, ψ) is admissible for the point (g, f, F), where

$$\varphi(x) = \begin{cases} \varphi_-(2x), & x \in [0, \frac{1}{2}) \\ \varphi_+(2x-1), & x \in [\frac{1}{2}, 1], \end{cases} \qquad \psi(x) = \begin{cases} \psi_-(2x), & x \in [0, \frac{1}{2}) \\ \psi_+(2x-1), & x \in [\frac{1}{2}, 1]. \end{cases}$$

Proof. It is clear that $\langle \varphi \rangle_{I_0} = f$, $\langle \psi \rangle_{I_0} = g$, and $\langle |\varphi| \rangle_{I_0} = F$. All we need to prove is that for any interval I it is true that

$$|(\psi, h_I)| = |(\varphi, h_I)|.$$

For any interval $I \neq I_0$ it is obvious, since pairs $(\varphi^\pm, \psi_\pm)$ are admissible for corresponding points. Thus, we need to show that

$$|(\varphi, h_{I_0})| = |(\psi, h_{I_0})|.$$

But

$$(\varphi, h_{I_0}) = \langle \varphi \rangle_{[\frac{1}{2},1]} - \langle \varphi \rangle_{[0,\frac{1}{2}]} = \langle \varphi_+ \rangle_{[0,1]} - \langle \varphi_- \rangle_{[0,1]} = f^+ - f^-,$$

$$(\psi, h_{I_0}) = \langle \psi \rangle_{[\frac{1}{2},1]} - \langle \psi \rangle_{[0,\frac{1}{2}]} = \langle \psi_+ \rangle_{[0,1]} - \langle \psi_- \rangle_{[0,1]} = g^+ - g^-,$$

which finishes our proof. $\square$

We generalize this lemma a little.

Lemma 3.3.10. *Suppose two pairs $(\varphi_\pm, \psi_\pm)$ are admissible for points $(g^\pm, f^\pm, F^\pm)$ correspondingly. Suppose also that*

$$F = \frac{F^+ + F^-}{2}, \quad f = \frac{f^+ + f^-}{2}, \quad , g = \frac{g^+ + g^-}{2}, \quad |f^+ - f^-| = |g^+ - g^-|.$$

Suppose I is a dyadic interval with "sons" $I_\pm$. Suppose that a pair (Φ, Ψ) is admissible for some point (g^0, f^0, F^0). Suppose that

$$\forall\, x \in I \;\; \Phi(x) = \varphi^I(x), \quad \Psi(x) = \psi^I(x),$$

where the pair (φ, ψ) is admissible for the point (g, f, F). Then the pair (Φ_1, Ψ_1), defined below, is admissible for the point (g^0, f^0, F^0):

$$\Phi_1(x) = \begin{cases} \Phi(x), & x \notin I \\ \varphi_+^{I_+}(x), & x \in I_+ \\ \varphi_-^{I_-}(x), & x \in I_- \end{cases}, \quad \Psi_1(x) = \begin{cases} \Psi(x), & x \notin I \\ \psi_+^{I_+}(x), & x \in I_+ \\ \psi_-^{I_-}(x), & x \in I_- \end{cases}$$

Essentially this lemma says that if we have pairs $(\varphi_\pm, \psi_\pm)$, and and a pair (φ, ψ) defined in the Lemma 3.3.9, then we can split this pair into $(\varphi^\pm, \psi^\pm)$, defined on $I^\pm$ correspondingly. The proof of the Lemma 3.3.10 is essentially the same as the proof of the Lemma 3.3.9.

3.3.2.1 Change of variables

It will be more convenient for us to work in variables

$$y_1 = \frac{f-g}{2}, \quad y_2 = \frac{-f-g}{2}, \quad F.$$

We define $M(y_1, y_2, F) = B(g, f, F)$. Then all properties of B are easily translated to properties of M. Moreover, the "splitting" lemmas 3.3.9, 3.3.10 remain true for fixed y_1 or fixed y_2.

If we have a point (y_1, y_2, F) then by $(\varphi_{(y_1,y_2,F)}, \psi_{(y_1,y_2,F)})$ we denote an admissible pair for this point. An individual function $\varphi_{(y_1,y_2,F)}$ is always such that there is a function $\psi_{(y_1,y_2,F)}$, such that the pair $(\varphi_{(y_1,y_2,F)}, \psi_{(y_1,y_2,F)})$ is admissible for (y_1, y_2, F).

3.3.2.2 The proof of $\mathcal{B} \geqslant B$

We will work in the y-variables. In these variables it is true that the function M is concave when y_1 or y_2 is fixed. This is proved in the Theorem 3.3.4. Analogously to the previous definition, we define

$$\mathcal{M}(y_1, y_2, F) = \mathcal{B}(g, f, F).$$

We first prove that

$$\mathcal{M}(1, 1, F) \geqslant M(1, 1, F).$$

Fix a large integer r and set $\delta = \frac{1}{2^r}$. We notice the following chain of inequalities:

$$\mathcal{M}(1,1,F) \geqslant \frac{1}{2}\left(\mathcal{M}(1,1-\delta,F+\delta(1-F)) + \mathcal{M}(1,1+\delta,F-\delta(1-F))\right) =$$
$$= \frac{1}{2}\left(\mathcal{M}(1,1-\delta,F+\delta(1-F)) + \mathcal{M}(1+\delta,1,F-\delta(1-F))\right). \quad (3.14)$$

Applying the same concavity we see that

$$\mathcal{M}(1,1-\delta,F+\delta(1-F)) \geqslant \delta\mathcal{M}(1,0,1) + (1-\delta)\mathcal{M}(1,1,F) = \delta + (1-\delta)\mathcal{M}(1,1,F).$$

Moreover, by the concavity

$$\mathcal{M}(1+\delta,1,F-\delta(1-F)) \geqslant$$
$$(\delta-\delta^2)\mathcal{M}(1+\delta,0,1+\delta) + (1-\delta)\mathcal{M}(1+\delta,1+\delta,(1+\delta)(F-\delta(2-F)))+$$
$$\delta^2\mathcal{M}(1+\delta,1,F-\delta(1-F)) \geqslant \delta-\delta^2+(1-\delta)\mathcal{M}(1,1,F-\delta(2-F)) \quad (3.15)$$

Therefore, we get

$$\mathcal{M}(1,1,F) \geqslant \frac{1}{2}\Big(\delta + (1-\delta)\mathcal{M}(1,1,F) + \delta - \delta^2 + (1-\delta)\mathcal{M}(1,1,F-\delta(2-F))\Big),$$

or

$$\mathcal{M}(1,1,F) \geqslant \frac{2\delta-\delta^2}{1+\delta} + \frac{1-\delta}{1+\delta}\mathcal{M}(1,1,F-\delta(2-F)).$$

Notice that it is true for any F. We now denote

$$F^k = 2-(2-F)(1+\delta)^k.$$

Then, clearly, $F^0 = F$, and $F^{k+1} = F^k - \delta(2 - F^k)$. With this notation we get

$$\mathcal{M}(1,1,F) \geqslant \frac{2\delta - \delta^2}{1+\delta} \sum_{k=0}^{K} \left(\frac{1-\delta}{1+\delta}\right)^k + \left(\frac{1-\delta}{1+\delta}\right)^{K+1} \cdot \mathcal{M}(1,1,F^{K+1}).$$

3.3.2.3 The case $F \geqslant 2$

In this case we have $F^{k+1} \geqslant F^k$, and therefore the point $(1, 1+\delta, F^k - \delta(1 - F^k))$ always lies in Ω. Thus, we can take K as huge as we want. Therefore,

$$\mathcal{M}(1,1,F) \geqslant \frac{2\delta - \delta^2}{1+\delta} \sum_{k=0}^{\infty} \left(\frac{1-\delta}{1+\delta}\right)^k = \frac{2\delta - \delta^2}{2\delta}.$$

This is true for arbitrary small δ, and thus $\mathcal{M}(1,1,F) \geqslant 1$.

3.3.2.4 The case $F \leqslant 2$

In this case to assure that $(1, 1+\delta, F^k - \delta(1 - F^k)) \in \Omega$ we need $F^k - \delta(1 - F^k) \geqslant \delta$, which implies

$$(1+\delta)^{K+1} \leqslant \frac{2}{2-F}.$$

Take $K \in [\frac{\log \frac{2}{2-F}}{\log(1+\delta)} - 10, \frac{\log \frac{2}{2-F}}{\log(1+\delta)} + 10]$, such that this inequality holds. Then we get

$$\mathcal{M}(1,1,F) \geqslant \frac{2\delta - \delta^2}{1+\delta} \sum_{k=0}^{K} \left(\frac{1-\delta}{1+\delta}\right)^k = \frac{2\delta - \delta^2}{2\delta} \left(1 - \left(\frac{1-\delta}{1+\delta}\right)^{K+1}\right).$$

It is only left to notice that with our choise of K we have

$$\left(\frac{1-\delta}{1+\delta}\right)^{K+1} \to \frac{(2-F)^2}{4}, \quad \delta \to 0,$$

and therefore

$$\mathcal{M}(1,1,F) \geqslant 1 - \frac{(2-F)^2}{4} = M(1,1,F).$$

We leave the proof of the general inequality $\mathcal{M}(y_1, y_2, F) \geqslant M(y_1, y_2, F)$ to the reader. In fact, it is a simple use of the concavity of $\mathcal{M}$ along the line that connects $(y_1, 0, y_1)$ with (y_1, y_2, F).

3.3.3 Building the extremal sequense for points $(1, 1, F)$

The aim of this Section is to prove that $B(g, f, F) \leqslant \mathcal{B}(g, f, F)$ by a construction of an extremal sequense of pairs (φ_n, ψ_n). For the sake of simplicity, we do it only for the case $f - g = 2$.

Due to the homogeneity and symmetry of the function B it is enough to prove that

$$B(g, f, F) \leqslant \mathcal{B}(g, f, F)$$

for $f \geqslant 0$, $f - g = 2$. In the new variables it means that we consider the case $y_1 = 1$, and $y_2 \leqslant y_1 = 1$. As we have seen, for $f \geqslant -g$ we have $B(g, f, F) = \mathcal{B}(g, f, F) = 1$, and so we need to consider the case $f \leqslant -g$, i.e. $y_2 \geqslant 0$. We first build the ε-extremizer for the point $(F, 1, 1)$.

Fix a large integer r and let $\delta = 2^{-r}$. As before, denote $I_0 = [0, 1]$. Also denote $J_i = [2^{-i}, 2^{-i+1})$, Denote $m_i(x) = 2^i x - 1$ — the linear function from J_k onto I_0.

We need the following lemma.

Lemma 3.3.11. *Suppose $\delta = 2^{-r}$ is small enough. Also, fix a small number $\varepsilon > 0$. Suppose $F^1 = F - \delta(2 - F)$, and the pair $(\varphi_{(1,1,F^1)}, \psi_{(1,1,F^1)})$ is admissible. Then there exists an* *admissible pair* $(\varphi_{(1,1,F)}, \psi_{(1,1,F)})$ *such that*

$$|\{x\colon \psi_{(1,1,F)} \geqslant 0\}| \geqslant \frac{2\delta-\delta^2}{1+\delta} + \frac{1-\delta}{1+\delta}|\{x\colon \psi_{(1,1,F^1)} \geqslant 0\}| - \varepsilon. \tag{3.16}$$

Proof. First, we explain our strategy. In what follows, we always assume that functions on the right-hand side are already defined. We specify their definition later; however, we clearly indicate points to which the functions are admissible.

We define

$$\varphi_{(1,1,F)}(x) = \begin{cases} \varphi_{(1,1-\delta,F+\delta(1-F))}(m_1(x)), & x \in J_1 \\ \varphi_{(1,1+\delta,F-\delta(1-F))}(2x), & x \in [0,\frac{1}{2}). \end{cases}$$

$$\psi_{(1,1,F)}(x) = \begin{cases} \psi_{(1,1-\delta,F+\delta(1-F))}(m_1(x)), & x \in J_1 \\ \psi_{(1,1+\delta,F-\delta(1-F))}(2x), & x \in [0,\frac{1}{2}). \end{cases}$$

By the Lemma 3.3.10 we see that $\psi_{(1,1,F)}$ is a martingale transform of $\varphi_{(1,1,F)}$. We define next

$$\varphi_{(1,1,F)}(x) = \begin{cases} \varphi_{(1,0,1)}(\frac{m_1(x)}{\delta})), & x \in m_1^{-1}(\delta I_0) \\ \varphi_{(1,1,F)}(m_k(m_1(x))), & x \in m_1^{-1} m_k^{-1}(I_0),\ k = 1 \ldots r \\ -\varphi_{(1+\delta,1,F-\delta(1-F))}(2x), & x \in [0, \frac{1}{2}). \end{cases}$$

$$\varphi_{(1,1,F)}(x) = \begin{cases} \psi_{(1,0,1)}(\frac{m_1(x)}{\delta})), & x \in m_1^{-1}(\delta I_0) \\ \psi_{(1,1,F)}(m_k(m_1(x))), & x \in m_1^{-1} m_k^{-1}(I_0),\ k = 1 \ldots r \\ \psi_{(1+\delta,1,F-\delta(1-F))}(2x), & x \in [0, \frac{1}{2}). \end{cases} \tag{3.17}$$

By the Lemma 3.3.8 and a multiple application of the Lemma 3.3.10, we still get an admissible pair for the point $(1, 1, F)$.

Finally, define

$$\varphi_{(1+\delta,1,F-\delta(1-F))}(x) = \begin{cases} \varphi_{(1+\delta,0,1+\delta)}(m_k(\frac{x}{\delta})), & x \in \delta \cdot J_k,\ k = 1 \ldots r \\ \varphi_{(1+\delta,1,F-\delta(1-F))}(\frac{x}{\delta^2}), & x \in [0, \delta^2) \\ (1+\delta)\varphi_{(1,1,F-\delta(2-F))}(m_k(x)), & x \in J_k,\ k = 1 \ldots r \end{cases}$$

$$\psi_{(1+\delta,1,F-\delta(1-F))}(x) = \begin{cases} \psi_{(1+\delta,0,1+\delta)}(m_k(\frac{x}{\delta})), & x \in \delta \cdot J_k,\ k = 1 \ldots r \\ \psi_{(1+\delta,1,F-\delta(1-F))}(\frac{x}{\delta^2}), & x \in [0, \delta^2) \\ (1+\delta)\psi_{(1,1,F-\delta(2-F))}(m_k(x)), & x \in J_k,\ k = 1 \ldots r \end{cases} \tag{3.18}$$

Again, the Lemma 3.3.8 and the Lemma 3.3.10 assure that the defined pair is admissible.

Bringing everything together, we get

$$\varphi_{(1,1,F)}(x) = \begin{cases} \varphi_{(1,0,1)}(\frac{m_1(x)}{\delta})), & x \in m_1^{-1}(\delta I_0) \\ \varphi_{(1,1,F)}(m_k(m_1(x))), & x \in m_1^{-1}m_k^{-1}(I_0),\ k = 1\dots r \\ -\varphi_{(1+\delta,0,1+\delta)}(m_k(\frac{2x}{\delta})), & x \in \frac{\delta}{2}m_k^{-1}(I_0),\ k = 1\dots r \\ -\varphi_{(1+\delta,1,F-\delta(1-F))}(\frac{2x}{\delta^2}), & x \in [0, \frac{\delta^2}{2}) \\ -(1+\delta)\varphi_{(1,1,F-\delta(2-F))}(m_k(2x)), & x \in \frac{1}{2}m_k^{-1}(I_0),\ k = 1\dots r. \end{cases}$$

$$\psi_{(1,1,F)}(x) = \begin{cases} \psi_{(1,0,1)}(\frac{m_1(x)}{\delta})), & x \in m_1^{-1}(\delta I_0) \\ \psi_{(1,1,F)}(m_k(m_1(x))), & x \in m_1^{-1}m_k^{-1}(I_0),\ k = 1\dots r \\ \psi_{(1+\delta,0,1+\delta)}(m_k(\frac{2x}{\delta})), & x \in \frac{\delta}{2}m_k^{-1}(I_0),\ k = 1\dots r \\ \psi_{(1+\delta,1,F-\delta(1-F))}(\frac{2x}{\delta^2}), & x \in [0, \frac{\delta^2}{2}) \\ (1+\delta)\psi_{(1,1,F-\delta(2-F))}(m_k(2x)), & x \in \frac{1}{2}m_k^{-1}(I_0),\ k = 1\dots r. \end{cases} \tag{3.19}$$

We now specify definitions of functions on the right-hand side. The pair $(\varphi_{(1,0,1)}, \psi_{(1,0,1)})$ is a $\frac{\varepsilon}{2}$-extremizer for the point $(1, 0, 1)$. The pair $(\varphi_{(1+\delta,0,1+\delta)}, \psi_{(1+\delta,0,1+\delta)})$ is a $\frac{\varepsilon}{\delta-\delta^2}$-extremizer for the point $(1+\delta, 0, 1+\delta)$.

The pair $(\varphi_{(1,1,F-\delta(2-F))}, \psi_{(1,1,F-\delta(2-F))})$ is given in the lemma. As for the pair $(\varphi_{(1+\delta,1,F-\delta(1-F))}, \psi_{(1+\delta,1,F-\delta(1-F))})$ — we take any admissible pair for this point.

It is an easy calculation that the function $\psi_{(1,1,F)}$ satisfies the inequality (3.16). Moreover, it is easy to see that for **any** pair, defined by (3.19) we have $\langle\varphi_{(1,1,F)}\rangle_{I_0} - \langle\psi_{(1,1,F)}\rangle_{I_0} = 2$. Thus, what we need to show is that there exists an admissible pair $(\varphi_{(1,1,F)}, \psi_{(1,1,F)})$

that satisfies the self-similarity condition (3.19)

To do that, we first take any admissible pair $(\tilde{\varphi}_{(1,1,F)}, \tilde{\psi}_{(1,1,F)})$ and define

$$\varphi^0_{(1,1,F)}(x) = \begin{cases} \varphi_{(1,0,1)}(\frac{m_1(x)}{\delta})), & x \in m_1^{-1}(\delta I_0) \\ \tilde{\varphi}_{(1,1,F)}(m_k(m_1(x))), & x \in m_1^{-1} m_k^{-1}(I_0),\ k = 1 \dots r \\ -\varphi_{(1+\delta,0,1+\delta)}(m_k(\frac{2x}{\delta})), & x \in \frac{\delta}{2} m_k^{-1}(I_0),\ k = 1 \dots r \\ -\varphi_{(1+\delta,1,F-\delta(1-F))}(\frac{2x}{\delta^2}), & x \in [0, \frac{\delta^2}{2}) \\ -(1+\delta)\varphi_{(1,1,F-\delta(2-F))}(m_k(2x)), & x \in \frac{1}{2} m_k^{-1}(I_0),\ k = 1 \dots r. \end{cases}$$

$$\psi^0_{(1,1,F)}(x) = \begin{cases} \psi_{(1,0,1)}(\frac{m_1(x)}{\delta})), & x \in m_1^{-1}(\delta I_0) \\ \tilde{\psi}_{(1,1,F)}(m_k(m_1(x))), & x \in m_1^{-1} m_k^{-1}(I_0),\ k = 1 \dots r \\ \psi_{(1+\delta,0,1+\delta)}(m_k(\frac{2x}{\delta})), & x \in \frac{\delta}{2} m_k^{-1}(I_0),\ k = 1 \dots r \\ \psi_{(1+\delta,1,F-\delta(1-F))}(\frac{2x}{\delta^2}), & x \in [0, \frac{\delta^2}{2}) \\ (1+\delta)\psi_{(1,1,F-\delta(2-F))}(m_k(2x)), & x \in \frac{1}{2} m_k^{-1}(I_0),\ k = 1 \dots r. \end{cases} \tag{3.20}$$

Then the pair $(\varphi^0_{(1,1,F)}, \psi^0_{(1,1,F)})$ is admissible to point $(1, 1, F)$. It is true by the Lemma 3.3.10, and by an easy calculation that shows that all averages are as we need. We now define inductively

$$\varphi^{n+1}_{(1,1,F)}(x) = \begin{cases} \varphi_{(1,0,1)}(\frac{m_1(x)}{\delta})), & x \in m_1^{-1}(\delta I_0) \\ \varphi^n_{(1,1,F)}(m_k(m_1(x))), & x \in m_1^{-1}m_k^{-1}(I_0),\ k = 1\dots r \\ -\varphi_{(1+\delta,0,1+\delta)}(m_k(\frac{2x}{\delta})), & x \in \frac{\delta}{2}m_k^{-1}(I_0),\ k = 1\dots r \\ -\varphi_{(1+\delta,1,F-\delta(1-F))}(\frac{2x}{\delta^2}), & x \in [0, \frac{\delta^2}{2}) \\ -(1+\delta)\varphi_{(1,1,F-\delta(2-F))}(m_k(2x)), & x \in \frac{1}{2}m_k^{-1}(I_0),\ k = 1\dots r. \end{cases}$$

$$\psi^{n+1}_{(1,1,F)}(x) = \begin{cases} \psi_{(1,0,1)}(\frac{m_1(x)}{\delta})), & x \in m_1^{-1}(\delta I_0) \\ \psi^n_{(1,1,F)}(m_k(m_1(x))), & x \in m_1^{-1}m_k^{-1}(I_0),\ k = 1\dots r \\ \psi_{(1+\delta,0,1+\delta)}(m_k(\frac{2x}{\delta})), & x \in \frac{\delta}{2}m_k^{-1}(I_0),\ k = 1\dots r \\ \psi_{(1+\delta,1,F-\delta(1-F))}(\frac{2x}{\delta^2}), & x \in [0, \frac{\delta^2}{2}) \\ (1+\delta)\psi_{(1,1,F-\delta(2-F))}(m_k(2x)), & x \in \frac{1}{2}m_k^{-1}(I_0),\ k = 1\dots r. \end{cases} \tag{3.21}$$

Then for any n we get an admissible pair to the point $(1,1,F)$.

We need to notice that

$$\int\limits_{I_0} |\varphi^{n+1}_{(1,1,F)} - \varphi^n_{(1,1,F)}|^2 dx = \sum_k \frac{|J_k|}{2} \int\limits_{I_0} |\varphi^n_{(1,1,F)} - \varphi^{n-1}_{(1,1,F)}|^2 dx =$$
$$= \frac{1-\delta}{2} \int\limits_{I_0} |\varphi^n_{(1,1,F)} - \varphi^{n-1}_{(1,1,F)}|^2 dx =$$
$$= (\frac{1-\delta}{2})^n \int\limits_{I_0} |\varphi^1_{(1,1,F)} - \varphi^0_{(1,1,F)}|^2 dx. \tag{3.22}$$

Thus, we can take

$$\varphi_{(1,1,F)} = \lim \varphi^{n+1}_{(1,1,F)} \quad \text{in} \quad L^2(I_0).$$

Similarly

$$\psi_{(1,1,F)} = \lim \psi^{n+1}_{(1,1,F)} \quad \text{in} \quad L^2(I_0).$$

It is clear that the pair $(\varphi_{(1,1,F)}, \psi_{(1,1,F)})$ satisfies the self-similarity conditions (3.19). Moreover, since the limit in L^2 implies the limin in L^1, we get that all the averages are as needed. Moreover, for every interval I:

$$|(\varphi_{(1,1,F)}, h_I)| = \lim |(\varphi^n_{(1,1,F)}, h_I)| = |(\psi^n_{(1,1,F)}, h_I)| = |(\psi_{(1,1,F)}, h_I)|,$$

and thus we get an admissible pair. The proof of the lemma is finished. □

We are now ready to finish the whole construction. We consider a sequence

$$F^k = 2 - (2-F)(1+\delta)^k.$$

Then it is clear the $F^0 = F$ and $F^{k+1} = F^k - \delta(2 - F^k)$.

3.3.3.1 The case $F \geqslant 2$

We take a huge number N and a small number ε. For a point $(1, 1, F^N)$ we take any admissible pair $(\varphi_{(1,1,F^N)}, \psi_{(1,1,F^N)})$. Using the Lemma 3.3.11 N times we build an admissible pair $(\varphi_{(1,1,F)}, \psi_{(1,1,N)})$. Moreover, we get

$$|\{x \colon \psi_{(1,1,F)}(x) \geqslant 0\}| \geqslant \frac{2\delta - \delta^2}{1+\delta} \sum_{k=0}^{N} \left(\frac{1-\delta}{1+\delta}\right)^k - N\varepsilon.$$

We now specify the choise of δ, N and ε. We first fix a small δ, so that $\frac{2\delta-\delta^2}{2\delta} = 1-\sigma$. Then fix a huge number N, such that $\sum\limits_{k=0}^{N}\left(\frac{1-\delta}{1+\delta}\right)^k > \frac{1+\delta}{2\delta} - \sigma\frac{1+\delta}{2\delta-\delta^2}$. Finally, fix a very small number ε, such that $N\varepsilon < \sigma$. Then we get

$$|\{x\colon \psi_{(1,1,F)}(x) \geqslant 0\}| \geqslant \frac{2\delta-\delta^2}{1+\delta}\left(\frac{1+\delta}{2\delta} - \sigma\frac{1+\delta}{2\delta-\delta^2}\right) - \sigma = 1-3\sigma.$$

where σ is an arbitrary small number.

3.3.3.2 The case $F < 2$

We remind that our very first step requires that the point $(1, 1+\delta, F-\delta(1-F))$ to be in our domain. Thus, the on the N-th iteration we need that the point $(1, 1+\delta, F^N - \delta(1-F^N))$ is in the domain $\Omega = \{(y_1, y_2, F)\colon F \geqslant |y_1 - y_2|\}$. This yields to the inequality

$$(1+\delta)^{N+1} < \frac{2}{2-F}.$$

Thus, we should stop at the K-th step with

$$(1+\delta)^{N+1} \approx \frac{2}{2-F}.$$

Here the sign "$\approx$" means that

$$N \in [\frac{\log\frac{2}{2-F}}{\log(1+\delta)} - 10, \frac{\log\frac{2}{2-F}}{\log(1+\delta)} + 10].$$

We again apply the Lemma 3.3.11 N times and get

$$|\{x\colon \psi_{(1,1,F)}(x) \geqslant 0\}| \geqslant \frac{2\delta-\delta^2}{1+\delta}\sum_{k=0}^{N}\left(\frac{1-\delta}{1+\delta}\right)^k - N\varepsilon = \frac{2\delta-\delta^2}{2\delta}\left(1-\left(\frac{1-\delta}{1+\delta}\right)^{N+1}\right) - N\varepsilon$$

Finally, since

$$N \in [\frac{\log\frac{2}{2-F}}{\log(1+\delta)} - 10, \frac{\log\frac{2}{2-F}}{\log(1+\delta)} + 10]$$

we get that $\delta \to 0$ implies $1-\left(\frac{1-\delta}{1+\delta}\right)^{N+1} \to 1-\frac{(2-F)^2}{4}$, which finishes our proof.

3.3.4 How to find the Bellman function $\mathcal{B}$

In this section we explain how did we search for the function $\mathcal{B}$ and find it. We start with the following lemma. Let $x^{\pm}$ be two points in Ω such tat $|f^+ - f^-| = |g^+ - g^-|$ and $x = \frac{1}{2}(x^+ + x^-)$. Then

$$\mathcal{B}(x) - \frac{\mathcal{B}(x^+) + \mathcal{B}(x^-)}{2} \geq 0\,. \tag{3.23}$$

Proof. Fix $x^{\pm} \in \Omega$, and let $(\varphi^{\pm}, \psi^{\pm})$ be two pairs of functions giving the supremum for $\mathcal{B}(x^+)$, $\mathcal{B}(x^-)$ respectively up to a small number $\eta > 0$. Write

$$\varphi^{\pm} = f^{\pm} + \sum_{I\subseteq I_0,\, I\in D} (\varphi, h_I)h_I\,, \qquad \psi^{\pm} = g^{\pm} + \sum_{I\subseteq I_0,\, I\in D} \varepsilon_I(\varphi, h_I)h_I\,,$$

Consider

$$\varphi(t) := \begin{cases} \varphi^+(2t-1)\,, & \text{if } t \in [\frac{1}{2}, 1] \\ \varphi^-(2t)\,, & \text{if } t \in [0, \frac{1}{2}). \end{cases}$$

and

$$\psi(t) := \begin{cases} \psi^+(2t-1)\,, & \text{if } t \in [\frac{1}{2}, 1] \\ \psi^-(2t)\,, & \text{if } t \in [0, \frac{1}{2}) \end{cases}$$

Since $|x_1^+ - x_1^-| = |x_2^+ - x_2^-|$, the function ψ is a martingale transform of φ, and the pair (φ, ψ) is an admissible pair of the test functions corresponding to the point x. Therefore,

$$\begin{aligned} \mathcal{B}(x) &\geq \frac{1}{|I_0|}\big|\{t \in I_0 \colon \psi(t) \geq 0\}\big| \\ &= \frac{1}{2|I_0^+|}\big|\{t \in [\frac{1}{2}, 1] \colon \psi(t) \geq 0\}\big| + \frac{1}{2|I_0^-|}\big|\{t \in [0, \frac{1}{2}) \colon \psi(t) \geq 0\}\big| \\ &\geq \frac{1}{2}\mathcal{B}(x^+) + \frac{1}{2}\mathcal{B}(x^-) - 2\eta. \end{aligned}$$

Since this inequality holds for an arbitrary small η, we can pass to the limit $\eta \to 0$, what gives us the required assertion. □

Corollary 3.3.12. *The lemma means that if we change variables $f = y_1 - y_2$, $g = -y_1 - y_2$, and introduce a function $M(y_1, y_2, F) := B(g, f, F)$ defined in the domain $G := \{y = (y_1, y_2, F) \in \mathbb{R}^3 \colon |y_1 - y_2| \leq F\}$, then we get that for each fixed y_2, $M(F, y_1, \cdot)$ is concave and for each fixed y_1, $M(F, \cdot, y_2)$ is concave.*

3.3.4.1 The boundary $F = y_1 - y_2$

We start with considering a boundary case $F = f$ or, in the y variables, $F = y_1 - y_2$. It means that we consider only non-negative functions φ. By the homogeneity of the function M we need to find a function S of variable $s = \frac{y_1}{y_2}$, such that

$$\left(S(\frac{y_1}{y_2})\right)''_{y_1 y_1} \leqslant 0, \text{ and } \left(S(\frac{y_1}{y_2})\right)''_{y_2 y_2} \leqslant 0. \tag{3.24}$$

We notice that when $g \to 0$ we have $s \to -1$ and we must have $S \to 0$. Thus, we get a condition

$$S(s) \to 0, \quad \text{as} \quad s \to -1. \tag{3.25}$$

Moreover, we have seen that if $f \geqslant -g$ then $\mathcal{B}(g, f, F) = 1$. In particular, it holds when $f = -g$. Therefore, we have $M(y_1, -y_1, 0) = 1$. This implies that

$$S(s) \to 1, \quad \text{as} \quad s \to -\infty.$$

From inequalities (3.24) we get that

$$S''(s) \leqslant 0, \quad , s^2 S''(s) + 2S'(s) \leqslant 0, \quad s \in (-\infty, -1].$$

Make the second inequality an equation (we are looking for the **best** nontrivial S). We get

$$S(s) = c_1 + \frac{c_2}{s}.$$

The boundary conditions imply that

$$S(s) = 1 - \frac{1}{s},$$

and therefore

$$M(y_1, y_2, y_1 - y_2) = 1 - \frac{y_2}{y_1} = \frac{y_1 - y_2}{y_1},$$

or

$$B(g, f, f) = \frac{2f}{f - g}.$$

Thus, we get an answer

$$M(y_1, y_2, y_1 - y_2) = \begin{cases} 1, & y_2 \leqslant 0 \\ \frac{y_1 - y_2}{y_1}, & y_2 \geqslant 0, \end{cases} \tag{3.26}$$

or

$$B(g, f, f) = \begin{cases} 1, & f \geqslant -g \\ \frac{2f}{f-g}, & f \leqslant -g. \end{cases}$$

3.3.4.2 The domain Ω

We remind the reader that for a fixed y_1 the function M is concave in variables (F, y_2). We also remind the symmetry condition, i.e.

$$M(y_1, y_2, F) = M(y_2, y_1, F).$$

Let us differentiate this equation in y_2 and set $y_2 = y_1$. Then we get an equation:

$$M_{y_1}(y_1, y_1, F) = M_{y_2}(y_1, y_1, F).$$

Moreover, due to the symmetry it is enough to find M for $y_2 \leqslant y_1$. As before, we saw that for $f \geqslant -g$ we have $B(g, f, F) = 1$, i.e.

$$\text{for } y_2 \leqslant 0, \text{ we have } M(y_1, y_2, F) = 1. \tag{3.27}$$

Thus, it is enough to consider the case $0 \leqslant y_2 \leqslant y_1$. Denote $\Omega_{y_1} = \{(y_2, F) \colon F \geqslant |y_2 - y_1|\}$ — the section of Ω for fixed y_1. We want to find M satisfying concavity in this hyperplane–we are going to look for M (and we will check later that it is concave) that solves Monge–Ampère (MA) equation in Ω_{y_1} with boundary conditions (3.26) and (3.27). In Ω_{y_1}, there is a point $P := (0, y_1, y_1)$. Let us make a guess that the characteristics (and we know by Pogorelov's theorem that they form the foliation of Ω_{y_1} by straight lines) of our MA equation in Ω_{y_1} form the fan of lines with common point $P = (y_1, y_1, 0)$. By Pogorelov's theorem we also know that there exists functions t_1, t_2, t constant on characteristics such that

$$M = t_1 F + t_2 y_2 + t\,, \tag{3.28}$$

such that $t_1 = t_1(t; y_1), t_2 = t_2(t; y_1)$ (we think that y_1 is a parameter), that

$$0 = (t_1)'_t F + (t_2)'_t y_2 + 1\,, \tag{3.29}$$

that

$$t_1 = \frac{\partial M(\cdot, y_2, F)}{\partial F}\,,\, t_1 = \frac{\partial M(\cdot, y_2, F)}{\partial y_2}\,. \tag{3.30}$$

Let us call characteristics L_t. Extend one of them from P till $y_2 = y_1$. We recall another boundary condition:

$$\text{If } y_2 = y_1 \Rightarrow \frac{\partial M}{\partial y_2} = \frac{\partial M}{\partial y_1}\,. \tag{3.31}$$

Or if we denote the intersection of L_t with $y_2 = y_1$ by $(y_1, y_1, F(t))$ we get

$$t_2(t; y_1) = \frac{\partial M}{\partial y_1}(y_1, y_1, F(t))\,. \tag{3.32}$$

We want to prove now that

$$\text{On the whole } L_t \text{ we have } F(t)t_1 + 2y_1t_2 = 0\,. \tag{3.33}$$

In fact, our M is 0 homogeneous. So everywhere $FM'_F + y_1M'_{y_1} + y_2M'_{y_2} = 0$. Apply this to point $(y_1, y_1, F(t))$, where we can use (3.32) and get $F(t)t_1 + t_2y_1 + t_2y_1 = 0$, which is (3.33) in one point. But then all entries are constants on L_t, therefore, (3.33) follows.

Now use our guess that L_t fan from $P = (y_1, y_1, 0)$. Plug this coordinates into $0 = (t_1)'_tF + (t_2)'_ty_2 + 1$, which is (3.29). Then we get the crucial (and trivial) ODE

$$t_1'(t) = -\frac{1}{y_1} \Rightarrow t_1(t) = -\frac{1}{y_1}t + C_1(y_1)\,. \tag{3.34}$$

Let boundary line $F = y_1 - u$ corresponds to $t = t_0$. Then we use (3.28) and (3.26):

$$(-\frac{1}{y_1}t_0 + C_1(y_1))(y_1 - u) + t_2u + t_0 = 1 - \frac{u}{y_1}\,.$$

Using (3.33) we can plug t_2 expressed via $F(t)$. But by definition $F(t_0) = 0$. So we get

$$(-\frac{1}{y_1}t_0 + C_1(y_1))(y_1 - u) + t_0 = 1 - \frac{u}{y_1}\,.$$

Or

$$C_1(y_1)y_1 - (t_0 + C_1(y_1)y_1)\frac{u}{y_1} = 1 - \frac{u}{y_1}\,.$$

Varying u we get $C_1(y_1) = \frac{1}{y_1}$, $t_0 = 0$. Now from (3.34) we get

$$t_1(t) = \frac{1}{y_1}(1 - t)\,. \tag{3.35}$$

After that (3.29) and (3.33) become the system of two linear "ODE"s on $F(t)$ and $t_2(t)$:

$$\begin{cases} -\frac{1}{y_1}F(t) + y_1 t_2'(t) + 1 = 0 \\ 2y_1 t_2(t) + F(t)\frac{1}{y_1}(1-t) = 0\,. \end{cases} \tag{3.36}$$

We find $t_2 = -\frac{1}{y_1}(1-t)t$. We find the arbitrary constant for t_2 by noticing that the second equation of (3.36) at $t_0 = 0$ implies that $t_2(0) = 0$ as $F(t_0) = F(0) = 0$ by definition.

Hence (3.29) becomes

$$-\frac{1}{y_1}F + \frac{1}{y_1}(2t-1)y_2 + 1 = 0\,. \tag{3.37}$$

Given $(y_1, y_2, F) \in \Omega_{y_1} \cap \{0 \leq y_2 \leq y_1\}$, we find t from (3.37) and plug it into (3.28), in which we know already $t_(t)$ and $t_2(t)$. Namely, we know that

$$M(y_1, y_2, F) = \frac{1}{y_1}F - \frac{1}{y_1}t(1-t)y_2 + t\,. \tag{3.38}$$

Plugging $t = \frac{1}{2}\frac{F-(y_1-y_2)}{y_2}$ from (3.37) into this equation we finally obtain

$$M(y_1, y_2, F) = 1 - \frac{(F - y_1 - y_2)^2}{4y_1 y_2}. \tag{3.39}$$

We notice that on the line $F = y_2 + y_1$ we get $M = 1$. Thus, we get the following answer for M:

$$M(y_1, y_2, F) = \begin{cases} 1 - \frac{(F-y_1-y_2)^2}{4y_1y_2}, & F \leqslant y_1 + y_2 \\ 1, & F \geqslant y_1 + y_2. \end{cases} \tag{3.40}$$

In our initial coordinates we get

$$B(g,f,F)=\begin{cases}1-\frac{(F+g)^2}{g^2-f^2}, & F\leqslant -g\\ \\ 1, & F\geqslant -g.\end{cases}$$

3.4 The weighted estimate

3.4.1 Differential properties of $\mathcal{B}$ translated to differential properties of B

It is convenient to introduce an auxiliary functions of 4 and 3 variables:

$$\widetilde{B}(x,y,f,\lambda):=B(\frac{x}{\lambda},y,\frac{f}{\lambda})\,.$$

Of course

$$\mathcal{B}(F,w,m,f,\lambda)=m\widetilde{B}(\frac{F}{m},\frac{w}{m},f,\lambda)=mB(\frac{F}{m\lambda},\frac{w}{m},\frac{f}{\lambda})\,. \tag{3.41}$$

Lemma 3.4.1. *Function B increases in the first and in the second variable.*

Proof. We know that by definition the RHS of (3.41) is getting bigger if λ is getting smaller. So let us consider $\lambda_1>\lambda_2, \lambda_1=\lambda_2+\delta$, and variables F,w,m,f fixed, and choose ϕ_1 (and a weight ω), $\langle\phi_1\rangle=f+\varepsilon, \langle|\phi_1|\omega\rangle=F$, which almost realizes the supremum $\mathcal{B}(F,w,m,f+\varepsilon,\lambda_1)$. Consider ϕ_2 such that $\phi_2=\phi_1-h$. Function h will be chosen later, however we say now that h is equal to a certain constant a on a small dyadic interval ℓ and is zero otherwise. Constant a and interval ℓ we will chose later. But $\varepsilon:=\langle h\rangle$ will be chosen very soon. Function ϕ_2 competes for supremizing $\mathcal{B}$ at $(\langle|\phi_2|\omega\rangle,w,m,f,\lambda_2)$. We choose ε in such a way that

$$\frac{\langle\phi_1\rangle}{\lambda_1} = \frac{f+\varepsilon}{\lambda_1} = \frac{f}{\lambda_1-\delta} = \frac{\langle\phi_2\rangle}{\lambda_2}\,. \tag{3.42}$$

Let us prove that (3.42) implies that

$$\frac{\langle|\phi_1|\omega\rangle}{\lambda_1} \le \frac{\langle|\phi_2|\omega\rangle}{\lambda_2}\,. \tag{3.43}$$

By (3.42) this is the same as

$$\frac{\langle|\phi_2+h|\omega\rangle}{\langle|\phi_2|\omega\rangle} \le \frac{\langle\phi_1\rangle}{\langle\phi_2\rangle} = \frac{\langle\phi_2\rangle+\varepsilon}{\langle\phi_2\rangle}\,.$$

The previous inequality becomes

$$\frac{\langle|\phi_2+h|\omega\rangle}{\langle|\phi_2|\omega\rangle} \le 1 + \frac{\langle h\rangle}{\langle\phi_2\rangle}\,.$$

By triangle inequality the latter inequality would follow from the following one

$$\langle|\phi_2|\,\omega\rangle \ge \langle\phi_2\rangle\frac{\langle|h|\,\omega\rangle}{\langle h\rangle}\,.$$

We can think that the minimum m of ω is attained on a whole tiny dyadic interval ℓ (we are talking about *almost* supremums). Put h to be a certain $a>0$ on this interval and zero otherwise. Of course we choose a to have $\langle h\rangle = \varepsilon$, where ε was chosen before. Now the previous display inequality becomes

$$\langle|\phi_2|\,\omega\rangle \ge \langle\phi_2\rangle\cdot m\,,$$

which is obvious.

Notice that $\mathcal{B}(\langle|\phi_2|\rangle, w, m, f, \lambda_2)$ as a supremum is larger than the ω-measure of the level set $> \lambda_2$ of the martingale transform of ϕ_2. But this is also the martingale transform of ϕ_1. The λ_1-level set for any martingale transform of ϕ_1 is smaller, as $\lambda_1 > \lambda_2$. But recall that we already said that ϕ_1 (and weight ω) almost realizes its own supremum $\mathcal{B}(F, w, m, f+\varepsilon, \lambda_1) = \mathcal{B}(\langle|\phi_1|\rangle, w, m, \langle\phi_1\rangle, \lambda_1)$ So

$$\mathcal{B}(\langle|\phi_1|\rangle, w, m, \langle\phi_1\rangle, \lambda_1) \leq \mathcal{B}(\langle|\phi_2|\rangle, w, m, \langle\phi_2\rangle, \lambda_2)\,.$$

In other notations we get

$$B(\frac{\langle|\phi_1|\rangle}{m\lambda_1}, \frac{w}{m}, \frac{\langle\phi_1\rangle}{\lambda_1}) \leq B(\frac{\langle|\phi_2|\rangle}{m\lambda_2}, \frac{w}{m}, \frac{\langle\phi_2\rangle}{\lambda_2})\,.$$

Let us denote the argument on the LHS as (x_1, y_1, z_1), and on the RHS as (x_2, y_2, z_2). Notice that $y_1 = y_2 =: y$ trivially and $z_1 = z_2 =: z$ by (3.42). Notice also that $x_1 < x_2$ by (3.43). Moreover by choosing δ very small we can realize any $x_1 < x_2$ as close to x_2 as we want. Then the last display inequality reads as

$$B(x_1, y, z) \leq B(x_2, y, z)\,.$$

So we proved that function B increases in the first variable.

The increase in the second variable is easy. Choose a dyadic interval I on which $\inf_I \omega > m$, but $\langle\omega\rangle_I / \inf_I \omega < Q =: [\omega]_{A_1}$. For non-constant ω this is always possible, just take a small interval containing a point x_0, where $\omega(x_0) > m$. Then augment ω on I slightly to get ω_1 with $\langle\omega_1\rangle = w + \varepsilon$. It is easy to see that as a result we have the new weight with the A_1 norm at most Q, the same global infimum m but a larger global average $\langle\omega\rangle$. The ω_1 measure of the level set of the martingale transform will be bigger than ω measure of the same level set of the same martingale transform, and w/m also grows to $(w+\varepsilon)/m$. All other variables stay the same. So if the original ω (and some ϕ) were (almost) realizing supremum, we would get

$$B(x, y_1, z) \le B(x, y_2, z)$$

for $y_1 = w/m, y_2 = (w+\varepsilon)/m$. □

Theorem 3.4.2. *Function B from* (3.3) *satisfies*

$$t \to t^{-1} B(\alpha t, \beta t, \gamma) \textit{ is increasing for } \frac{|\gamma|}{\alpha} \le t \le \frac{Q}{\beta}\,. \tag{3.44}$$

$$B \textit{ is concave}\,. \tag{3.45}$$

$$B(\frac{x}{\lambda}, y, \frac{f}{\lambda}) - \frac{1}{4}\Big[B(\frac{x-dx}{\lambda-d\lambda}, y-dy, \frac{f-d\lambda}{\lambda-d\lambda}) + B(\frac{x-dx}{\lambda-d\lambda}, y-dy, \frac{f+d\lambda}{\lambda-d\lambda}) +$$
$$B(\frac{x+dx}{\lambda+d\lambda}, y+dy, \frac{f-d\lambda}{\lambda+d\lambda}) + B(\frac{x+dx}{\lambda+d\lambda}, y+dy, \frac{f+d\lambda}{\lambda+d\lambda})\Big] \ge 0\,. \tag{3.46}$$

Proof. These relations follow from Theorem 3.2.3, Theorem 3.2.2, and Theorem 3.2.1 (actually from (3.7)) correspondingly. □

We can choose extremely small ε_0 and inside the domain Ω we can mollify $\mathcal{B}$ by a convolution of it with ε_0-bell function ψ supported in a ball of radius $\varepsilon_0/10$.

Multiplicative convolution can be viewed as the integration with $\frac{1}{\delta^5}\psi(\frac{x-x_0}{\delta})$, where $\delta = \varepsilon_0/10$. Here x_0 is a point inside the domain of definition Ω for function $\mathcal{B}$.

This new function we call $\mathcal{B}$ again. It is exactly as the initial function $\mathcal{B}$, and it obviously satisfies all the same relationships, in particular it satisfies Theorems 3.2.1, 3.2.2, 3.2.3. Only its domain of definitionΩ_{ε_0} is smaller (slightly) than Ω. The advantage however is that the new $\mathcal{B}$ is smooth. We build B by this new $\mathcal{B}$. A new function B defined by the new $\mathcal{B}$ as in (3.41) will be smooth. Actually the new B should be denoted B^{ϵ_0}, where superscript denotes our operation of mollification, but we drop the superscript for the sake of brevity. In fact, all these mollifications are for the sake of convenience, the new functions satisfy the old inequalities in the uniform way, independently of ε_0. Property (3.46) can be now rewritten by the use of Taylor's formula:

Theorem 3.4.3.

$$-\alpha^2 B_{\alpha\alpha}\left(\frac{dx}{x}-\frac{d\lambda}{\lambda}\right)^2-\beta^2 B_{\beta\beta}\left(\frac{dy}{y}\right)^2-(1+\gamma^2)B_{\gamma\gamma}\left(\frac{d\lambda}{\lambda}\right)^2-$$

$$-2\alpha\beta B_{\alpha\beta}\left(\frac{dx}{x}-\frac{d\lambda}{\lambda}\right)\frac{dy}{y}+2\beta\gamma B_{\beta\gamma}\frac{dy}{y}\frac{d\lambda}{\lambda}+2\alpha\gamma B_{\alpha\gamma}\left(\frac{dx}{x}-\frac{d\lambda}{\lambda}\right)\frac{d\lambda}{\lambda}+$$

$$+2\alpha B_{\alpha}\left(\frac{dx}{x}-\frac{d\lambda}{\lambda}\right)\frac{d\lambda}{\lambda}-2\gamma B_{\gamma}\left(\frac{d\lambda}{\lambda}\right)^2\geq 0\,.$$

Proof. This is just Taylor's formula applied to (3.46). □

Denoting

$$\xi=\frac{dx}{x}=\frac{dy}{y}\,,\ \eta=\frac{d\lambda}{\lambda}$$

we obtain the following quadratic form inequality

Theorem 3.4.4.

$$-\xi^2\left[\alpha^2 B_{\alpha\alpha}+\beta^2 B_{\beta\beta}+2\alpha\beta B_{\alpha\beta}\right]-\eta^2\left[\alpha^2 B_{\alpha\alpha}+(1+\gamma^2)B_{\gamma\gamma}+2\alpha\gamma B_{\alpha\gamma}+2\alpha B_{\alpha}+2\gamma B_{\gamma}\right]+$$

$$+2\xi\eta\left[\alpha^2 B_{\alpha\alpha}+\alpha\beta B_{\alpha\beta}+\beta\gamma B_{\beta\gamma}+\alpha\gamma B_{\alpha\gamma}+\alpha B_{\alpha}\right]\geq 0\,.$$

Now let us combine Theorem 3.4.4 and Theorem 3.2.2. In fact, Theorem 3.2.2 implies

$$-2\alpha\gamma B_{\alpha\gamma}\eta^2 \le -\alpha^2\gamma B_{\alpha\alpha}\eta^2 - \gamma B_{\gamma\gamma}\eta^2\,.$$

We plug it into the second term above. Also Theorem 3.2.2 implies

$$2\alpha\gamma B_{\alpha\gamma}\xi\eta \le -\alpha^2\gamma B_{\alpha\alpha}\xi^2 - \gamma B_{\gamma\gamma}\eta^2\,,$$

$$2\beta\gamma B_{\beta\gamma}\xi\eta \le -\beta^2\gamma B_{\beta\beta}\xi^2 - \gamma B_{\gamma\gamma}\eta^2\,,$$

We will plug it into the third term above. Then using the notation

$$\psi(\alpha,\beta,\gamma) := -\alpha^2 B_{\alpha\alpha} - 2\alpha\beta B_{\alpha\beta} - \beta^2 B_{\beta\beta}$$

(which is non-negative by the concavity of B in its first two variables by the way) we introduce the notations

$$K := \psi(\alpha,\beta,\gamma) + (-\alpha^2 B_{\alpha\alpha} - \beta^2 B_{\beta\beta})\gamma\,,$$

$$L := -\psi(\alpha,\beta,\gamma) + (\alpha^2 B_\alpha)_\alpha - \beta^2 B_{\beta\beta}\,,$$

$$N := -(1+3\gamma+\gamma^2)B_{\gamma\gamma} - 2\gamma B_\gamma - (\alpha^2 B_\alpha)_\alpha - \alpha^2 B_{\alpha\alpha}\gamma\,.$$

And we get that the following quadratic form is non-negative:

$$\xi^2\,K + \xi\eta\,L + \eta^2\,N :=$$

$$\xi^2\,[\psi(\alpha,\beta,\gamma) + (-\alpha^2 B_{\alpha\alpha} - \beta^2 B_{\beta\beta})\gamma] +$$

$$\xi\eta\,[-\psi(\alpha,\beta,\gamma)+(\alpha^2B_\alpha)_\alpha-\beta^2B_{\beta\beta}]+$$

$$\eta^2\,[-(1+3\gamma+\gamma^2)B_{\gamma\gamma}-2\gamma B_\gamma-(\alpha^2B_\alpha)_\alpha-\alpha^2B_{\alpha\alpha}\gamma]\geq 0\,.$$

Therefore, K is positive, and

$$N\geq\frac{L^2}{4K}\,. \tag{3.47}$$

Now we will estimate L from below, K from above and as a result we will obtain the estimate of N from below, which will bring us our proof.

But first we need some a priori estimates, and for that we will need to mollify $B=B^{\epsilon_0}$ in variables α,β. Again we make a multiplicative convolution with a bell-type function. Let us explain why we need it. Let

$$\hat{Q}:=\sup_G B/\alpha\,.$$

We want to prove that

$$\hat{Q}/Q\to\infty\,. \tag{3.48}$$

First we need to notice that

$$\int_{1/2}^1\psi(\alpha t,\beta t,\gamma)\,dt\leq C\,(\hat{Q}\gamma+\frac{\hat{Q}}{Q}\alpha),\;\;\psi(\alpha,\beta,\gamma):=-\alpha^2B_{\alpha\alpha}-2\alpha\beta B_{\alpha\beta}-\beta^2B_{\beta\beta}\,. \tag{3.49}$$

In fact, consider $\beta\in[Q/4,Q/2]$, $b(t):=B(\alpha t,\beta t,\gamma)$ on the interval $\frac{|\gamma|}{\alpha}=:t_0\leq t\leq 1$. Let $\ell(t)=b(1)t\leq\hat{Q}t\alpha$. We saw that $b(t)/t$ is increasing and b is concave, and b is under ℓ, and so by elementary picture of concave function having property $b(\cdot)/\cdot$ increasing and $b(\cdot)$ concave on the interval $[t_0',1]$ we get that the maximum of $\ell(\cdot)-b(\cdot)$ is attained on the left end-point. The left end-point t_0' is the maximum of $t_0=|\gamma|/\alpha$ and $1/\beta$ which is c/Q.

Therefore,

$$\ell(t) - b(t)|(t = (\max(\frac{\gamma}{\alpha}, \frac{c}{Q})) \leq \ell(\max(\frac{\gamma}{\alpha}, \frac{c}{Q})) \leq C\hat{Q}\alpha \max(\frac{\gamma}{\alpha}, \frac{1}{Q}) \leq \hat{Q}\gamma + \frac{\hat{Q}}{Q}\alpha\,,$$

and the above value is maximum of $g(t) := \ell(t) - b(t)$ on $[t_0', 1]$. By the same property that $b(t)/t$ is increasing we get that

$$g'(1) = \ell'(1) - b'(1) = b(1) - b'(1) \leq 0\,.$$

Combining this with Taylor's formula on $[t_0, 1]$ we get for $g := \ell - b$ (g is convex of course):

$$-(1-t_0)g'(1) + \int_{t_0}^1 dt \int_t^1 g''(s)ds = \text{positive} + \int_{t_0}^1 (s-t_0)g''(s)ds \leq \sup g \leq \hat{Q}\gamma + \frac{\hat{Q}}{Q}\alpha\,. \tag{3.50}$$

This implies (3.49) because $g''(t) = \frac{1}{t^2}\psi(\alpha t, \beta t, \gamma)$, $t \in [1/2, 1]$.

Consider now function $a(t) := B(\alpha t, \beta, \gamma)$ We also have the same type of consideration applied to convex function $\hat{Q}\alpha - a(t)$ bringing us

$$\int_{1/2}^1 -\alpha^2 B_{\alpha\alpha}(\alpha t, \beta, \gamma)\, dt \leq C\hat{Q}\alpha\,. \tag{3.51}$$

Similarly,

$$\int_{1/2}^1 -\beta^2 B_{\beta\beta}(\alpha, \beta t, \gamma)\, dt \leq C\hat{Q}\alpha\,. \tag{3.52}$$

We used here that $B_\alpha \geq 0, B_\beta \geq 0$, which is not difficult to see.

For the future estimates we want (3.49), (3.51), (3.52) to hold not in average but pointwise.

To achieve the replacement of "in-average" estimates (3.49), (3.51), (3.52) by their pointwise analogs let us consider yet another mollification, now it is of B:

$$B_{new}(\alpha, \beta, \gamma) := 2 \int_{1/2}^{1} B(\alpha t, \beta t, \gamma)\, dt.$$

The domain of definition of B_{new} is only in tiny difference with the domain of definition of B. In fact, the latter is $\{(\alpha, \beta, \gamma) : |\gamma| \le \alpha, 1 \le \beta \le Q\}$, and the former is just $G := \{(\alpha, \beta, \gamma) : |\gamma| \le \frac{1}{2}\alpha, 2 \le \beta \le Q\}$.

If we replace (α, β, γ) by $(\alpha t, \beta t, \gamma), 1/2 \le t \le 1$, everywhere in the inequality of Theorem 3.4.4, and then integrate the inequality with $2 \int_{1/2}^{1} \dots dt$, we will get Theorem 3.4.4 but for B_{new}.

It is not difficult to see that (3.49) becomes a pointwise estimate for B_{new} (just differentiate the formula for B_{new} in α, β, γ and multiply by α, β, γ appropriately):

$$-\alpha^2 (B_{new})_{\alpha\alpha} - 2\alpha\beta (B_{new})_{\alpha\beta} - \beta^2 (B_{new})_{\beta\beta} \le C(\hat{Q}\gamma + \frac{\hat{Q}}{Q}\alpha). \tag{3.53}$$

This pointwise estimate automatically imply new "average" estimate:

$$2 \int_{1/2}^{1} \big(- \alpha^2 s^2 (B_{new})_{\alpha\alpha}(\alpha s, \beta, \gamma) - 2\alpha s \beta (B_{new})_{\alpha\beta}(., \beta, .) - \beta^2 (B_{new})_{\beta\beta} \big) \le C(\hat{Q}\gamma + \frac{\hat{Q}}{Q}\alpha).$$

This means exactly that the function

$$\tilde{B} := (B_{new})_{new} := 2 \int_{1/2}^{1} B(\alpha s, \beta, \gamma)\, ds$$

still satisfies (3.53). It also clearly satisfies the inequality of Theorem 3.4.4 because (as we noticed above) B_{new} satisfies this inequality. To see this fact just replace all α's in the inequality of Theorem 3.4.4 applied to B_{new} by αs and integrate $2\int_{1/2}^{1} \dots \, ds$.

Now let us see that $\tilde{B} = (B_{new})_{new}$ also satisfies a pointwise analog of (3.51), namely, that

$$-\alpha^2 \tilde{B}_{\alpha\alpha}(\alpha, \beta, \gamma) \le C\hat{Q}\alpha \,. \tag{3.54}$$

To show (3.54) we just repeat what has been done above. Let $\tilde{g}(t) := \hat{Q}\alpha - B_{new}(\alpha t, \beta, \gamma)$. Then we have: 1) $0 \le \tilde{g} \le \hat{Q}\alpha$ on $[t_0, 1]$, 2) $\tilde{g}'(1) \le 0$ (we saw that B, and hence B_{new}, are increasing in the first argument), 3) $\tilde{g}$ is convex. Then we saw in (3.50) that

$$\int_{1/2}^{1} s^2 \, \tilde{g}''(s) \, ds \le \int_{1/2}^{1} \tilde{g}''(s) \, ds \le C\hat{Q}\alpha.$$

But this is exactly (3.54).

So far we constructed a function $\tilde{B} = (B_{new})_{new}$ that satisfies pointwise inequalities (3.53), (3.54) and the inequality of Theorem 3.4.4. We are left to see that by introducing

$$\hat{B} := 2\int_{1/2}^{1} \tilde{B}(\alpha, \beta s, \gamma) \, ds$$

we keep (3.53), (3.54) and the inequality of Theorem 3.4.4 valid and also ensure

$$-\beta^2 \hat{B}_{\beta\beta}(\alpha, \beta, \gamma) \le C\hat{Q}\alpha \,. \tag{3.55}$$

W already just saw that (3.53), (3.54) and the inequality of Theorem 3.4.4 are valid for $\hat{B}$ just by averaging the same inequalities for $\tilde{B}$. We can see that (3.55) holds by the repetition of what has been just done. Namely, consider $\hat{g}(t) := \hat{Q}\alpha - \tilde{B}(\alpha, \beta t, \gamma)$. Then we have: 1)

$0 \le \hat{g} \le \hat{Q}\alpha$ on $[t_0, 1]$, 2) $\hat{g}'(1) \le 0$ (we saw that B, and hence B_{new}, $\tilde{B}$ are increasing in the first argument), 3) $\hat{g}$ is convex. Using (3.50) again in exactly the same manner as we did with proving (3.54) we get

$$\int_{1/2}^{1} s^2\, \hat{g}''(s)\, ds \le \int_{1/2}^{1} \hat{g}''(s)\, ds \le C\hat{Q}\alpha.$$

But this is exactly (3.55).

We drop "hat", and from now on $\hat{B}$ is just denoted by B. We can summarize its properties as follows.

$$0 \le \psi(\alpha,\beta,\gamma) \le C(\hat{Q}\gamma + \frac{\hat{Q}}{Q}\alpha)\,. \tag{3.56}$$

$$0 \le -\alpha^2 B_{\alpha\alpha}(\alpha,\beta,\gamma) \le C\hat{Q}\alpha\,. \tag{3.57}$$

$$0 \le -\beta^2 B_{\beta\beta}(\alpha,\beta,\gamma) \le C\hat{Q}\alpha\,. \tag{3.58}$$

Recall that (now with this mollified B):

$$\xi^2\, K + \xi\eta\, L + \eta^2\, N :=$$

$$\xi^2\, [\psi(\alpha,\beta,\gamma) + (-\alpha^2 B_{\alpha\alpha} - \beta^2 B_{\beta\beta})\gamma]$$

$$\xi\eta\, [-\psi(\alpha,\beta,\gamma) + (\alpha^2 B_\alpha)_\alpha - \beta^2 B_{\beta\beta}]$$

$$\eta^2\, [-(1 + 3\gamma + \gamma^2)B_{\gamma\gamma} - 2\gamma B_\gamma - (\alpha^2 B_\alpha)_\alpha - \alpha^2 B_{\alpha\alpha}\gamma] \ge 0\,.$$

We will choose soon appropriate $\alpha_0, \alpha_1 \le \frac{1}{100}\alpha_0$ and $\gamma \le \tau\alpha_0$ with some small τ. Let us introduce

$$k := \int_{\alpha_1}^{\alpha_0} K\, d\alpha = \int_{\alpha_1}^{\alpha_0} [\psi(\alpha,\beta,\gamma) + (-\alpha^2 B_{\alpha\alpha} - \beta^2 B_{\beta\beta})\gamma]\, d\alpha\,,$$

$$n := \int_{\alpha_1}^{\alpha_0} N\, d\alpha = \int_{\alpha_1}^{\alpha_0} [-(1+3\gamma+\gamma^2)B_{\gamma\gamma} - 2\gamma B_\gamma - (\alpha^2 B_\alpha)_\alpha - \alpha^2 B_{\alpha\alpha}\gamma]\, d\alpha\,,$$

$$\ell := \int_{\alpha_1}^{\alpha_0} [-\psi(\alpha,\beta,\gamma) + (\alpha^2 B_\alpha)_\alpha - \beta^2 B_{\beta\beta}]\, d\alpha\,.$$

Estimate of k from above. The integrand of k is obviously positive and ψ term dominates other terms (by (3.56), (3.57), (3.58) and the smallness of γ). Therefore,

$$0 \le k \le C_1\,(\hat{Q}\gamma\alpha_0 + C\frac{\hat{Q}}{Q}\alpha_0^2) + C_2\,\hat{Q}\gamma\alpha_0^2 \le C\,(\hat{Q}\gamma\alpha_0 + C\frac{\hat{Q}}{Q}\alpha_0^2)\,, \tag{3.59}$$

if Q is very large. We choose (we are sorry for a strange way of writing α_0, why we do that will be seen in the next section)

$$\alpha_0 = c\left(\frac{Q}{\hat{Q}}\right)^\rho,\ \rho = 1\,,\ \alpha_1 = \frac{1}{100}\sqrt{\frac{Q}{\hat{Q}}}\alpha_0\,. \tag{3.60}$$

Here c is a small positive constant. We also choose to have γ running only on the following interval

$$\gamma \in [0,\gamma_0]\,,\ \gamma_0 := \tau\left(\frac{Q}{\hat{Q}}\right)^\rho \alpha_0\,,\ \rho = 1\,, \tag{3.61}$$

where τ is a small positive constant.

Estimate of ℓ from below. Estimating from below we can skip the non-negative term $-\beta^2 B_{\beta\beta}$. Also

$$\int_{\alpha_1}^{\alpha_0} -\psi(\alpha,\beta,\gamma) \ge -C\hat{Q}\gamma\alpha_0 - C\frac{\hat{Q}}{Q}\alpha_0^2\,.$$

On the other hand,

$$\int_{\alpha_1}^{\alpha_0} (\alpha^2 B_\alpha)_\alpha \, d\alpha \geq \alpha_0^2 B_\alpha(\alpha_0, \beta, \gamma) - \alpha_1^2 \hat{Q} \,,$$

as mollification gives a pointwise estimate

$$B_\alpha \leq C\hat{Q} \,. \tag{3.62}$$

Recall that $\beta \in [Q/4, Q/2]$. We also will prove soon the obstacle condition (3.74), which says that

$$B(1, \beta, \gamma) \geq \frac{\beta}{8} \,. \tag{3.63}$$

If $B_\alpha(\alpha_0, \beta, \gamma)$ would be smaller than $Q/40$ (and then $B_\alpha(s, \beta, \gamma) \leq Q/40$ for all $s \in [\alpha_0, 1]$ by concavity of B in its first variable) we would not be able to reach at least $\frac{Q}{4 \cdot 8}$. In fact, by our choice of α_0 in (3.60) we have

$$B(\alpha_0, \beta, \gamma) \leq \hat{Q}\alpha_0 \leq c\, Q \,. \tag{3.64}$$

If $B_\alpha(\alpha_0, \beta, \gamma) \leq \frac{Q}{40}$, and so this derivative $B_\alpha(s, \beta, \gamma) \leq \frac{Q}{40}$ on $s \in [\alpha_0, 1]$ (concavity), we cannot reach $Q/(4 \cdot 8)$ for $s = 1$ if we start with value of B in (3.64) at $s = \alpha_0$. But the fact that we cannot reach $Q/(4 \cdot 8)$ contradicts to (3.63). Therefore,

$$B_\alpha(\alpha_0, \beta, \gamma) \geq \frac{Q}{40} \,, \tag{3.65}$$

and

$$\ell \geq \frac{\alpha_0^2}{40} Q - \alpha_1^2 \hat{Q} - C\, \hat{Q} \gamma \alpha_0 - C \frac{\hat{Q}}{Q} \alpha_0^2 \,. \tag{3.66}$$

As $\alpha_1 = \frac{1}{100}\alpha_0\sqrt{\frac{Q}{\hat{Q}}}$ (see (3.60)), the second term is dominated by the first; the third term is dominated by the first because of the choice of γ_0 in (3.61), the fourth term is dominated by the first one because $Q^2 >> \hat{Q}$, see [P] for a much better estimate.

Finally,

$$\ell \geq \frac{\alpha_0^2}{80} Q \geq c\,\alpha_0^2\, Q\,. \tag{3.67}$$

And k is

$$0 \leq k \leq C\,(\hat{Q}\gamma\alpha_0 + C\frac{\hat{Q}}{Q}\alpha_0^2) = \alpha_0\hat{Q}\,(\gamma + \frac{1}{Q}\alpha_0)\,.$$

We got

$$n \geq \frac{\ell^2}{4k} \geq c\,\frac{\alpha_0^4 Q^2}{\alpha_0\hat{Q}\,(\gamma + \frac{1}{Q}\alpha_0)}\,. \tag{3.68}$$

Estimate of n from above. By (3.65), (3.62) and (3.57) we get

$$\int_{\alpha_1}^{\alpha_0} -(\alpha^2 B_\alpha)_\alpha\, d\alpha - \gamma\int_{\alpha_1}^{\alpha_0} \alpha^2 B_{\alpha\alpha}\, d\alpha \leq -cQ\alpha_0^2 + C\hat{Q}\alpha_1^2 + c\hat{Q}\alpha_0^2\gamma \leq 0\,.$$

Negativity is by the choice of α_1 in (3.60) and by the fact that

$$\gamma \leq c\sqrt{\frac{Q}{\hat{Q}}}\,, \tag{3.69}$$

which is much overdone in (3.61).

Therefore, we get, combining with (3.68) (here $\eta >$ is an absolute constant and it is at least the maximum of all our $3\gamma + \gamma^2$)

$$c\,\frac{\alpha_0^3 Q^2}{\hat{Q}\,(\gamma + \frac{1}{Q}\alpha_0)} \leq n \leq -(1+\eta)\int_{\alpha_1}^{\alpha_0} (e^{\frac{1}{1+\eta}\gamma^2} B_\gamma)_\gamma\, d\alpha\,,$$

or

$$\int_{\alpha_1}^{\alpha_0} (-e^{\frac{1}{1+\eta}\gamma^2} B_\gamma)_\gamma \, d\alpha \geq C \, \frac{\alpha_0^3 Q^3}{\hat{Q}\,(Q\gamma + \alpha_0)} \,. \tag{3.70}$$

Function B is smooth, concave in γ and symmetric in γ (the latter is by definition). In particular $B_\gamma(\alpha, \beta, 0) = 0$. So after integrating in γ on $[0, \gamma], \gamma < \gamma_0$ we get

$$\int_{\alpha_1}^{\alpha_0} (-B_\gamma) \, d\alpha \geq C \, \alpha_0^3 \frac{Q^2}{\hat{Q}} [\log(\alpha_0 + Q\gamma) - \log \alpha_0] = C \, \alpha_0^3 \frac{Q^2}{\hat{Q}} \log(1 + \frac{Q}{\alpha_0}\gamma) \,. \tag{3.71}$$

Integrate again in γ on $[0, \gamma_0]$. We get the integral over $[\alpha_1, \alpha_0]$ of the oscillation of B, which is

$$\int_{\alpha_1}^{\alpha_0} [B(\alpha, \beta, 0) - B(\alpha, \beta, \gamma_0)] \, d\alpha \geq C \, \alpha_0^3 \frac{Q^2}{\hat{Q}} \cdot \frac{\alpha_0}{Q} (1 + Q\frac{\gamma_0}{\alpha_0}) \log(1 + Q\frac{\gamma_0}{\alpha_0}) \,.$$

But this oscillation is smaller than $C\hat{Q}\alpha_0^2$. We get the inequality

$$C \, \alpha_0^4 \frac{Q}{\hat{Q}} \, (1 + Q\frac{\gamma_0}{\alpha_0}) \log(1 + Q\frac{\gamma_0}{\alpha_0}) \leq \alpha_0^2 \hat{Q} \,. \tag{3.72}$$

Notice that α_0, γ_0, γ_0/α_0 are all powers of $\frac{Q}{\hat{Q}}$, which we expect to be a sort of $\frac{1}{(\log Q)^p}$. Then we get the estimate in terms of *powers* of $\frac{Q}{\hat{Q}}$:

$$C \, \alpha_0^2 \frac{Q^2}{\hat{Q}^2} \frac{\gamma_0}{\alpha_0} \log(1 + Q\frac{\gamma_0}{\alpha_0}) \leq 1 \,. \tag{3.73}$$

Let us count the powers of $\frac{Q}{\hat{Q}}$: α_0^2 brings power 2—by (3.60), $\frac{\gamma_0}{\alpha_0}$ brings power 1 by (3.61), so totally we have $\frac{1}{(\log Q)^{5p}} \log \frac{Q}{(\log Q)^{\cdots}}$ in the left hand side.

We can see that if $\hat{Q} \leq Q \log^p Q$ with $p < \frac{1}{5}$, then (3.73) leads to a contradiction. So we proved

Theorem 3.4.5. *The weighted weak norm of the martingale transform for weights* $w \in A_1^{dyadic}$ *can reach* $c\,[w]_{A_1} \log^p [w]_{A_1}$ *for any positive* $p < 1/5$.

3.4.2 Obstacle conditions for B.

Now we want to show the following obstacle condition for B, which we already used:

$$\text{if } |\gamma| < \frac{1}{4}\,,\ \text{then } B(1,\beta,\gamma) \geq \frac{\beta}{8}\,. \tag{3.74}$$

Let $I := [0,1]$. Given numbers $|f| < \lambda/4, \frac{F}{m} = \lambda$ it is enough to construct functions φ, ψ, w on I such that

Put $\varphi = -a$ on I_{--}, $= b$ on I_{++}, zero otherwise. And $w = 1$ on $I_{--} \cup I_{++}$, and $w = Q$ otherwise. Then put

$$\psi := (\varphi, h_{I_-})h_{I_-} - (\varphi, h_{I_+})h_{I_+}\,.$$

Let $0 < a < b$ and a is close to b. Put $\lambda = (a+b)/4$. Then average of φ is small with respect to λ and we can prescribe it. $F = (a+b)/4, m = 1$. On the other hand, function ψ (which is a martingale transform of $\varphi - \langle\varphi\rangle$) is at least $-(\varphi, h_{I_+})h_{I_+} \geq \frac{1}{2}b \geq \lambda$ on I_{+-}, whose w-measure is more than $\frac{1}{3}w(I)$. So

$$B(1,\beta,\gamma) \geq \frac{1}{3}\beta\,, \tag{3.75}$$

for all small γ and $\beta \asymp Q$. This is what we wanted to prove.

Chapter 4

One weight estimates in metric spaces

4.1 Main results

Here we give a proof of the A_2 conjecture in geometrically doubling metric spaces (GDMS), i.e. a metric space where one can fit not more than a fixed amount of disjoint balls of radius r in a ball of radius $2r$.

Our main result is the following.

Theorem 4.1.1 (A_2 theorem for a geometrically doubling metric space)**.** *Let X be a geometrically doubling metric space, μ and T as above, $w \in A_{2,\mu}$. In addition we assume that μ is a doubling measure. Then*

$$\|T\|_{L^2(wd\mu)\to L^2(wd\mu)} \leqslant C(T)[w]_{2,\mu}. \tag{4.1}$$

The proof is organized as follows:

1. A construction of a probability space of random "dyadic" lattice in a metric space is given in Section 4.2;

2. Averaging trick of Hytönen [HPTV] (but we think we simplified it) is given in Section 4.4.3;

3. A linear estimate of weighted dyadic shift on metric space from [NV], which uses

Bellman function technique, is given in Sections 4.4.1 and 4.4.2. For another proof of the linear estimate for weighted dyadic shifts, which can be easily adjusted to the metric case, we refer to [T].

4.2 The main construction

4.2.1 First Step

Consider a compact doubling metric space X with metric d and doubling constant A. Instead of $d(x, y)$ we write $|xy|$. Precisely, the definition is the following.

Definition 24. Suppose $(X, |.|)$ is a metric space. We call it geometrically doubling with constant A, if for any $x \in X$ and $r > 0$ we can fit no more than A disjoint balls of radius 2 in the ball $B(x, r)$.

As authors of [HM], we essentially use the idea of Michael Christ [Chr], but randomize his construction in a different way. Therefore, we want to guard the reader that even though on the surface the proof below is very close to the proof from [HM], however, our construction is essentially different, and so the proof of the assertion in our main lemma, which was not hard in [HM], becomes much more subtle here.

We now proceed to the construction.

For a number $k > 0$ we say that a set G is a k-grid if G is maximal (with respect to inclusion) set, such that for any $x, y \in G$ we have $d(x, y) > k$.

Let from now on $\operatorname{diam} X = 1$. Take a small positive number $\delta \ll 1$ depending on the doubling constant of X and a large natural number N, and for every $M \geqslant N$ fix $G_M = \{z_M^\alpha\}$, a certain δ^M-grid of X. Now take G_N and randomly choose a $G_{N-1} = \delta^{N-1}$-grid in G_N.

Then take G_{N-1} and randomly choose a $G_{N-2} = \delta^{N-2}$-grid in G_{N-1}. Do this N times. Notice that G_0 consists of just one random point of G_N.

We explain what is "randomly". Since X is a compact metric space, all G_k's are finite. Therefore, there are finitely many $(N-1)$-grids in G_N. We choose one of them with a probability

$$\frac{1}{\text{number of } (N-1)\text{-grids in } G_N}.$$

Our first lemma is the following.

Lemma 4.2.1. *For* $k = 0, \ldots, N$

$$\bigcup_{y \in G_{N-k}} B(y, 3\delta^{N-k}) = X.$$

Remark 11. For $N+k, k \geq 0$, instead of $N-k$ this is obvious.

Proof. Take $x \in X$. Then, since G_N is maximal, there exists a point $y_0 \in G_N$, such that $|xy_0| \leqslant \delta^N$. Since G_{N-1} is maximal in G_N, there is a point $y_1 \in G_{N-1}$, such that $|y_0 y_1| \leqslant \delta^{N-1}$. Similarly we get $y_2, \ldots, y_k$ and then

$$|xy_k| \leqslant |xy_0| + \ldots + |xy_k| \leqslant \delta^N + \ldots + \delta^{N-k} = \delta^{N-k}(1 + \delta + \ldots + \delta^k) \leqslant \frac{\delta^{N-k}}{1-\delta} \leqslant 2\delta^{N-k}.$$

□

Once we have all our sets G_N, we introduce a relationship $\prec$ between points. We follow [HM] and [Chr].

Take a point $y_{k+1} \in G_{k+1}$. There exists at most one $y_k \in G_k$, such that $|y_{k+1} y_k| \leqslant \frac{\delta^k}{4}$.

This is true since if there are two such points $y_k^1,\ y_k^2$, then

$$|y_k^1 y_k^2| \leqslant \frac{\delta^k}{2},$$

which is a contradiction, since G_k was a δ^k-grid in G_{k+1}.

Also there exists at least one $z_k \in G_k$ such that $|y_{k+1} z_k| \leqslant 3\delta^k$. This is true by the lemma.

Now, if there exists an y_k as above, we set $y_{k+1} \prec y_k$. If no, then we pick one of z_k as above and set $y_{k+1} \prec z_k$. For all other $x \in G_k$ we set $y_{k+1} \not\prec x$. Then extend by transitivity.

We also assume that $y_k \prec y_k$. This is if y_k on the left happened to belong already to G_{k+1}.

We do this procedure randomly and independently, and treat same families of G_k's with different $\prec$-law as different families.

Take now a point $y_k \in G_k$ and define

$$Q_{y_k} = \bigcup_{z \prec y_k, z \in G_\ell} B(z, \frac{\delta^\ell}{100}).$$

Lemma 4.2.2. *For every k we have*

$$X = \bigcup_{y_k \in G_k} \operatorname{clos}(Q_{y_k})$$

Remark 12. There is only one point in G_0, and $clos(Q_y), y \in G_0$, is just X. But for small δ, $X = \bigcup_{y_1 \in G_1} \operatorname{clos}(Q_{y_1})$ is a genuine (and random) splitting of X.

Proof. Take any $x \in X$. By the previous lemma, for every $m > k$ there exists a point

$x_m \in G_m$, such that $|xx_m| \leqslant 3\delta^m$. In particular, $x_m \to x$. Fix for a moment x_m. Then there are points $y_{m-1} \in G_{m-1}, \ldots, y_k \in G_k$, such that $x_m \prec y_{m-1} \prec \ldots \prec y_k$. In particular, $x_m \in Q_{y_k}$, where y_k depends on x_m. Then

$$|y_k x| \leqslant |y_k x_m| + |x_m x| \leqslant |y_k x_m| + 3\delta^m \leqslant |y_k x_m| + 3\delta^k.$$

Moreover, by the chain of $\prec$'s, we know that $|y_k x_m| \leqslant 10\delta^k$. Therefore,

$$|y_k x| \leqslant 15\delta^k.$$

We claim that the set $\{y_k\} = \{y_k(x_m)\}_{m \geqslant k}$ is finite independently on k. This is true since all y_k's are separated from each other and by the doubling of our space (we are "stuffing" the ball $B(x, 15\delta^k)$ with balls $B(y_k, \delta^k)$).

So, take an infinite subsequence x_m that corresponds to one point $y_k \in G_k$. Then we get $x_m \in Q_{y_k}$, $x_m \to x$, so $x \in \mathrm{clos} Q_{y_k}$, and we are done. □

Remark 13. Since the space X is compact, our random procedure consists of finitely many steps. Therefore, our probability space is discreet. We suggest to think about all probabilities just as number of good events divided by number of all events.

However, all our estimates will not depend on number of steps (and, therefore, diameter of X), which is essential.

Remark 14. We notice that in the Euclidian space, say, $\mathbb{R}$, this procedure does not give a standard dyadic lattice.

4.2.2 Second step: technical lemmata

Define

$$\tilde{Q}_{y_k} = X \setminus \bigcup_{z_k \neq y_k, z_k \in G_k} \operatorname{clos} Q_{z_k}.$$

In particular,

$$Q_{y_k} \subset \tilde{Q}_{y_k} \subset \operatorname{clos}(Q_{y_k}).$$

Lemma 4.2.3 (Lemma 4.5 in [HM])**.** *Let m be a natural number, $\varepsilon > 0$, and $\delta^m \geqslant 100\varepsilon$. Suppose $x \in \operatorname{clos} Q_{y_k}$ and $dist(x, X \setminus \tilde{Q}_{y_k}) < \varepsilon\delta^k$. Then for any chain*

$$z_{k+m} \prec z_{k+m-1} \prec \ldots \prec z_{k+1} \prec z_k,$$

such that $x \in \operatorname{clos} Q_{z_{k+m}}$, the following relationships hold

$$|z_i z_j| \geqslant \frac{\delta^j}{100}, \quad k \leqslant j < i \leqslant k+m.$$

Proof. Suppose $|z_i z_j| < \frac{\delta^j}{100}$. We first consider a case when $z_k = y_k$. Since $z_j \prec z_k = y_k$, we have $B(z_j, \frac{\delta^j}{200}) \subset Q_{y_k} \subset \tilde{Q}_{y_k}$. Therefore,

$$\frac{\delta^j}{200} \leqslant dist(z_j, X \setminus \tilde{Q}_{y_k}) \leqslant dist(x, X \setminus \tilde{Q}_{y_k}) + dist(x, z_i) + dist(z_i, z_j) < \varepsilon\delta^k + 5\delta^i + \frac{\delta^j}{100}$$

If δ is less than, say, $\frac{1}{1000}$, then we get a contradiction.

The only not obvious estimate is that $dist(x, z_i) < 5\delta_i$. It is true since $x \in \operatorname{clos} Q_{z_{k+m}}$.

We have proved the lemma with assumption that $z_k = y_k$. Let us get rid of this assumption. We know that

$$x \in \operatorname{clos} Q_{z_{k+m}} \subset \operatorname{clos} Q_{z_k}.$$

Also we have $x \in \operatorname{clos} Q_{y_k}$, so, since

$$\tilde{Q}_{z_k} = X \setminus \bigcup_{u_k \neq z_k} \operatorname{clos} Q_{u_k} \subset X \setminus \operatorname{clos} Q_{y_k},$$

we get $x \in X \setminus \tilde{Q}_{z_k}$. In particular, $dist(x, X \setminus \tilde{Q}_{z_k}) = 0 < \varepsilon\delta^k$, and we are in the situation of the first part. This finishes our proof. □

Lemma 4.2.4. *Fix $x_k \in G_k$. Then*

$$\mathbb{P}(\exists x_{k-1} \in G_{k-1} \colon |x_k x_{k-1}| < \frac{\delta^{k-1}}{1000}) \geqslant a \tag{4.2}$$

for some $a \in (0, 1)$.

Proof. We remind that we are in a compact metric situation. By rescaling we can think that we work with G_1 and choose G_0. We can even think that the metric space consists of finitely many points, it is $X := G_2$. The finite set $G_1 \subset X$ consists of points having the following properties:

1. $\forall x, y \in G_1$ we have $|xy| \geq \delta$;
2. if $z \in X \setminus G_1$ then $\exists x \in G_1$ such that $|zx| < \delta$.

These two properties are equivalent to saying that the subset G_1 of X consists of points such that $\forall x, y \in G_1$ we have $|xy| \geq \delta$ and we cannot add any point from X to G_1 without violating that property. In other words: G_1 is a *maximal* set with property 1.

Recall that here the word "maximal" means maximal with respect to inclusion, not maximal in the sense of the number of elements.

Now we consider the new metric space $Y = G_1$ and G_0 is any maximal subset such that

$$\forall x, y \in G_0\,,\ |xy| \geq 1\,. \tag{4.3}$$

In other words, we have 1. $\forall x, y \in G_0$ we have $|xy| \geq 1$;

2. if $z \in Y \setminus G_0$ then $\exists x \in G_0$ such that $|zx| < 1$.

There are finitely many such maximal subsets G_0 of Y. We prescribe for each choice the same probability. Now we want to prove the claim that is even stronger than (4.2). Namely, we are going to prove that given $y \in Y$

$$\mathbb{P}(\exists x_0 \in G_0 \colon x_0 = y) \geqslant a\,, \tag{4.4}$$

where a depends only on δ and the constants of geometric doubling of our compact metric space.

Let Y be any metric space with finitely many elements. We will color the points of Y into red and green colors. The coloring is called proper if

1. every red point does not have any other red point at distance < 1;
2. every green point has at least one red point at distance < 1.

Given *a proper coloring* of Y the collection of red points is called 1-*lattice*. It is a maximal (by inclusion) collection of points at distance ≥ 1 from each other.

What we need to finish the proof is

Lemma 4.2.5. *Let Y be a finite metric space as above. Assume Y has the following property:*

$$\textit{In every ball of radius less than } 1 \textit{ there are at most } d \textit{ elements}\,. \tag{4.5}$$

Let $\mathcal{L}$ be a collection of 1*-lattices in* Y*. Elements of* $\mathcal{L}$ *are called* L*. Let* $v \in Y$*. Then*

$$\frac{\text{the number of 1-lattices } L \text{ such that } v \text{ belongs to } L}{\text{the total number of 1-lattices } L} \geq a > 0\,,$$

where a *depends only on* d*.*

Proof. Given $v \in Y$ consider all subsets of $B(v,1) \setminus v$, this collection is called $\mathcal{S}$. Let $S \in \mathcal{S}$. We call W_S the collection of all proper colorings such that v is green, all elements of S are red, and all elements of $B(v,1) \setminus S$ are green. We call $\tilde{S}$ all points in Y, which are not in $B(v,1)$, but at distance < 1 from some point in S.

All proper colorings of Y such that v is red are called B. Let us show that

$$\operatorname{card} W_S \leq \operatorname{card} B\,. \tag{4.6}$$

Notice that if (4.6) were proved, we would be done with Lemma 4.2.5, $a \geq 2^{-d+1}$, and, consequently, the proof of the main lemma would be finished, $a \geq 2^{-\delta^{-D}}$, where D is a geometric doubling constant.

To prove (4.6) let us show that we can recolor any proper coloring from W_S into the one from B, and that this map is injective. Let $L \in W_S$. We

1. Color v into red;

2. Color S into green;

3. Elements of $\tilde{S}$ were all green before. We leave them green, but we find among them all those y that now in the open ball $B(y,1)$ in Y all elements are green. We call them yellow (temporarily) and denote them Z;

4. We enumerate Z in any way (non-uniqueness is here, but we do not care);

5. In the order of enumeration color yellow points to red, ensuring that we skip recoloring of a point in Z if it is at < 1 distance to any previously colored yellow-to-red point from Z. After several steps all green and yellow elements of $\tilde{S}$ will have the property that at distance < 1 there is a red point;

6. Color the rest of yellow (if any) into green and stop.

We result in a proper coloring (it is easy to check), which is obviously B. Suppose L_1, L_2 are two different proper coloring in W_S. Notice that the colors of $v, S, B(v,1) \setminus S$, $\tilde{S}$ are the same for them. So they differ somewhere else. But our procedure does not touch "somewhere else". So the modified colorings L_1', L_2' that we obtain after the algorithm 1-6 will differ as well may be even more). So our map $W_S \to B$ (being not uniquely defined) is however injective. We proved (4.6).

□

Thus, the proof of the Lemma 4.2.4 is finished.

□

Remark. We are grateful to Michael Shapiro and Dapeng Zhan who helped us to prove Lemma 4.2.4.

4.2.3 Main definition and theorem

Fix a number γ, $0 < \gamma < 1$. Later the choice of γ will be dictated by the Calderón-Zygmund properties of the operator T. Also fix a sufficiently big r. The coice of r will be made in this section.

Definition 25 (Bad cubes)**.** Take a "cube" $Q = Q_{x_k}$. We say that Q is good if there exists a cube $Q_1 = Q_{x_n}$, such that if

$$\delta^k \leqslant \delta^r \delta^n \quad (k \geqslant n + r)$$

then either

$$dist(Q, Q_1) \geqslant \delta^{k\gamma} \delta^{n(1-\gamma)}$$

or

$$dist(Q, X \setminus Q_1) \geqslant \delta^{k\gamma} \delta^{n(1-\gamma)}.$$

Remark 15. Notice that $\delta^k = \ell(Q)$ just by definition.

If Q is not good we call it bad.

Theorem 4.2.6. *Fix a cube Q_{x_k}. Then*

$$\mathbb{P}(Q_{x_k} \text{ is bad}) \leqslant \frac{1}{2}.$$

Remark 16 (Discussion)**.** This theorem makes sense because when we fix a cube Q_k, say, $k \geqslant N$, so the grid G_k is not even random, we can make big cubes random. And we claim that for big quantity of choices, our big cubes will have Q_k either "in the middle" or far away, but not close to the boundary.

Definition 26. For $Q = Q_{x_k}$ define

$$\delta_Q(\varepsilon) = \delta_Q = \{x \colon dist(x, Q) \leqslant \varepsilon\delta^k \text{ and } dist(x, X \setminus Q) \leqslant \varepsilon\delta^k\}$$

Lemma 4.2.7. *Let us start with level N by fixing a δ^N-grid (non-random), and let $k < N$,*

x_k denoting the points of the (random) grid G_k. Fix a point $x \in X$.

$$\mathbb{P}(\exists x_k \in G_k : x \in \delta_{Q_{x_k}}) \leqslant \varepsilon^\eta$$

for some $\eta > 0$.

Proof of the theorem. Take the cube Q_{x_k}. There is a unique (random!) point x_{k-s} such that $x_k \in Q_{x_{k-s}}$. Then

$$dist(Q_{x_k}, X \setminus Q_{x_{k-s}}) \geqslant dist(x_k, X \setminus Q_{x_{k-s}}) - diam(Q_{x_k}) \geqslant dist(x_k, X \setminus Q_{x_{k-s}}) - C\delta^k.$$

Assume that $dist(x_k, X \setminus Q_{x_{k-s}}) > 2\delta^{k\gamma}\delta^{(k-s)(1-\gamma)}$ and that $s \geqslant r$ (this assumption is obvious, otherwise $Q_{x_{k-s}}$ does not affect goodness of Q_{x_k}).

Then, if r is big enough ($\delta^{r(1-\gamma)} < \frac{1}{C}$) we get

$$dist(Q_{x_k}, X \setminus Q_{x_{k-s}}) \geqslant \delta^{k\gamma}\delta^{(k-s)(1-\gamma)},$$

and so Q_{x_k} is good. Therefore,

$$\mathbb{P}(Q_{x_k} \text{ is bad }) \leqslant C \sum_{s \geqslant r} \mathbb{P}(x_k \in \delta_{Q_{k-s}}(\varepsilon = 2\delta^{s\gamma})) \leqslant C \sum_{s \geqslant r} \delta^{\eta\gamma s} \leqslant 100C\delta^{\eta\gamma r}.$$

By the choice of η, for sufficiently large r this is less than $\frac{1}{2}$. □

Proof of the lemma. Let x_k be such that $x \in \operatorname{clos} Q_{x_k}$ (see Lemma 4.2.2). We will estimate $\mathbb{P}(dist(x, X \setminus \tilde{Q}_k) < \varepsilon\delta^k) \,|\, x \in \operatorname{clos} Q_{x_k})$. Fix the largest m such that $500\varepsilon \leqslant \delta^m$. Choose a point x_{k+m} such that $x \in \operatorname{clos} Q_{x_{k+m}}$. Then by the main lemma

$$\mathbb{P}(\exists x_{k+m-1} \in G_{k+m-1} \colon |x_{k+m}x_{k+m-1}| < \frac{\delta^{k+m-1}}{1000}) \geqslant a.$$

Therefore,

$$\mathbb{P}(\forall x_{k+m-1} \in G_{k+m-1} \colon |x_{k+m}x_{k+m-1}| \geqslant \frac{\delta^{k+m-1}}{1000}) \leqslant 1-a.$$

Let now

$$x_{k+m} \prec x_{k+m-1}.$$

Then

$$\mathbb{P}(\forall x_{k+m-2} \in G_{k+m-2} \colon |x_{k+m-1}x_{k+m-2}| \geqslant \frac{\delta^{k+m-2}}{1000}) \leqslant 1-a.$$

So by Lemma 4.2.3

$$\mathbb{P}(dist(x, X \setminus \tilde{Q}_k) < \varepsilon\delta^k) \leqslant \mathbb{P}(|x_{k+j}x_{k+j-1}| \geqslant \frac{\delta^{k+j-1}}{1000}\ \forall j = 1, \dots, m) \leqslant (1-a)^m \leqslant C\varepsilon^\eta$$

for

$$\eta = \frac{\log{(1-a)}}{\log(\delta)}.$$

□

4.2.4 Probability to be "good" is the same for every cube

We make the last step to make the probability to be "good" not just bounded away from zero, but the same for all cubes. We use the idea from [M].

Take a cube $Q(\omega)$. Take a random variable $\xi_Q(\omega')$, which is equally distributed on $[0, 1]$.

We know that

$$\mathbb{P}(Q \text{ is good}) = p_Q > a > 0.$$

We call Q "really good" if

$$\xi_Q \in [0, \frac{a}{p_Q}].$$

Otherwise Q joins bad cubes. Then

$$\mathbb{P}(Q \text{ is really good}) = a,$$

and we are done.

4.3 The Haar shift decomposition

Take two step functions, f and g. We first fix an N-grid G_N in X, and "cubes" on level N, such that f and g are constants on every such cube. Then we start our randomization process.

As we mentioned, this process consists of finitely many steps, so all probabilistic terminology becomes trivial: we have a finite probability space.

Starting from G_N, we go "up" and on each level get dyadic cubes (random Christ's cubes). They have the usual structure of being either disjoint or one containing the other. For each dyadic cube Q we have several dyadic sons, they are denoted by $s_i(Q)$, $i = 1, \ldots, M(Q) \leq M$. The number M here is universal and depends only on geometric doubling constants of the space X.

Definition 27. By $\mathcal{E}_k$ we denote set of all dyadic "cubes" of generation k. We call $Q_k^i \subset Q_{k-1}^j$, $Q_k^i \in \mathcal{E}_k$ **sons** of Q_{k-1}^j.

With every cube $Q = Q_{x_k}$ we associate **Haar functions** h^j_Q, $j = 1, \dots, M-1$, with following properties:

1. h^j_Q is supported on Q;

2. h^j_Q takes constant values on each "son" of Q;

3. For any two cubes Q and R, we have $(h^j_Q, h^i_R) = 0$, and $(h^j_Q, 1) = 0$;

4. $\|h^j_Q\|_\infty \leqslant \frac{C}{\sqrt{\mu(Q)}}$.

We notice that the last property implies that $\|h^j_Q\|_2 \leqslant C$.

We use angular brackets to denote the average: $\langle f\rangle_{Q,\mu} := \frac{1}{\mu(Q)}\int_Q f\,d\mu$. When we average over the whole space X, we drop the index and write $\langle f\rangle = \frac{1}{\mu(X)}\int_X f d\mu$.

Our main "tool" is going to be the famous "dyadic shifts". Precisely, we call by $\mathbb{S}_{m,n}$ the operator given by the kernel

$$f \to \sum_{L\in\mathcal{D}} \int_L a_L(x,y) f(y) dy\,,$$

where

$$a_L(x,y) = \sum_{\substack{I\subset L, J\subset L\\ g(I)=g(L)+m,\, g(J)=g(L)+n}} c_{L,I,J} h^j_J(x) h^i_I(y)\,,$$

where h^i_I, h^j_J are Haar functions normalized in $L^2(d\mu)$ and satisfying (iv), and $|c_{L,I,J}| \leq \frac{\sqrt{\mu(I)}\sqrt{\mu(J)}}{\mu(L)}$. Often we will skip superscripts i, j.

Definition 28. We call the number $m+n+1$ the **complexity** of a shift $\mathbb{S}_{m,n}$.

Our next aim is to decompose the bilinear form of the operator T into bilinear forms of dyadic shifts, which are estimated in the Section 4.4.2. The rest will be the so-called "paraproducts", estimated in the Section 4.4.1.

Functions $\{\chi_X\} \cup \{h_Q^j\}$ form an orthogonal basis in the space $L^2(X,\mu)$. Therefore, we can write

$$f = \langle f\rangle \chi_X + \sum_Q \sum_j (f, h_Q^j) h_Q^j, \qquad g = \langle g\rangle \chi_X + \sum_R \sum_i (g, h_R^i) h_R^i.$$

First, we state and proof the theorem, that says that essential part of bilinear form of T can be expressed in terms of pair of cubes, where the smallest one is good. We follow the idea of Hytönen [H]. In fact, the work [H] improved on "good-bad" decomposition of [NTV], [NTV2], [NTV3] by replacing inequalities by an equality.

Theorem 4.3.1. *Let T be any linear operator. Then the following equality holds:*

$$\pi_{good} \mathbb{E} \sum_{\substack{Q,R,i,j \\ \ell(Q) \geqslant \ell(R)}} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i) = \mathbb{E} \sum_{\substack{Q,R,i,j \\ \ell(Q)\geqslant \ell(R),\, R \text{ is good}}} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i).$$

The same is true if we replace $\geqslant$ by $>$.

Proof. We denote

$$\sigma_1(T) = \sum_{\ell(Q)\geqslant \ell(R)} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i).$$

$$\overline{\sigma_1(T)} = \sum_{\substack{\ell(Q)\geqslant \ell(R) \\ R \text{ is good}}} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i).$$

We would like to get a relationship between $\mathbb{E}\sigma_1(T)$ and $\mathbb{E}\overline{\sigma_1(T)}$.

We fix R and write (using $g_{good} := \sum_{R \text{ is good}} (g, h^i_R) h^i_R$)

$$\sum_Q \sum_{R \text{ is good}} (Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R) = \left(T(f - \langle f \rangle \chi_X), \sum_{R \text{ is good}} (g, h^i_R) h^i_R \right) =$$

$$\big(T(f - \langle f \rangle \chi_X), g_{good} \big) \,.$$

Taking expectations, we obtain

$$\mathbb{E} \sum_{Q,R} (Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R) \mathbf{1}_{R \text{ is good}} =$$

$$\mathbb{E}(T(f - \langle f \rangle \chi_X), g_{good}) = (T(f - \langle f \rangle \chi_X), \mathbb{E}\, g_{good}) =$$

$$\pi_{good}(T(f - \langle f \rangle \chi_X), g) = \pi_{good} \mathbb{E} \sum_{Q,R} (Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R). \quad (4.7)$$

Next, suppose $\ell(Q) < \ell(R)$. Then goodness of R does not depend on Q, and so

$$\pi_{good}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R) = \mathbb{E} \left((Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R) \mathbf{1}_{R \text{ is good}} | Q, R \right) \,.$$

Let us explain this equality. The right hand side is conditioned: meaning that the left hand side involves the fraction of the number of all lattices containing Q, R in this lattice and such that R (the larger one) is good to the number of lattices containing Q, R in it. This fraction is exactly π_{good}. Now we fix a pair of Q, R, $\ell(Q) < \ell(R)$, and multiply both sides by the probability that this pair is in the same dyadic lattice from our family. This probability is just the ratio of the number of dyadic lattices in our family containing elements Q and R to the number of all dyadic lattices in our family. After multiplication by this ratio and the summation of all terms with $\ell(Q) < \ell(R)$ we get finally,

$$\pi_{good}\mathbb{E}\sum_{\ell(Q)<\ell(R)}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R) = \mathbb{E}\sum_{\ell(Q)<\ell(R)}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R)\mathbf{1}_{R\text{ is good}}\,. \tag{4.8}$$

Now we use first (4.7) and then (4.8):

$$\begin{aligned}\pi_{good}\mathbb{E}\sum_{Q,R}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R) &= \mathbb{E}\sum_{Q,R}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R)\mathbf{1}_{R\text{ is good}} = \\ &= \mathbb{E}\sum_{\ell(Q)<\ell(R)}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R)\mathbf{1}_{R\text{ is good}} + \\ &\mathbb{E}\sum_{\ell(Q)\geqslant\ell(R)}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R)\mathbf{1}_{R\text{ is good}} = \\ &= \pi_{good}\mathbb{E}\sum_{\ell(Q)<\ell(R)}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R) + \\ &\mathbb{E}\sum_{\ell(Q)\geqslant\ell(R), R\text{ is good}}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R),\end{aligned} \tag{4.9}$$

and therefore

$$\mathbb{E}\sum_{\ell(Q)\geqslant\ell(R), R\text{ is good}}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R) = \pi_{good}\mathbb{E}\sum_{\ell(Q)\geqslant\ell(R)}(Th^j_Q, h^i_R)(f, h^j_Q)(g, h^i_R). \tag{4.10}$$

□

This is the main trick. To have the whole sum expressed as the multiple of the sum, where the **smaller** in size cube is good, is **very useful** as we will see. It gives extra decay on matrix coefficients (Th^j_Q, h^i_R) and allows us to represent our operator as "convex combination of dyadic shifts".

So, we have obtained that

$$\mathbb{E}\sigma_1(T) = \pi_{good}^{-1} \cdot \mathbb{E}\overline{\sigma_1(T)}.$$

Thus, to estimate $\mathbb{E}\sigma_1(T)$ it is enough to estimate $\mathbb{E}\overline{\sigma_1(T)}$. Absolutely the same symmetrically holds for $\sigma_2(T)$.

4.3.1 Paraproducts

In this subsection we take care of the terms $\langle f\rangle\chi_X$ and $\langle g\rangle\chi_X$. These terms will lead to so called paraproducts. In fact, let us introduce three auxiliary operators:

$$\pi(f) := \pi_{T\chi_X}(f) := \sum_{Q,j} \langle f\rangle_Q (T\chi_X, h_Q^j) h_Q^j; \tag{4.11}$$

$$\pi_*(f) := \sum_{Q,j} (f, h_Q^j)(T^*\chi_X, h_Q^j) \frac{\chi_Q}{\mu(Q)} = (\pi_{T^*\chi_X})^*(f); \tag{4.12}$$

$$o(f) := \langle f\rangle \langle T\chi_X\rangle \chi_X. \tag{4.13}$$

Recall that $\langle\varphi\rangle$ denotes $\frac{1}{\mu(X)}\int_X \varphi\, d\mu$. These operators depend on the dyadic grid we chose. We shall need the following technical lemma.

Lemma 4.3.2.

$$(\pi(f), g) = \langle f\rangle (T\chi_X, g - \langle g\rangle\chi_X) + \sum (\pi h_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i),$$

$$(\pi_*(f), g) = \langle g\rangle (T^*\chi_X, f - \langle f\rangle\chi_X) + \sum (\pi_* h_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i).$$

Proof. The second equality follows from the first one and the definition of π_*. We prove the first equality. We will not write superscripts i and j in Haar functions.

We write

$$\pi(f) = \langle f\rangle \pi(\chi_X) + \sum (f, h^i_Q)\pi(h^i_Q).$$

Notice that

$$\pi(\chi_X) = \sum (T\chi_X, h^i_Q)h^i_Q = T\chi_X - \langle T\chi_X\rangle,$$

and that $\pi(f)$ is orthogonal to χ_X. Thus,

$$(\pi(f), g) = (\pi(f), g - \langle g\rangle\chi_X) = \langle f\rangle(\pi(\chi_X), \sum (g, h^j_R)h^j_R) + \sum (\pi h^i_Q, h^j_R)(f, h^i_Q)(g, h^j_R) =$$

$$= \langle f\rangle(T\chi_X, g - \langle g\rangle\chi_X) + \sum (\pi h^i_Q, h^j_R)(f, h^i_Q)(g, h^j_R),$$

as desired. The last equality is true because $\langle T\chi_X\rangle$ is orthogonal to $g - \langle g\rangle\chi_X$. □

Notice that π, π^* depend on the random dyadic grid. We introduce a random operator

$$\tilde{T} = Tf - \pi(f) - \pi_*(f).$$

Now we state the following very useful lemma.

Lemma 4.3.3.

$$(Tf, g) = \pi_{good}^{-1}\mathbb{E} \sum_{\substack{Q,R \\ \textit{smaller is good}}} (\tilde{T}h^i_Q, h^j_R)(f, h^i_Q)(g, h^j_R) +$$

$$\mathbb{E}(\pi(f), g) + \mathbb{E}(\pi_*(f), g) + \langle f\rangle\langle g\rangle(T\chi_X, \chi_X).$$

Proof. First, we write

$$(Tf,g)=\sum(Th^i_Q,h^j_R)(f,h^i_Q)(g,h^j_R)+\langle f\rangle(T\chi_X,g)+\langle g\rangle(T^*\chi_X,f-\langle f\rangle\chi_X).$$

We take expectations now. Notice that only the first term in the right-hand side depends on a dyadic grid. Therefore,

$$(Tf,g)=\mathbb{E}\sum(Th^i_Q,h^j_R)(f,h^i_Q)(g,h^j_R)+\langle f\rangle(T\chi_X,g)+\langle g\rangle(T^*\chi_X,f-\langle f\rangle\chi_X).$$

We focus on the first term. By the Theorem 4.3.1, we know that

$$\begin{aligned}\mathbb{E}\sum(Th^i_Q,h^j_R)(f,h^i_Q)(g,h^j_R)=\pi_{good}^{-1}\mathbb{E}\sum_{\text{smaller is good}}(Th^i_Q,h^j_R)(f,h^i_Q)(g,h^j_R)=\\ =\pi_{good}^{-1}\mathbb{E}\sum_{\text{smaller is good}}(\tilde{T}h^i_Q,h^j_R)(f,h^i_Q)(g,h^j_R)+\\ +\pi_{good}^{-1}\mathbb{E}\sum_{\text{smaller is good}}(\pi h^i_Q,h^j_R)(f,h^i_Q)(g,h^j_R)+\\ \pi_{good}^{-1}\mathbb{E}\sum_{\text{smaller is good}}(\pi_* h^i_Q,h^j_R)(f,h^i_Q)(g,h^j_R).\end{aligned}\tag{4.14}$$

The first term is one of those that we want to get in the right-hand side.

On the other hand, we want to get a result for paraproducts, similar to the Theorem 4.3.1. Indeed, it is clear that

$$(\pi h^i_Q,h^j_R)=\langle h^i_Q\rangle_R(T\chi_X,h^j_R),$$

which is non-zero only if $R \subset Q$, and $R \neq Q$. So,

$$\mathbb{E} \sum_{\text{smaller is good}} (\pi h_Q^i, h_R^j)(f, h_Q^i)(g, h_R^j) =$$

$$\mathbb{E} \sum_{R \subset Q} \langle h_Q^i \rangle_R (T\chi_X, h_R^j)(f, h_Q^i)(g, h_R^j) \mathbf{1}_{R \text{ is good}} =$$

$$= \mathbb{E} \sum_R (T\chi_X, h_R^j)(g, h_R^j) \mathbf{1}_{R \text{ is good}} \sum_{Q : R \subsetneq Q} (f, h_Q^i) \langle h_Q^i \rangle_R. \quad (4.15)$$

We now see that since $f = \langle f \rangle \chi_X + \sum_Q (f, h_Q^i) h_Q^i$, we have

$$\langle f \rangle_R - \langle f \rangle = (f, \mu(R)^{-1} \chi_R) - \langle f \rangle = \sum_{Q : R \subsetneq Q} (f, h_Q^i) \langle h_Q^i \rangle_R = \sum_Q (f, h_Q^i) \langle h_Q^i \rangle_R.$$

Therefore,

$$\mathbb{E} \sum_R (T\chi_X, h_R^j)(g, h_R^j) \mathbf{1}_{R \text{ is good}} \sum_Q (f, h_Q^i) \langle h_Q^i \rangle_R =$$

$$\mathbb{E} \sum_R (T\chi_X, h_R^j)(g, h_R^j) \mathbf{1}_{R \text{ is good}} (\langle f \rangle_R - \langle f \rangle). \quad (4.16)$$

Now it is clear that we can take the expectation inside (we have no Q anymore, which was preventing us from doing that), and so we get

$$\mathbb{E} \sum_{\text{smaller is good}} (\pi h_Q^i, h_R^j)(f, h_Q^i)(g, h_R^j) = \pi_{good} \mathbb{E} \sum_R (T\chi_X, h_R^j)(g, h_R^j)(\langle f \rangle_R - \langle f \rangle).$$

Making all above steps backwards, we get

$$\mathbb{E} \sum_{\text{smaller is good}} (\pi h_Q^i, h_R^j)(f, h_Q^i)(g, h_R^j) = \pi_{good} \mathbb{E} \sum (\pi h_Q^i, h_R^j)(f, h_Q^j)(g, h_R^j)$$

Therefore,

$$\pi_{good}^{-1}\mathbb{E}\sum_{\text{smaller is good}}(\pi h_Q^i, h_R^j)(f, h_Q^i)(g, h_R^j)+$$

$$\pi_{good}^{-1}\mathbb{E}\sum_{\text{smaller is good}}(\pi_* h_Q^i, h_R^j)(f, h_Q^i)(g, h_R^j) =$$

$$= \mathbb{E}\sum(\pi h_Q^i, h_R^j)(f, h_Q^i)(g, h_R^j) + \mathbb{E}\sum(\pi_* h_Q^i, h_R^j)(f, h_Q^i)(g, h_R^j) =$$

$$= \mathbb{E}(\pi(f), g) + \mathbb{E}(\pi_*(f), g) - \mathbb{E}[\langle f\rangle(T\chi_X, g - \langle g\rangle\chi_X)] - \mathbb{E}[\langle g\rangle(T^*\chi_X, f - \langle f\rangle\chi_X)]. \quad (4.17)$$

We now use that last two terms do not depend on the dyadic grid, and so we drop expectations. Finally,

$$(Tf, g) = \mathbb{E}\sum_{\text{smaller is good}}(\tilde{T}h_Q^i, h_R^j)(f, h_Q^i)(g, h_R^j) + \mathbb{E}(\pi(f), g) + \mathbb{E}(\pi_*(f), g) -$$

$$- \langle f\rangle(T\chi_X, g - \langle g\rangle\chi_X) - \langle g\rangle(T^*\chi_X, f - \langle f\rangle\chi_X) + \langle f\rangle(T\chi_X, g) + \langle g\rangle(T^*\chi_X, f - \langle f\rangle\chi_X) =$$

$$= \mathbb{E}\sum_{\text{smaller is good}}(\tilde{T}h_Q^i, h_R^j)(f, h_Q^i)(g, h_R^j) + \mathbb{E}(\pi(f), g) + \mathbb{E}(\pi_*(f), g) + \langle f\rangle\langle g\rangle(T\chi_X, \chi_X).$$

(4.18)

This is what we want to prove. □

The following lemma, which will be proved later, takes care of paraproducts.

Lemma 4.3.4. *The operators π, π_* are bounded on $L^2(X, wd\mu)$, and*

$$\|\pi\|_{2,w} \leqslant C \cdot [w]_{2,\mu}.$$

The same is true for π_*.

We postpone the proof of this lemma. We also notice that the operator

$$o(f) = \langle f \rangle \langle T\chi_X \rangle \chi_X$$

is clearly bounded with desired constant. In fact, as T is bounded in the unweighted L^2, we have $\langle T\chi_X \rangle^2 \leq \|T\|_{L^2}^2 =: C_0$

$$\|o(f)\|_{2,w}^2 = \langle f \rangle^2 \langle T\chi_X \rangle^2 w(X) \leqslant C_0 \langle f^2 w \rangle \langle w^{-1} \rangle w(X) \leqslant C_0 [w]_2 \|f\|_{2,w}^2.$$

We, therefore, should take care only of the first term, with $\tilde{T}$. We now erase the tilde, and write T instead of $\tilde{T}$. Even though T is not a Calderon-Zygmund operator anymore, all further estimates are true for T (i.e., for a CZO minus paraproducts), see, for example, [HM] or [HPTV].

4.3.2 Estimates of σ_1

Our next step is to decompose σ_1 into random dyadic shifts. We write

$$\overline{\sigma_1(T)} = \sum_{\substack{\ell(Q)\geqslant\ell(R)\\ R \text{ is good}}} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i) =$$

$$= \mathbb{E} \sum_{\substack{\ell(Q)\geqslant\delta^{-r_0}\ell(R),\\ R\subset Q,\\ R \text{ is good}}} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i) +$$

$$+ \mathbb{E} \sum_{\substack{\ell(R)\leqslant\ell(Q)<\delta^{-r_0}\ell(R),\\ R\subset Q,\\ R \text{ is good}}} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i) +$$

$$+ \mathbb{E} \sum_{\substack{\ell(R)\leqslant\ell(Q),\\ R\cap Q=\emptyset,\\ R \text{ is good}}} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i). \quad (4.19)$$

Essentially, we will prove that the norm of every expectation is bounded by

$$C(T)\cdot \mathbb{E}\sum_n \delta^{-\varepsilon(T)\cdot n}\|\mathbb{S}_n\|.$$

First, we state our choice for γ, which we have seen in the definition of good cubes.

Definition 29. Put

$$\gamma = \frac{\varepsilon}{2\cdot(\varepsilon + \log_2(C))},$$

where C is the doubling constant of the function λ.

Remark 17. We remark that this choice of γ make Lemmata 4.3.5 and 4.3.6 true.

The estimate of the second sum is easy. In fact,

$$\mathbb{E} \sum_{\substack{\ell(R)\leqslant\ell(Q)<\delta^{-r_0}\ell(R),\\ R\subset Q,\\ R \text{ is good}}} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i) \leqslant C r_0\, [w]_2 \|f\| \|g\|.$$

This is bounded by at most r_0 expressions for shifts of bounded complexity, so just see [NV]. For more details, see [HPTV]

We denote

$$\Sigma_{in} = \mathbb{E} \sum_{\substack{\ell(Q)\geqslant\delta^{-r_0}\ell(R),\\ R\subset Q,\\ R \text{ is good}}} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i),$$

$$\Sigma_{out} = \mathbb{E} \sum_{\substack{\ell(R)\leqslant\ell(Q),\\ R\cap Q=\emptyset,\\ R \text{ is good}}} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i).$$

4.3.3 Estimate of Σ_{in}.

We use the following lemma.

Lemma 4.3.5. *Let T be as before; suppose $\ell(Q) \geqslant \delta^{-r_0}\ell(R)$ and $R \subset Q$. Let Q_1 be the son of Q that contains R. Then*

$$|(Th_Q^j, h_R^i)| \lesssim \frac{\ell(R)^{\frac{\varepsilon}{2}}}{\ell(Q)^{\frac{\varepsilon}{2}}} \left(\frac{\mu(R)}{\mu(Q_1)}\right)^{\frac{1}{2}}.$$

We notice that $\mu(Q_1) \asymp \mu(Q)$.

We write

$$\Sigma_{in} = \sum_{n\geqslant r_0} \sum_{\ell(Q)=\delta^{-n}\ell(R), R \text{ is good}, R\subset Q} (Th_Q^j, h_R^i)(f, h_Q^j)(g, h_R^i),$$

$$\begin{aligned}
|\Sigma_{in}| \leqslant \sum_{n\geqslant r_0} \sum_{\substack{\ell(Q)=\delta^{-n}\ell(R),\\ R \text{ is good},\\ R\subset Q}} |(Th_Q^j, h_R^i)||(f, h_Q^j)||(g, h_R^i)| \leqslant \\
\leqslant C \sum_{n\geqslant r_0} \sum_{\substack{\ell(Q)=\delta^{-n}\ell(R),\\ R \text{ is good},\\ R\subset Q}} \frac{\ell(R)^{\frac{\varepsilon}{2}}}{\ell(Q)^{\frac{\varepsilon}{2}}} \left(\frac{\mu(R)}{\mu(Q)}\right)^{\frac{1}{2}} |(f, h_Q^j)||(g, h_R^i)| = \\
= C \sum_{n\geqslant r_0} \delta^{\frac{n\varepsilon}{2}} \sum_{\substack{\ell(Q)=\delta^{-n}\ell(R),\\ R \text{ is good},\\ R\subset Q}} \left(\frac{\mu(R)}{\mu(Q)}\right)^{\frac{1}{2}} |(f, h_Q^j)||(g, h_R^i)|. \quad (4.20)
\end{aligned}$$

We **fix** functions f and g and define S_n as an operator with the following quadratic form:

$$(S_n u, v) = \sum_{\substack{\ell(Q)=\delta^{-n}\ell(R),\\ R \text{ is good},\\ R\subset Q}} \pm \left(\frac{\mu(R)}{\mu(Q)}\right)^{\frac{1}{2}} (u, h_Q^j)(v, h_R^i),$$

where $\pm$ is chosen so $|(f, h_Q^j)||(g, h_R^i)| = \pm(f, h_Q^j)(g, h_R^i)$. Then clearly S_n is a dyadic shift of complexity n, and so, see Section 4.4.2,

$$|(S_n f, g)| \leqslant C n^a [w]_2 \|f\|_w \|g\|_{w^{-1}}.$$

Therefore,

$$|\Sigma_{in}| \leqslant \sum_n Cn^a \delta^{\frac{n\varepsilon}{2}} [w]_2 \|f\|_w \|g\|_{w^{-1}} \leqslant C[w]_2 \|f\|_w \|g\|_{w^{-1}}.$$

4.3.4 Estimates for Σ_{out}

We use the following lemma from [HM].

Lemma 4.3.6. *Let T be as before, $\ell(R) \leqslant \ell(Q)$ and $R \cap Q = \emptyset$. Then the following holds*

$$|(Th^j_Q, h^i_R)| \lesssim \frac{\ell(Q)^{\frac{\varepsilon}{2}} \ell(R)^{\frac{\varepsilon}{2}}}{D(Q,R)^{\varepsilon} \sup_{z \in R} \lambda(z, D(Q,R))} \mu(Q)^{\frac{1}{2}} \mu(R)^{\frac{1}{2}},$$

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

Remark 18. We should clarify one thing here. If T was a Calderon-Zygmund operator, this estimate would be standard, see [NTV], [NTV2] or, for metric spaces, [HM]. We, however, subtracted from T two operators: paraproduct and adjoint to paraproduct. However, an easy argument (see [HPTV]) shows that if $R \cap Q = \emptyset$, then $(Th^j_Q, h^i_R) = (\tilde{T} h^j_Q, h^i_Q)$ (for the definition of $\tilde{T}$ see Lemma 4.3.4 and thereon).

Suppose now that $D(Q,R) \sim \delta^{-s} \ell(Q)$. We ask the question: what is the probability

$$\mathbb{P}(R \subset Q^{(s+s_0+10)} | Q, R \in D_\omega),$$

where s_0 is a sufficiently big number. We use the Lemma 4.2.7. Suppose that

$$R \cap Q^{(s+s_0+10)} = \emptyset.$$

Suppose also $R = R_x$ (so x is the "center" of R). Then

$$dist(x, Q^{(s+s_0+10)}) \leqslant dist(x, Q) \leqslant dist(Q, R) \leqslant C\delta^{-s}\ell(Q) =$$
$$= C\delta^{-s}\delta^{s+s_0+10}\ell(Q^{(s+s_0+10)}) = C\delta^{s_0+10}\ell(Q^{(s+s_0+10)}). \quad (4.21)$$

So $x \in \delta_{Q^{(s+s_0+10)}}(\delta^{s_0+10}))$, and the probability of this is estimated by $\delta^{\eta(s_0+10)} < \frac{1}{2}$ for sufficiently big s_0 (we remind that $\eta = \log_\delta(1-a)$). Therefore,

$$\mathbb{P}(R \subset Q^{(s+s_0+10)} | Q, R \in D_\omega) \geqslant \frac{1}{2}.$$

So

$$|\Sigma_{out}| \leqslant 2\mathbb{E}\sum_{t,s}\sum_{\substack{\ell(Q)=\delta^{-t}\ell(R),\\ D(Q,R)\sim\delta^{-s}\ell(Q),\\ R\cap Q=\emptyset}} |(Th_Q^j,h_R^i)||(f,h_Q^j)||(g,h_R^i)|\mathbf{1}_{R\,\text{is good}}\mathbf{1}_{R\subset Q^{(s+s_0+10)}} \leqslant$$

$$2\mathbb{E}\sum_{t,s}\sum_{\substack{\ell(Q)=\delta^{-t}\ell(R),\\ D(Q,R)\sim\delta^{-s}\ell(Q),\\ R\cap Q=\emptyset\\ R,Q\subset Q^{s+s_0+10}}} \frac{\ell(Q)^{\frac{\varepsilon}{2}}\ell(R)^{\frac{\varepsilon}{2}}}{D(Q,R)^{\varepsilon}\sup_{z\in R}\lambda(z,D(Q,R))}\mu(Q)^{\frac{1}{2}}\mu(R)^{\frac{1}{2}}\times$$

$$|(f,h_Q^j)||(g,h_R^i)|\mathbf{1}_{R\,\text{is good}} \leqslant$$

$$\leqslant 2\mathbb{E}\sum_{t,s}\sum_{\substack{\ell(Q)=\delta^{-t}\ell(R),\\ D(Q,R)\sim\delta^{-s}\ell(Q),\\ R\cap Q=\emptyset,\\ R,Q\subset Q^{s+s_0+10}}} \delta^{\frac{t\varepsilon}{2}}\left(\frac{\ell(Q)}{D(Q,R)}\right)^{\varepsilon}\frac{\mu(Q)^{\frac{1}{2}}\mu(R)^{\frac{1}{2}}}{\sup_{z\in R}\lambda(z,D(Q,R))}\times$$

$$|(f,h_Q^j)||(g,h_R^i)|\mathbf{1}_{R\,\text{is good}} \leqslant$$

$$\leqslant C2\mathbb{E}\sum_{t,s}\delta^{\frac{t\varepsilon}{2}}\delta^{s\varepsilon}\sum_{\substack{\ell(Q)=\delta^{-t}\ell(R),\\ D(Q,R)\sim\delta^{-s}\ell(Q),\\ R\cap Q=\emptyset,\\ R,Q\subset Q^{s+s_0+10}}} \frac{\mu(Q)^{\frac{1}{2}}\mu(R)^{\frac{1}{2}}}{\sup_{z\in R}\lambda(z,D(Q,R))}|(f,h_Q^j)||(g,h_R^i)|\mathbf{1}_{R\,\text{is good}}. \tag{4.22}$$

We now define S_n as we did before:

$$(S_n u,v) = \sum_{\substack{\ell(Q)=\delta^{-t}\ell(R),\\ D(Q,R)\sim\delta^{-s}\ell(Q),\\ R\cap Q=\emptyset,\\ R,Q\subset Q^{s+s_0+10}}} \pm\frac{\mu(Q)^{\frac{1}{2}}\mu(R)^{\frac{1}{2}}}{\sup_{z\in R}\lambda(z,D(Q,R))}(u,h_Q^j)(v,h_R^i)\mathbf{1}_{R\,\text{is good}}.$$

We need to estimate the coefficient. We write

$$\lambda(z, D(Q,R)) \sim \lambda(z, \delta^{-s}\ell(Q)) \sim \lambda(z, \delta^{-s-s_0-10}\ell(Q)) \sim$$
$$\sim \lambda(z, \ell(Q^{(s+s_0+10)})) \sim \lambda(z, diam(Q^{(s+s_0+10)})) \geqslant \mu(B(z, diam(Q^{(s+s_0+20)}))) \geqslant$$
$$\geqslant \mu(Q^{(s+s_0+10)}), \quad (4.23)$$

and therefore

$$|\pm \frac{\mu(Q)^{\frac{1}{2}}\mu(R)^{\frac{1}{2}}}{\sup_{z\in R}\lambda(z, D(Q,R))}| \leqslant C\frac{\mu(Q)^{\frac{1}{2}}\mu(R)^{\frac{1}{2}}}{\mu(Q^{s+s_0+10})}.$$

We notice that C does not depend on s since we used the doubling property of λ only for transmission from $\delta^{-s}\ell(Q)$ to $\delta^{-s-s_0-10}\ell(Q)$.

We conclude that S_n is a dyadic shift of complexity at most $C(s+t)$. Therefore, see Section 4.4.2,

$$|\Sigma_{out}| \leqslant 2C\mathbb{E}\sum_{t,s}\delta^{\frac{t\varepsilon}{2}}\delta^{s\varepsilon}(s+t)^a[w]_2\|f\|_w\|g\|_{w^{-1}} \leqslant C[w]_2\|f\|_w\|g\|_{w^{-1}},$$

and our proof is completed.

4.4 The rest of the proof

4.4.1 Paraproducts and Bellman function

Now we will prove the Lemma 4.3.4.

We remind that the quadratic form of our paraproduct π is the following:

$$(\pi(f), g) := \sum_R \sum_i \langle f\rangle_{\mu,R}(T\chi_X, h^i_R)(g, h^i_R)\,.$$

Operator T is bounded in $L^2(\mu)$ and μ is doubling. Therefore, it is well known that coefficients $b_R := b^i_R := (T\chi_X, h^i_R)$ satisfy Carleson condition for any of our lattices of Christ's dyadic cubes:

$$\forall Q \in \mathcal{D} \quad \sum_{R\in\mathcal{D},\, R\subset Q} |b_R|^2 \le B\,\mu(Q)\,. \tag{4.24}$$

The best constant B here is called the Carleson constant and it is denoted by $\|b\|_C$. It is known that for our $b_R := (T\chi_X, h^i_R)$ the Carleson constant is bounded by $B_T := C\,\|T\|_{L^2(\mu)\to L^2(\mu)}$.

If we would be on the line with Lebesgue measure μ and w would be a usual weight in A_2, then the sum would follow the estimate of O. Beznosova [B]:

$$|\pi_{T\chi_X}(f, g)| \le C\,\sqrt{B_T}[w]_{A_2}\,. \tag{4.25}$$

But the same is true in our situation. To prove that, one should analyze the proof in [B] and see that it used always conditions on w and b separately. They were always split by Cauchy–Schwarz inequality. The only inequality, where w and b meet was of the type: let Q be a Christ's cube of a certain lattice, then

$$\sum_{R\subset Q,\, R\in\mathcal{D}} \langle w\rangle_{\mu,R} b_R^2 \le [w]_{A_\infty}\|b\|_C \int_Q w\,d\mu\,, \tag{4.26}$$

where

$$[w]_{A_\infty} = \sup \frac{1}{\mu(B)} \int\limits_B w d\mu \cdot \exp\left(-\frac{1}{\mu(B)} \int\limits_B w d\mu \right).$$

Let us explain the last inequality. We write

$$\langle w \rangle_{\mu,R} \leqslant [w]_{A_\infty} \cdot \exp\left(\langle w \rangle_{\mu,R}\right) = [w]_{A_\infty} \cdot \exp\left(2\langle w^{\frac{1}{2}} \rangle_{\mu,R} \right) \leqslant [w]_{A_\infty} \langle w^{\frac{1}{2}} \rangle^2_{\mu,R}$$
$$\leqslant [w]_{A_\infty} \inf_{x\in R} M(w^{\frac{1}{2}} \chi_R)^2.$$

Finally, we notice that $\{b_R^2\}$ is a Carleson sequence, and finish our explanation with the following well known theorem.

Theorem 4.4.1. *Suppose $\{\alpha_K\}$ is a Carleson sequence. Then for any positive function F the following inequality holds:*

$$\sum_K \alpha_K \inf_K F(x) \leqslant \int F(x) d\mu(x).$$

In all other estimates in [B] the sums with $\Delta_Q w$ (see the definition before Lemma 3.2 of [NV]) and the sums with b are always estimated separately. The sums where the terms contain the product of $\Delta_Q w$ and b_Q never got estimated by Bellman technique: they got split first. Then (4.25) follows in our metric situation as well.

4.4.2 Weighted estimates for dyadic shifts via Bellman function

This section is here just for the sake of completeness. In fact, it just repeats the article of Nazarov–Volberg [NV]. In this section we prove the following theorem.

Theorem 4.4.2. *Let $\mathbb{S}_{m,n}$ be a dyadic shift of complexity $m+n+1$. Then*

$$\|\mathbb{S}_{m,n}\|_{wd\mu} \leqslant C(m+n+1)^a[w]_{2,\mu}.$$

Remark 19. We notice that the best known a is equal to one. It can be gotten using the technique from [HLM+] or from [T]. However, for the application we made in the previous sections, namely, the linear A_2 bound for an arbitrary Calderón–Zygmund operator on geometrically doubling metric space, the actual value of a is not important.

We denote $\sigma = w^{-1}$. We begin with the following famous lemma.

Lemma 4.4.3.

$$h_I^j = \alpha_I^j h_I^{w,j} + \beta_I^j \chi_I\,,$$

where

1) $|\alpha_I^j| \leq \sqrt{\langle w\rangle_{\mu,I}}$,

2)$|\beta_I^j| \leq \frac{|(h_I^{w,j},w)_\mu|}{w(I)}$, *where* $w(I) := \int_I w\,d\mu$,

3) $\{h_I^{w,j}\}_I$ *is supported on* I, *orthogonal to constants in* $L^2(w\,d\mu)$,

4) $h_I^{w,j}$ *assumes on each son* $s_i(I)$ *a constant value,*

5) $\|h_I^{w,j}\|_{L^2(w\mu)} = 1$.

Definition. Let

$$\Delta_I w := \sum_{\text{sons of } I} |\langle w\rangle_{\mu,s(I)} - \langle w\rangle_{\mu,I}|\,.$$

It is a easy to see that the doubling property of measure μ implies

$$|(h_I^{w,j}, w)_\mu| \leq C\,(\Delta_I w)\,\mu(I)^{1/2}\,. \tag{4.27}$$

Therefore, the property 2) above can be rewritten as

2') $|\beta_I^j| \le C\,\frac{|\Delta_I w|}{\langle w\rangle_{\mu,I}}\frac{1}{\mu(I)^{1/2}}$.

Fix $\phi \in L^2(w\,d\mu), \psi \in L^2(\sigma d\mu)$. We need to prove

$$|(\mathbb{S}_{m,n}\phi w, \psi\sigma)| \le C\,(n+m+1)^a \|\phi\|_w \|\psi\|_\sigma\,. \tag{4.28}$$

Remark 20. In next calculations we drop the superscript i and j in Haar functions h_I^i and $h_I^{w,i}$. The reader should always assume that we sum up over all i's.

We estimate $(\mathbb{S}_{m,n}\phi w, \psi\sigma)$ as

$$|\sum_L \sum_{I,J} c_{L,I,J}(\phi w, h_I)_\mu (\psi\sigma, h_J)_\mu| \le$$

$$\sum_L \sum_{I,J} |c_{L,I,J}(\phi w, h_I^w)_\mu \sqrt{\langle w\rangle_{\mu,I}}(\psi\sigma, h_J^\sigma)_\mu| \sqrt{\langle\sigma\rangle_{\mu,J}}| +$$

$$\sum_L \sum_{I,J} |c_{L,I,J}\langle\phi w\rangle_{\mu,I}\frac{\Delta_I w}{\langle w\rangle_{\mu,I}}(\psi\sigma, h_J^\sigma)_\mu \sqrt{\langle\sigma\rangle_{\mu,J}}\sqrt{I}| +$$

$$\sum_L \sum_{I,J} |c_{L,I,J}\langle\psi\sigma\rangle_{\mu,J}\frac{\Delta_J \sigma}{\langle \sigma\rangle_{\mu,J}}(\phi w, h_I^w)_\mu \sqrt{\langle w\rangle_{\mu,I}}\sqrt{J}| +$$

$$\sum_L \sum_{I,J} |c_{L,I,J}\langle\phi w\rangle_{\mu,I}\langle\psi\sigma\rangle_{\mu,J}\frac{\Delta_I w}{\langle w\rangle_{\mu,I}}\frac{\Delta_J \sigma}{\langle \sigma\rangle_{\mu,J}}\sqrt{I}\sqrt{J}| =: I + II + III + IV\,.$$

We can notice that because $|c_{L,I,J}| \le \frac{\sqrt{\mu(I)}\sqrt{\mu(J)}}{\mu(L)}$ each sum inside L can be estimated by a perfect product of S and R terms, where

$$R_L(\phi w) := \sum_{I\subset L\ldots} \langle\phi w\rangle_{\mu,I}\frac{|\Delta_I w|}{\langle w\rangle_{\mu,I}}\frac{\mu(I)}{\sqrt{\mu(L)}}$$

$$S_L(\phi w) := \sum_{I\subset L...} (\phi w, h_I^w)_\mu \sqrt{\langle w\rangle_{\mu,I}} \frac{\sqrt{\mu(I)}}{\sqrt{\mu(L)}}$$

and the corresponding terms for $\psi\sigma$. So we have

$$I \le \sum_L S_L(\phi w) S_L(\psi\sigma),\ II \le \sum_L S_L(\phi w) R_L(\psi\sigma),$$

$$III \le \sum_L R_L(\phi w) S_L(\psi\sigma),\ \ IV \le \sum_L R_L(\phi w) R_L(\psi\sigma)\,.$$

Now

$$S_L(\phi w) \le \sqrt{\sum_{I\subset L...} |(\phi w, h_I^w)_\mu|^2} \sqrt{\langle w\rangle_{\mu,L}}\,,\ \ S_L(\psi\sigma) \le \sqrt{\sum_{J\subset L...} |(\psi\sigma, h_J^\sigma)|^2} \sqrt{\langle\sigma\rangle_{\mu,L}} \quad (4.29)$$

Therefore,

$$I \leqslant C[w]_{A_2}^{1/2} \|\phi\|_w \|\psi\|_\sigma\,. \quad (4.30)$$

Terms II, III are symmetric, so consider III. Using Bellman function $(xy)^\alpha$ one can prove now

Lemma 4.4.4. *Let $Q := [w]_{A_2}$ and $\alpha \in (0, 1/2)$.*

The sequence

$$\tau_I := \langle w\rangle_{\mu,I}^\alpha \langle\sigma\rangle_{\mu,I}^\alpha \left(\frac{|\Delta_I w|^2}{\langle w\rangle_{\mu,I}^2} + \frac{|\Delta_I \sigma|^2}{\langle\sigma\rangle_{\mu,I}^2} \right) \mu(I)$$

form a Carleson measure with Carleson constant at most $c_\alpha Q^\alpha$.

Proof. We need a very simple

Sublemma. Let $Q > 1$, $0 < \alpha < \frac{1}{2}$. In domain $\Omega_Q := \{(x, y) : X > o, y > 0, 1 < xy \le Q$

the function $B_Q(x,y) := x^\alpha y^\alpha$ satisfies the following estimate of its Hessian matrix:

$$-d^2 B_Q(x,y) \geq \alpha(1-2\alpha)x^\alpha y^\alpha \bigg(\frac{(dx)^2}{x^2} + \frac{(dy)^2}{y^2}\bigg)\,.$$

The form $-d^2 B_Q(x,y) \geq 0$ everywhere in $x>0, y>0$. Also obviously $0 \leq B_Q(x,y) \leq Q^\alpha$ in Ω_Q.

Proof. Direct calculation. □

Fix now a Christ's cube I and let $s_i(I), i = 1, ..., M$, be all its sons. Denote $a = (\langle w\rangle_{\mu,I}, \langle\sigma\rangle_{\mu,I})$, $b_i = (\langle w\rangle_{\mu,s_i(I)}, \langle\sigma\rangle_{\mu,s_i(I)})$, $i = 1,\dots,M$, be points–obviously–in Ω_Q, where Q temporarily means $[w]_{A_2}$. Consider $c_i(t) = a(1-t)+b_it, 0 \leq t \leq 1$ and $q_i(t) := B_Q(c_i(t))$. We want to use Taylor's formula

$$q_i(0) - q_i(1) = -q_i'(0) - \int_0^1 dx \int_0^x q_i''(t)\,dt\,. \tag{4.31}$$

Notice two things: Sublemma shows that $-q_i''(t) \geq 0$ everywhere. Moreover, it shows that if $t \in [0,1/2]$, then the following qualitative estimate holds

$$-q_i''(t) \geq c\,(\langle w\rangle_{\mu,I}\langle\sigma\rangle_{\mu,I})^\alpha \bigg(\frac{(\langle w\rangle_{\mu,s_i(I)} - \langle w\rangle_{\mu,I})^2}{\langle w\rangle^2_{\mu,I}} + \frac{(\langle\sigma\rangle_{\mu,s_i(I)} - \langle\sigma\rangle_{\mu,I})^2}{\langle\sigma\rangle^2_{\mu,I}}\bigg) \tag{4.32}$$

This requires a small explanation. If we are on the segment $[a,b_i]$, then the first coordinate of such a point cannot be larger than $C\,\langle w\rangle_{\mu,I}$, where C depends only on doubling of μ (not w). This is obvious. The same is true for the second coordinate with the obvious change of w to σ. But there is no such type of estimate from below on this segment: the first coordinate cannot be smaller than $k\,\langle w\rangle_{\mu,I}$, but k may (and will) depend on the doubling of w (so ultimately on its $[w]_{A_2}$ norm. In fact, at the "right" endpoint of $[a, b_i]$. The first coordinate is $\langle w\rangle_{\mu,s_i(I)} \le \int_I w\, d\mu/\mu(s_i(I)) \le C\, \int_I w\, d\mu/\mu(I)) = C\,\langle w\rangle_{\mu,I}$, with C only depending on the doubling of μ. But the estimate from below will involve the doubling of w, which we must avoid. But if $t\in[0,1/2]$, and we are on the "left half" of interval $[a,b_i]$ then obviously the first coordinate is $\ge \frac12\langle w\rangle_{\mu,I}$ and the second coordinate is $\ge \frac12\langle\sigma\rangle_{\mu,I}$.

We do not need to integrate $-q_i''(t)$ for all $t\in[0,1]$ in (4.31). We can only use integration over $[0,1/2]$ noticing that $-q_i''(t)\ge 0$ otherwise. Then the chain rule

$$q_i''(t) = (B_Q(c_i(t))'' = (d^2 B_Q(c_i(t)(b_i-a), b_i-a)$$

immediately gives us (4.32) with constant c depending on the doubling of μ but *independent* of the doubling of w.

Next step is to add all (4.31), with convex coefficients $\frac{\mu(s_i(I))}{\mu(I)}$, and to notice that $\sum_{i=1}^M \frac{\mu(s_i(I))}{\mu(I)} q_i'(0) = \nabla B_Q(a)\sum_{i=1}^M \cdot (a-b_i)\frac{\mu(s_i(I))}{\mu(I)} = 0$, because by definition

$$a = \sum_{i=1}^M \frac{\mu(s_i(I))}{\mu(I)}\, b_i\,.$$

Notice that the addition of all (4.31), with convex coefficients $\frac{\mu(s_i(I))}{\mu(I)}$ gives us now (we take into account (4.32) and positivity of $-q_i''(t)$)

$$B_Q(a) - \sum_{i=1}^M \frac{\mu(s_i(I))}{\mu(I)}\, B_Q(b_i) \ge$$
$$c\, c_1\,(\langle w\rangle_{\mu,I}\langle\sigma\rangle_{\mu,I})^{\alpha} \sum_{i=1}^M \Big(\frac{(\langle w\rangle_{\mu,s_i(I)} - \langle w\rangle_{\mu,I})^2}{\langle w\rangle_{\mu,I}^2} + \frac{(\langle \sigma\rangle_{\mu,s_i(I)} - \langle \sigma\rangle_{\mu,I})^2}{\langle \sigma\rangle_{\mu,I}^2}\Big)\,.$$

We used here the doubling of μ again, by noticing that $\frac{\mu(s_i(I))}{\mu(I)} \geq c_1$ (recall that $s_i(I)$ and I are almost balls of comparable radii). We rewrite the previous inequality using our definition of $\Delta_I w, \Delta_I \sigma$ listed above as follows

$$\mu(I)\, B_Q(a) - \sum_{i=1}^{M} \mu(s_i(I))\, B_Q(b_i) \geq c\, c_1\, (\langle w\rangle_{\mu,I}\langle\sigma\rangle_{\mu,I})^{\alpha} \left(\frac{(\Delta_I w)^2}{\langle w\rangle_{\mu,I}^2} + \frac{(\Delta_I \sigma)^2}{\langle\sigma\rangle_{\mu,I}^2} \right) \mu(I)\,.$$

Notice that $B_Q(a) = \langle \big(w\rangle_{\mu,I}\langle\sigma\rangle_{\mu,I}\big)^{\alpha}$. Now we iterate the above inequality and get for any of Christ's dyadic I's:

$$\sum_{J\subset I\,, J\in\mathcal{D}} (\langle w\rangle_{\mu,J}\langle\sigma\rangle_{\mu,J})^{\alpha} \left(\frac{(\Delta_J w)^2}{\langle w\rangle_{\mu,J}^2} + \frac{(\Delta_J \sigma)^2}{\langle\sigma\rangle_{\mu,J}^2} \right) \mu(J) \leq C\, Q^{\alpha} \mu(I)\,.$$

This is exactly the Carleson property of the measure $\{\tau_I\}$ indicated in our Lemma 4.4.4, with Carleson constant $C\, Q^{\alpha}$. The proof showed that C depended only on $\alpha \in (0, 1/2)$ and on the doubling constant of measure μ.

□

Now, using this lemma, we start to estimate our S_L's and R_L's. For $S_L(\psi\sigma)$ we already had estimate (4.29).

To estimate $R_L(\phi w)$ let us denote by $\mathcal{P}_L$ maximal stopping intervals $K \in \mathcal{D}, K \subset L$, where the stopping criteria are 1) either $\frac{|\Delta_K w|}{\langle w\rangle_{\mu,K}} \geq \frac{1}{m+n+1}$, or $\frac{|\Delta_K \sigma|}{\langle\sigma\rangle_{\mu,K}} \geq \frac{1}{m+n+1}$, or 2) $g(K) = g(L) + m$.

Lemma 4.4.5. *If K is any stopping interval then*

$$\sum_{I\subset K,\ell(I)=2^{-m}\ell(L)} |\langle \phi w\rangle_{\mu,I}| \frac{|\Delta_I w|}{\langle w\rangle_{\mu,I}} \frac{\mu(I)}{\sqrt{\mu(L)}} \leq 2e^{\alpha}(m+n+1)\langle|\phi|w\rangle_{\mu,K}\frac{\sqrt{\mu(K)}}{\sqrt{\mu(L)}}\sqrt{\tau_K}\langle w\rangle_{\mu,L}^{-\alpha/2}\langle\sigma\rangle_{\mu,L}^{-\alpha/2}. \quad (4.33)$$

Proof. If we stop by the first criterion, then

$$\sum_{I\subset K,\ell(I)=2^{-m}\ell(L)} |\langle \phi w\rangle_{\mu,I}| \frac{|\Delta_I w|}{\langle w\rangle_{\mu,I}} \frac{\mu(I)}{\sqrt{\mu(L)}} \leq 2\sum_{I\subset K,\ell(I)=2^{-m}\ell(L)} |\langle \phi w\rangle_{\mu,I}|\mu(I)\frac{1}{\mu(K)}\frac{\mu(K)}{\sqrt{\mu(L)}} \leq$$

$$\leq 2\,\langle|\phi|w\rangle_{\mu,K}\frac{\mu(K)}{\sqrt{\mu(L)}} \leq 2(m+n+1)\langle|\phi|w\rangle_{\mu,K}\left(\frac{|\Delta_K w|}{\langle w\rangle_{\mu,K}}+\frac{|\Delta_K\sigma|}{\langle\sigma\rangle_{\mu,K}}\right)\frac{\mu(K)}{\sqrt{\mu(L)}} \leq$$

$$\leq 2(m+n+1)\langle|\phi|w\rangle_{\mu,K}\frac{\sqrt{\mu(K)}}{\sqrt{\mu(L)}}\sqrt{\tau_K}\langle w\rangle_{\mu,K}^{-\alpha/2}\langle\sigma\rangle_{\mu,K}^{-\alpha/2}.$$

Now replacing $\langle w\rangle_{\mu,K}^{-\alpha/2}\langle\sigma\rangle_{\mu,K}^{-\alpha/2}$ by $\langle w\rangle_{\mu,L}^{-\alpha/2}\langle\sigma\rangle_{\mu,L}^{-\alpha/2}$ does not grow the estimate by more than e^{α} as all pairs of son/father intervals larger than K and smaller than L will have there averages compared by constant at most $1\pm\frac{1}{m+n+1}$. And there are at most m such intervals between K and L.

If we stop by the second criterion, then K is one of I's, $g(I)=g(L)+m$, and

$$|\langle \phi w\rangle_{\mu,I}| \frac{|\Delta_I w|}{\langle w\rangle_{\mu,I}} \frac{\mu(I)}{\sqrt{\mu(L)}} \leq |\langle \phi w\rangle_{\mu,K}|\frac{\mu(K)}{\sqrt{\mu(L)}}\frac{|\Delta_K w|}{\langle w\rangle_{\mu,K}} \leq \langle|\phi|w\rangle_{\mu,K}\frac{\sqrt{\mu(K)}}{\sqrt{\mu(L)}}\sqrt{\tau_K}\langle w\rangle_{\mu,K}^{-\alpha/2}\langle\sigma\rangle_{\mu,K}^{-\alpha/2}.$$

Now we replace $\langle w\rangle_{\mu,K}^{-\alpha/2}\langle\sigma\rangle_{\mu,K}^{-\alpha/2}$ by $\langle w\rangle_{\mu,L}^{-\alpha/2}\langle\sigma\rangle_{\mu,L}^{-\alpha/2}$ as before.

$\square$

Now

$$R_L(\phi w) \le C(m+n+1)\langle w\rangle_{\mu,L}^{-\alpha/2}\langle\sigma\rangle_{\mu,L}^{-\alpha/2}\sum_{K\in\mathcal{P}_L}\langle|\phi|w\rangle_{\mu,K}\frac{\sqrt{\mu(K)}}{\sqrt{\mu(L)}}\sqrt{\tau_K}$$

$$\le C(m+n+1)\langle w\rangle_{\mu,L}^{-\alpha/2}\langle\sigma\rangle_{\mu,L}^{-\alpha/2}\Big(\sum_{K\in\mathcal{P}_L}\langle|\phi|w\rangle_{\mu,K}^2\frac{\mu(K)}{\mu(L)}\Big)^{1/2}(\widetilde{\tau}_L)^{1/2},$$

where

$$\widetilde{\tau}_L = \sum_{K\in\mathcal{P}_L}\tau_K\,.$$

Notice that the sequence $\{\widetilde{\tau}_L\}_{L\in\mathcal{D}}$ form a Carleson sequence (measure) with constant at most $C(m+1)Q^\alpha$.

Now we make a trick! We will estimate the right hand side as

$$R_L(\phi w) \le C(m+n+1)\langle w\rangle_{\mu,L}^{-\alpha/2}\langle\sigma\rangle_{\mu,L}^{-\alpha/2}\Big(\sum_{K\in\mathcal{P}_L}\langle|\phi|w\rangle_{\mu,K}^p\frac{\mu(K)}{\mu(L)}\Big)^{1/p}(\widetilde{\tau}_L)^{1/2},$$

where $p = 2-\frac{1}{m+n+1}$. In fact,

$$\Big(\sum_{K\subset L,\, K\, is\ maximal}\langle|\phi|w\rangle_{\mu,K}^2\frac{\mu(K)}{\mu(L)}\Big)^{p/2} \le \sum_{K\in\mathcal{P}_L}\langle|\phi|w\rangle_{\mu,K}^p\Big(\frac{\mu(K)}{\mu(L)}\Big)^{p/2}.$$

But if if $0 \le j \le m$, then $(C^{-j})^{-\frac{1}{m+n+1}} \le C$, and therefore in the formula above $\left(\frac{\mu(K)}{\mu(L)}\right)^{1-\frac{1}{2(m+n+1)}} \le C\,\frac{\mu(K)}{\mu(L)}$, and C depends only on the doubling constant of μ. So the trick is justified. Therefore, using Cauchy inequality, one gets

$$R_L(\phi w) \le C(m+n+1)\langle w\rangle_{\mu,L}^{-\alpha/2}\langle \sigma\rangle_{\mu,L}^{-\alpha/2}\Big(\sum_{K\in\mathcal{P}_L}\langle|\phi|^p w\rangle_{\mu,K}\langle w\rangle_{\mu,K}^{p-1}\frac{\mu(K)}{\mu(L)}\Big)^{1/p}(\widetilde{\tau}_L)^{1/2}\,.$$

We can replace all $\langle w\rangle_{\mu,K}^{p-1}$ by $\langle w\rangle_{\mu,L}^{p-1}$ paying the price by constant. This is again because all intervals larger than K and smaller than L will have there averages compared by constant at most $1\pm\frac{1}{m+n+1}$. And there are at most m such intervals between K and L. Finally,

$$R_L(\phi w) \le C(m+n+1)\langle w\rangle_{\mu,L}^{-\alpha/2}\langle \sigma\rangle_{\mu,L}^{-\alpha/2}\Big(\sum_{K\in\mathcal{P}_L}\langle|\phi|^p w\rangle_{\mu,K}\frac{\mu(K)}{\mu(L)}\Big)^{1/p}\langle w\rangle_{\mu,L}^{1-\frac{1}{p}}(\widetilde{\tau}_L)^{1/2} \tag{4.34}$$

We need the standard notations: if ν is an arbitrary positive measure we denote

$$M_\nu f(x) := \sup_{r>0}\frac{1}{\nu(B(x,r))}\int_{B(x,r)}|f(x)|\,d\nu(x)\,.$$

In particular M_w will stand for this maximal function with $d\nu = w(x)\,d\mu$.

From (4.34) we get

$$R_L(\phi w) \le C(m+n+1)\langle w\rangle_{\mu,L}^{1-\alpha/2}\langle \sigma\rangle_{\mu,L}^{-\alpha/2}\inf_L\, M_w(|\phi|^p)^{1/p}(\widetilde{\tau}_L)^{1/2} \tag{4.35}$$

Now

$$S_L(\psi\sigma)R_L(\phi w) \le C(m+n+1)\langle w\rangle_{\mu,L}^{1-\alpha/2}\langle \sigma\rangle_{\mu,L}^{1-\alpha/2}\frac{\inf_L\, M_w(|\phi|^p)^{1/p}}{\langle\sigma\rangle_{\mu,L}^{1/2}}\times (\widetilde{\tau}_L)^{1/2}\sqrt{\sum_{J\subset L\ldots}|(\psi\sigma,h_J^\sigma)|^2}\,, \tag{4.36}$$

$$R_L(\psi\sigma)R_L(\phi w) \le C(m+n+1)\langle w\rangle_{\mu,L}^{1-\alpha}\langle\sigma\rangle_{\mu,L}^{1-\alpha}\inf_L\, M_w(|\phi|^p)^{1/p}\inf_L\, M_\sigma(|\psi|^p)^{1/p}\widetilde{\tau}_L\,. \quad (4.37)$$

Now we use the Carleson property of $\{\widetilde{\tau}_L\}_{L\in\mathcal{D}}$. We need a simple folklore Lemma.

Lemma 4.4.6. *Let $\{\alpha_L\}_{L\in\mathcal{D}}$ define Carleson measure with intensity B related to dyadic lattice $\mathcal{D}$ on metric space X. Let F be a positive function on X. Then*

$$\sum_L (\inf_L F)\,\alpha_L \le 2B\int_X F\,d\mu\,. \quad (4.38)$$

$$\sum_L \frac{\inf_L F}{\langle\sigma\rangle_{\mu,L}}\alpha_L \le C\,B\int_X \frac{F}{\sigma}d\mu\,. \quad (4.39)$$

Now use (4.36). Then the estimate of $III \le \sum_L S_L(\psi\sigma)R_L(\phi w)$ will be reduced to estimating

$$(m+n+1)Q^{1-\alpha/2}\Bigg(\sum_L \frac{\inf_L M_w(|\phi|^p)^{2/p}}{\langle\sigma\rangle_{\mu,L}}\widetilde{\tau}_L\Bigg)^{1/2} \le$$

$$(m+n+1)^2\,Q\left(\int_{\mathbb{R}}(M_w(|\phi|^p))^{2/p}w d\mu\right)^{1/2}$$

$$\le (\frac{1}{2-p})^{1/p}(m+n+1)^2\,Q\left(\int_{\mathbb{R}}\phi^2\,wd\mu\right)^{1/2} \le (m+n+1)^3\,Q\left(\int_{\mathbb{R}}\phi^2\,wd\mu\right)^{1/2}.$$

Here we used (4.39) and the usual estimates of maximal function M_μ in $L^q(\mu)$ when $q\approx 1$. Of course for II we use the symmetric reasoning.

Now IV: we use (4.37) first.

$$\sum_L R_L(\psi\sigma)R_L(\phi w) \le (m+n+1)Q^{1-\alpha}\sum_L \inf_L\, M_w(|\phi|^p)^{1/p}\inf_L\, M_\sigma(|\psi|^p)^{1/p}\widetilde{\tau}_L$$

$$\le C(m+n+1)^2 Q\int_{\mathbb{R}}(M_w(|\phi|^p))^{1/p}\,(M_\sigma(|\psi|^p))^{1/p}w^{1/2}\sigma^{1/2}d\mu$$

$$\leq C(m+n+1)^2 Q \left(\int_{\mathbb{R}} (M_w(|\phi|^p))^{2/p} \, w d\mu \right)^{1/2} \left(\int_{\mathbb{R}} (M_\sigma(|\psi|^p))^{2/p} \, \sigma d\mu \right)^{1/2}$$

$$\leq C(m+n+1)^4 \, Q \left(\int_{\mathbb{R}} \phi^2 \, w d\mu \right)^{1/2} \left(\int_{\mathbb{R}} \psi^2 \, \sigma d\mu \right)^{1/2}.$$

Here we used (4.38) and the usual estimates of maximal function M_μ in $L^{2/p}(\mu)$ when $p \approx 2$, $p < 2$.

BIBLIOGRAPHY

BIBLIOGRAPHY

[R1] Reznikov, Alexander **Sharp constants in the Paneyah-Logvinenko-Sereda theorem** C. R. Math. Acad. Sci. Paris 348 (2010), no. 3-4, 141–144.

[NR1] Nazarov, A. I.; Reznikov, A. B. **On the existence of an extremal function in critical Sobolev trace embedding theorem** J. Funct. Anal. 258 (2010), no. 11, 3906–3921.

[NR2] Nazarov A., Reznikov A., **Attainability of infima in the critical Sobolev trace embedding theorem on manifolds** — American Mathematical Society Translations–Series 2 Advances in the Mathematical Sciences 2010; 252 pp; hardcover Volume: 229.

[R2] Reznikov, Alexander **Sharp weak type estimates for weights in the class** A_{p_1,p_2}, Rev. Mat. Iberoam. **29** (2013), no. 2, 433–478; doi 10.4171/rmi/726; arXiv:1105.4848v1.

[RVV] Reznikov A., Vasyunin V., Volberg A. **An observation: cut-off of the weight** w **does not increase the** A_{p_1,p_2}**-"norm" of** w arXiv:1008.3635

[BR1] Beznosova O., Reznikov A. **Equivalent definitions of dyadic Muckenhoupt and Reverse Hölder classes in terms of Carleson sequences, weak classes, and comparability of dyadic** $L\log L$ **and** A_∞ **constants**, arXiv:1201.0520; accepted to Revista Matematica Iberoamericana.

[BR2] Beznosova O., Reznikov A. **Sharp estimates involving** A_∞ **and** $L\log L$ **constants, and their applications to PDE**, arXiv:1107.1885; accepted to the St. Petersburg Math Journal.

[RTV] Reznikov A., Treil S., Volberg A. **A sharp estimate of weighted dyadic shifts of complexity 0 and 1**, arXiv:1104.5347.

[RV2] Reznikov A., Volberg A. **Random "dyadic" lattice in geometrically doubling metric space and** A_2 **conjecture**, arXiv:1103.5246;

[NRV] Nazarov F., Reznikov A., Volberg A. **The proof of A_2 conjecture in a geometrically doubling metric space**, arXiv:1106.1342; accepted to the Indiana University Math Journal.

[NRVV] Nazarov F., Reznikov A., Vasyunin V., Volberg V. **A_1 conjecture and Bellman functions**, 2010. Available at http://sashavolberg.files.wordpress.com/2010/11/a11_7loghilb11_21_2010.pdf

[CURV] Cruz-Uribe D., Reznikov A., Volberg A. **Logarithmic bump conditions and the two weight boundedness of Caldern-Zygmund operators**, arXiv:1112.0676; submitted to the Advances in Mathematics.

[NRTV1] Nazarov, F., Reznikov, A., Treil, S., Volberg, A. **A Bellman function proof of the L^2 bump conjecture**, arXiv:1202.2406; accepted to the Journal d'Analyse Mathematique.

[NRTV2] Nazarov, F., Reznikov, A., Treil, S., Volberg, A. **Carleson–Buckley measures beyond the scope of A_∞ and their applications**, arXiv:1202.2931.

[NRV1] Nazarov F., Reznikov A., Volberg A., **Bellman approach to the one-sided bumping for weighted estimates of Calderón–Zygmund operators**, arXiv:1306.2653.

[RV] Reznikov, A., Volberg, A. **Cauchy independent measures and super-additivity of analytic capacity**, arXiv:1211.2675; submitted to the Math Research Letters.

[ACUM] Anderson T., Cruz-Uribe D., Moen K., **Logarithmic bump conditions for Calderón-Zygmund Operators on spaces of homogeneous type**, arXiv:1308.2026.

[B] O. Beznosova, Linear bound for the dyadic paraproduct on weighted Lebesgue space L2(w). J. Funct. Anal. 255 (2008), no. 4, 994–1007

[BMP] Beznosova O., Moraes J.C., Pereyra M.C., **Sharp bounds for T-Haar multipliers on L_2.** To appear in Proceedings El Escorial 2012

[Chr] Christ M., **A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral**, Colloq. Math. **60/61** (1990), no. 2, 601-628;

[CUMP] Cruz-Uribe D., Martell J. M. , Pérez C., **Sharp weighted estimates for classical operators**, Adv. in Math., 229 (2012), 408–441.

[1] D. Cruz-Uribe, J. M. Martell, C. Pérez, *Weights, Extrapolation and the Theory of Rubio de Francia*, Operator Theory: Advances and Applications, 215, Birkhauser, Basel, (2011).

[CWW] Chang A., Wilson M., Wolff T., **Some weighted norm inequalities concerning the Schrödinger operators**, Comment. Math. Helvetici, 60 (1985), 217–286.

[CUMP1] Cruz-Uribe D., Martell J. M., Pérez C., **Weights, Extrapolation and the Theory of Rubio de Francia**, Operator Theory: Advances and Applications, 215, Birkhauser, Basel, (2011)

[F] Fefferman C., **The uncertainty principle**, Bull. Amer. Math. Soc., 9 (1983), 129–206

[HavJ] Havin, V., Jöricke, B. **The uncertainty principle in harmonic analysis.** Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], 28. Springer-Verlag, Berlin, 1994. xii+543 pp. ISBN: 3-540-56991-X

[H] Hytönen T., **The sharp weighted bound for general Calderón-Zygmund operators**, arXiv:1007.4330v1;

[HyLa] Hytönen T., Lacey M., **The A_p-A_∞ inequality for general Calderón–Zygmund operators**, arXiv:1106.4797;

[HLM+] Hytönen, T., Lacey, M., Martikainen, H., Orponen, T., Reguera, M., Sawyer, E., Uriarte-Tuero, I., **Weak and strong type estimates for maximal truncations of Calderòn-Zygmund operators on A_p weighted spaces.** J. Anal. Math. 118 (2012), no. 1, 177–220;

[HM] Hytönen T., Martikainen H., **Non-homogeneous Tb theorem and random dyadic cubes on metric measure spaces**, arXiv:0911.4387;

[HP] Hytönen, T. Pérez, C. **Sharp weighted bounds involving A_∞**. Anal. PDE 6 (2013), no. 4, 777–818.

[HPTV] Hytönen T., Pérez C., Treil S., Volberg A., **Sharp weighted estimates for dyadic shifts and the A_2 conjecture**, arXiv:1010.0755v2;

[KV] Konyagin S., Volberg A., **On measures with the doubling condition.** (Russian) Izv. Akad. Nauk SSSR Ser. Mat. **51** (1987), no. 3, 666–675; translation in Math. USSR-Izv. **30** (1988), no. 3, 629–638;

[Lac1] Lacey, M. **The Two Weight Inequality for the Hilbert Transform: A Primer**, arXiv:1304.5004;

[Lac2] Lacey, M. **Two Weight Inequality for the Hilbert Transform: A Real Variable Characterization, II**, arXiv:1301.4663;

[Lac3] Lacey, M. **On the A_2 inequality for Calderón–Zygmund Operators**, arXiv:1106.4802;

[LPR] Lacey M., Petermichl S., Reguera M., **Sharp A_2 inequality for Haar shift operators**, arXiv:0906.1941;

[LSUT] Lacey, M., Sawyer, S., Uriarte-Tuero, **I. A Two Weight Inequality for the Hilbert transform Assuming an Energy Hypothesis**, arXiv:1001.4043;

[LSSUT] Lacey, M., Sawyer, S., Shen, C.Y., Uriarte-Tuero, I. **Two Weight Inequality for the Hilbert Transform: A Real Variable Characterization, I**, arXiv:1201.4319;

[L] Lerner A., **A simple proof of A2 conjecture**, Intern. Math. Res. Notices IMRN, 2012

[L1] Lerner A., **On an estimate of Caldern-Zygmund operators by dyadic positive operators**, arXiv:1202.1860.

[M] H. Martikainen, Vector-valued non-homogeneous Tb theorem on metric measure spaces, arXiv:1004.3176;

[MP] J. Moraes and M. C. Pereyra, *Weighted estimates for dyadic paraproducts and t-Haar multipliers with complexity* (m, n), ArXiv:1108.3109.

[NV] Nazarov, F., Volberg, A. **The Bellman function, the two-weight Hilbert transform, and embeddings of the model spaces K_θ.** Dedicated to the memory of Thomas H. Wolff. J. Anal. Math. 87 (2002), 385414.

[NV1] Nazarov F., Volberg A., **A simple sharp weighted estimate of the dyadic shifts on metric space with geometric doubling**, arxiv 1104.4893;

[NV2] Nazarov F., Volberg A., **On analytic capacity of portions of continuum and a question of T. Murai**, Proc. Sympos. Pure Math., 79, Amer. Math. Soc., Providence, RI, 2008;

[NTV] Nazarov F., Treil S., Volberg A., **The Tb-theorem on non-homogeneous spaces**, Acta Math., 190 (2003), 151-239;

[NTV2] Nazarov, F., Treil, S., Volberg, A. **The Bellman functions and two-weight inequalities for Haar multipliers**, J. Amer. Math. Soc. 12 (1999), no. 4, 909–928

[NTV3] Nazarov F., Treil S., Volberg A., **Two weight inequalities for individual Haar multipliers and other well localized operators**, Math. Res. Lett. **15** (2008), no. 4, 583-597;

[NTV4] Nazarov F., Treil S., Volberg A., **Two weight estimate for the Hilbert transform and corona decomposition for non-doubling measures**, Preprint 2005, arXiv:1003.1596;

[Pan] Panejah, B. **Some theorems of Paley-Wiener type.** Dokl. Akad. Nauk SSSR 138 47–50 (Russian); translated as Soviet Math. Dokl. 2 1961 533536.

[P] C. Pérez, *Weighted norm inequalities for singular integral operators.* J. London Math. Soc. (2) **49** (1994), no. 2, 296–308

[Pe] C. Pérez *On sufficient conditions for the boundedness of the Hardy-Littlewood maximal operator between weighted L^p-spaces with different weights.* Proc. London Math. Soc. **71** (1995), No. 3, 135–157.

[PTV] Pérez C., Treil S., Volberg A., **On A_2 conjecture and corona decomposition of weights**, arXiv:1006.2630;

[Pet1] Petermichl, S. **The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical A_p characteristic.** Amer. J. Math. 129 (2007), no. 5, 1355–1375.

[Pet2] Petermichl, S. **The sharp weighted bound for the Riesz transforms**. Proc. Amer. Math. Soc. 136 (2008), no. 4, 1237–1249.

[RegS] Reguera, M., Scurry, J., **On joint estimates for maximal functions and singular integrals on weighted spaces**. Proc. Amer. Math. Soc. 141 (2013), no. 5, 1705–1717.

[RegTh] Reguera, M., Thiele, C. **The Hilbert transform does not map $L^1(Mw)$ to $L^{1,\infty}(w)$.** Math. Res. Lett. 19 (2012), no. 1, 1–7;

[SlVa] Slavin L., Vasyunin V., **Sharp results in the integral-form John-Nirenberg inequality.** Trans. Amer. Math. Soc. 363 (2011), no. 8, 4135–4169

[St] Stein E., **Note on the class $L \log L$.** Studia Math., 31:305-310, 1969.

[Tol] Tolsa X., **Analytic capacity, the Cauchy transform, and non-homogeneous Calderón-Zygmund theory.** To appear (2012).

[T] Treil S., **Sharp A_2 estimates of Haar shifts via Bellman function**, arXiv:1105.2252v1;

[Va1] Vasyunin V., **The exact constant in the inverse Hölder inequality for Muckenhoupt weights**. St.Petersburg Mathematical Journal, 15:49–79, 2004.

[Va2] Vasyunin V., **Mutual estimates of L^p-norms and the Bellman function** Journal of mathematical science, Volume 156, Number 5, 766–798, DOI: 10.1007/s10958-009-9288-3;

[VaVo] V. Vasyunin, A. Volberg, *The Bellman function for the simplest two-weight inequality: an investigation of a particular case*, St. Petersburg Math. J. 18 (2007), no. 2, 201–222.

[V] Volberg V., **Calderón-Zygmund capacities and operators on nonhomogeneous spaces.** CBMS Regional Conf. Ser. in Math. 100, Amer. Math. Soc., Providence, 2003.

[Wil] Wilson M., **Weighted Littlewood-Paley Theory and Exponential-Square Integrability** Lecture Notes in Mathematics, 1924. Springer, Berlin, 2008. xiv+224 pp. ISBN: 978-3-540-74582-2

[W] Wittwer, J. **A sharp estimate on the norm of the martingale transform**. Math. Res. Lett. 7 (2000), no. 1, 1–12.