The hunt for a Bellman function: applications to estimates of singular integral operators and to other classical problems in harmonic analysis

F. Nazarov and S. Treil

Contents

Notation

$:=$	equal by definition;
$\asymp$	equivalent in a sense of two-sided estimate; $f \asymp g$ (for positive functions f and g) means that there exists a constant C, such that $C^{-1}f \le g \le Cf$
E	finite-dimensional Hilbert space
$\|I\|$	the length of the interval I;
$\langle\varphi\rangle_I$	the average of the function φ over the interval I; $\langle\varphi\rangle_I = \|I\|^{-1}\int_I \varphi(t)dt$
$\langle\varphi\rangle_{I,p}$	the power average of the function φ over the interval I $\langle\varphi\rangle_{I,p} = \Big(\|I\|^{-1}\int_I \varphi(t)^p dt\Big)^{1/p}$;
$\langle\varphi\rangle_{I,0}$	geometric mean of the function φ, $\langle\varphi\rangle_{I,0} = \exp\{\langle\log\varphi\rangle_I\}$;
$\mathcal{L}_+(E)$	the set of all positive ($(Ax,x) > 0\ \forall x \in E\setminus\{0\}$) self-adjoint operators in a Hilbert space E;
$L^p(W)$	weighted L^p space with (matrix-valued) weight W;
L^p_X	L^p space with values in a Banach space X;
$\ell^p(X)$	ℓ^p sequence space with values in a Banach space X;
$Y^\sharp$	the *formal dual* to an ideal sequence Banach space Y (see p. 29). The space $Y^\sharp$ consists of all numerical sequences $\{\beta_j\}_{j\in\mathcal{J}}$ for which the norm $$\\|\{\beta_j\}_{j\in\mathcal{J}}\\|_{Y^\sharp} = \sup\Big\{\sum_{j\in\mathcal{J}} \|\alpha_j\|\cdot\|\beta_j\| \ :\ \{\alpha_j\}_{j\in\mathcal{J}} \in Y\,,\ \\|\{\alpha_j\}_{j\in\mathcal{J}}\\|_Y \le 1\Big\}$$ is finite. Clearly $Y^\sharp \subset Y^*$;
$\\|x^*\\|_{(\mathcal{E}_j)^*}$	the norm of the linear functional $x^* \in X^*$ restricted to a subspace $\mathcal{E}_j \subset X$, see p. 29;
$\mathcal{TL}_p$	Triebel–Lizorkin space, see Section 9.1.2;
$\mathcal{N}_p(E)$	the set of all functions $U : E \to [0,+\infty)$ for which $U^{1/p}$ is a norm on E, see page 75;
$\mathbb{P}\{\dots\}$	the probability of the event $\dots$;
$\mathbb{E}\{\dots\}$	the expectation the event $\dots$;
(f,g)	standard duality: for scalar functions $(f,g) = \int f(t)\overline{g}(t)dt$; for functions with values in a Hilbert space E it is defined as $(f,g) = \int_{\mathbb{R}}(f(t),g(t))_E dt$;

ρ^* the norm dual to the norm ρ, see page 18;

$\langle\rho\rangle^*_{I,p}$ the norm dual to the norm $\langle\rho\rangle_{I,p}$.

1. What is the Bellman function and how to use it

The best way to explain the main idea of the Bellman function approach is to consider some simple and familiar examples first. Let us consider a couple of such examples below.

1.1. Carleson imbedding theorem (dyadic L^2 version)

Let with any dyadic interval I of a real line $\mathbb{R}$ (in this section we assume unless otherwise specified that all intervals are dyadic) we associate a corresponding value $\mu_I \geq 0$. Given a function $\varphi \in L^1_{\text{loc}}$, denote $\langle\varphi\rangle_I := |I|^{-1}\int_I \varphi$. We shall always use the angular brackets $\langle . \rangle_I$ to distinguish the average over the interval I from the case when I is used simply as an index.

The well known Carleson imbedding theorem says that if the values μ_I satisfy the Carleson measure condition, that means

$$\sum_{\substack{J \text{ is dyadic} \\ J\subset I}} \mu_J \leq |I| \qquad \text{for any dyadic interval } I,$$

then for any $\varphi \in L^2$

$$\sum_{I \text{ is dyadic}} \mu_I |\langle\varphi\rangle_I|^2 \leq C\|\varphi\|^2_{L^2} .$$

Suppose that we want to prove this famous theorem, but somehow we have completely forgotten all standard machinery (maximal function, stopping times, interpolation, etc ...) So we are left only with our common sense and a strong belief that the theorem is true. Is it possible to get a proof based only on these two things? And the answer is "yes"!

1.1.1. The Bellman function

Let us use the belief first. Consider for the simplicity the case of real valued functions.

Fix a dyadic interval I and real numbers F, f, M. Consider all sequences $\{\mu_J\}_{J\in\mathcal{D}}$ satisfying the Carleson condition

$$\frac{1}{|J|}\sum_{J'\subset J} \mu_{J'} \leq 1 \qquad \forall J \in \mathcal{D} \tag{1.1}$$

and such that

$$\frac{1}{|I|}\sum_{J\subset I} \mu_J = M\,. \tag{1.2}$$

Consider also all functions $\varphi \in L^2$ for which the quantities

$$\langle\varphi^2\rangle_I := \frac{1}{|I|}\int_I \varphi^2 = F\,, \qquad \langle\varphi\rangle_I := \frac{1}{|I|}\int_I \varphi = f \tag{1.3}$$

are fixed. Then if we believe that the theorem is true, the quantity

$$\mathcal{B}(F,f,M) = \frac{1}{|I|}\sup\Big\{\sum_{J\subset I} \mu_J \cdot \langle\varphi\rangle_J^2 \ : \ \varphi,\ \{\mu_J\}_{J\in\mathcal{D}} \text{ satisfy (1.1), (1.2), (1.3)} \Big\}$$

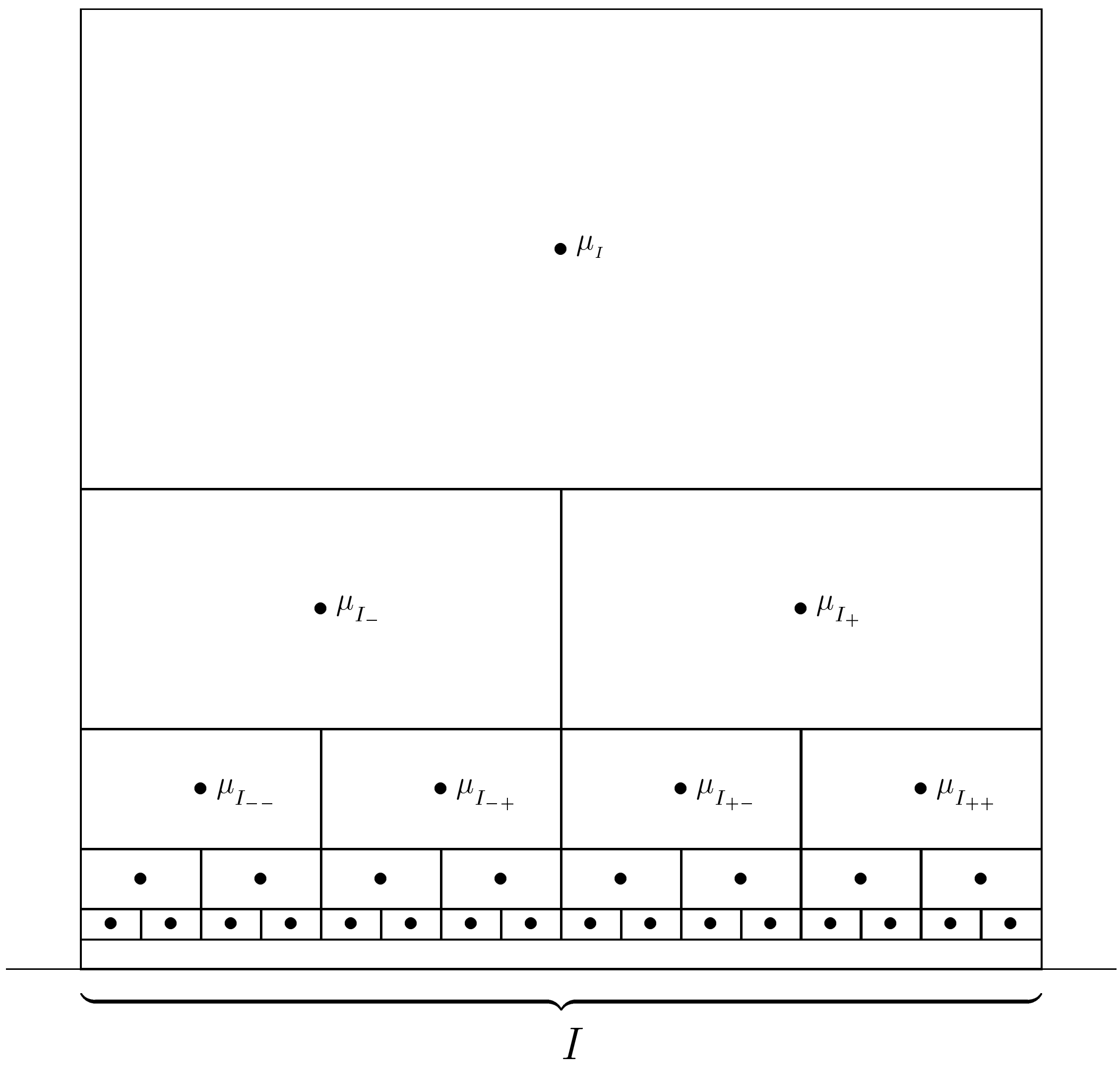

Figure 1: Dyadic Carleson imbedding

is finite, and moreover, satisfies the inequality $\mathcal{B}(F,f,M)\le CF$.

Note that $\mathcal{B}(F,f,M)$ does not depend on the choice of an interval I. Thus we get a function $\mathcal{B}(F,f,M)$ of three real variables F, f, M, which we call the *Bellman function* (associated with the Carleson imbedding theorem). The reader familiar with optimal control can easily recognize the origin of the term.

Now the common sense comes into play:

1.1.2. Properties of $\mathcal{B}(F,f,M)$

1) Domain: $f^2\le F$, $0\le M\le 1$ (Cauchy inequality and the Carleson condition);

2) Range: $0\le \mathcal{B}(F,f,M)\le CF$ (our belief that the theorem is true);

3) Key property: for all $0\le \vartriangle\! M\le M$

$$\mathcal{B}(F,f,M)\ge \vartriangle\! M f^2+\frac{1}{2}\Big\{\mathcal{B}(F_+,f_+,M_+)+\mathcal{B}(F_-,f_-,M_-)\Big\},$$

whenever all triples (F,f,M), $(F_\pm,f_\pm,M_\pm)$ belong to the domain and

$$F=\frac{1}{2}(F_++F_-),\qquad f=\frac{1}{2}(f_++f_-),\quad \text{and}\quad M=\frac{1}{2}(M_++M_-)+\vartriangle\! M\,.$$

Let us explain the properties a little bit. By domain we will always understand the set of variables for which the *supremum* in the definition of the Bellman function is taken over a non-empty set.

It is quite a bit of work to show that for all triples (F,f,M) from our domain one can find functions with corresponding averages. But formally, to prove the Carleson imbedding theorem we need only to be sure that nothing is left outside the domain (there is no φ, $\{\mu_I\}$ with the averages outside the domain), which has already been explained.

To prove the key property 3 we have to look at how the quantities F, f and M can be shared between the left half I_- and the right half I_+ of the interval I. Fix all parameters from 3, and consider all functions φ satisfying

$$\langle\varphi\rangle_{I_+}=f_+\,,\qquad \langle\varphi\rangle_{I_-}=f_-\,,\qquad \langle\varphi^2\rangle_{I_+}=F_+\,,\qquad \langle\varphi^2\rangle_{I_-}=F_-\,.$$

Clearly for any such φ

$$\langle\varphi\rangle_I=\frac{1}{2}\big(f_++f_-\big)=f\,,\qquad \langle\varphi^2\rangle_I=\frac{1}{2}\big(F_++F_-\big)=F\,.$$

Consider also all sequences $\{\mu_J\}$ satisfying the Carleson measure condition (1.1) and

$$\frac{\mu_I}{|I|}=\vartriangle\! M,\qquad \frac{1}{|I_+|}\sum_{J\,:\,J\subset I_+}\mu_J=M_+\,,\qquad \frac{1}{|I_-|}\sum_{J\,:\,J\subset I_-}\mu_J=M_-\,.$$

For any such $\{\mu_J\}$

$$\frac{1}{|I|}\sum_{J\,:\,J\subset I}\mu_J=\frac{1}{2}\big(M_+ + M_-\big)+\frac{\mu_I}{|I|}=\frac{1}{2}\big(M_+ + M_-\big)+\vartriangle M=M\,.$$

Taking $\sup\sum_{J\subset I}\mu_J$ over all φ and $\{\mu_J\}$ as above, we get

$$\frac{1}{2}\Big\{\mathcal{B}(F_+,f_+,M_+)+\mathcal{B}(F_-,f_-,M_-)\Big\}+\frac{\mu_I}{|I|}f^2\,.$$

Since we took the supremum over a smaller set than the set in the definition of $\mathcal{B}$, the above expression is at most $\mathcal{B}(F,f,M)$. Recalling that $\mu_I/|I|=\vartriangle M$, we get the desired property.

1.1.3. Proof of the Carleson imbedding theorem via Bellman type function

And now the critical point comes: note that the whole machine can be run backward. Namely, if we have just any function $\mathcal{B}$ of three real variables satisfying properties 1–3 (we will call any such function a Bellman type function for the Carleson imbedding theorem), the proof of the theorem follows immediately.

Indeed, given a function $\varphi\in L^2$ and a sequence $\{\mu_J\}$ satisfying the Carleson measure condition (1.1), define for a dyadic interval I

$$F_I:=\langle\varphi^2\rangle_I\,,\qquad f_I:=\langle\varphi\rangle_I\,,\qquad M_I:=\frac{1}{|I|}\sum_{J\subset I}\mu_J\,.$$

The key property of $\mathcal{B}$ implies (we again denote by I_+ and I_- the left and right halves of I respectively)

$$\mu_I {f_I}^2\le|I|\cdot\mathcal{B}(F,f,M)-|I_+|\mathcal{B}(F_{I_+},f_{I_+},M_{I_+})-|I_-|\cdot\mathcal{B}(F_{I_-},f_{I_-},M_{I_-});$$

because

$$M_I=\frac{1}{2}(M_{I_+}+M_{I_-})+\frac{\mu_I}{|I|}\,,$$

and all other assumptions of the key property 3 are also satisfied.

Thus, if we draw the picture as on Figure 2, the quantity on each vertical edge does not exceed the difference between the value at the upper vertex and the sum of values at the two lower vertices.

Going n levels down on this picture, we get the inequality

$$\sum_{\substack{J:J\subset I\\ |J|\ge 2^{-n+1}|I|}}\mu_J\cdot(f_J)^2\le|I|\cdot\mathcal{B}(F_I,f_I,M_I)-\sum_{\substack{J:J\subset I\\ |J|=2^{-n}|I|}}|J|\cdot\mathcal{B}(F_J,f_J,M_J)\,.$$

Since $\mathcal{B}(F_J,f_J,M_J)\ge 0$ for every J, this gives

$$\sum_{J\subset I}\mu_J\langle\varphi\rangle_J^2=\sum_{J\subset I}\mu_J\cdot(f_J)^2\le C|I|F_I=C\int_I\varphi^2\,.$$

Applying the above estimate for intervals $[-2^n,0]$ and $[0,2^n]$ and taking the limit as $n\to\infty$, one can immediately obtain the Carleson imbedding theorem. $\square$

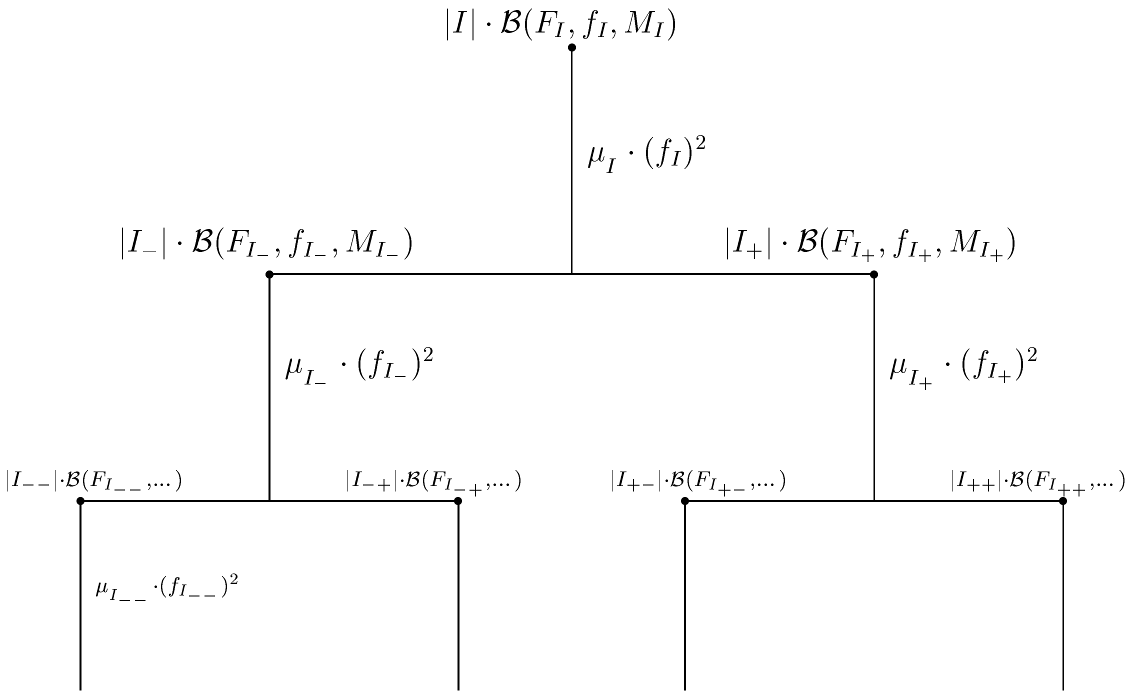

Figure 2: Imbedding theorem from Bellman function. Vertices correspond to dyadic intervals and are denoted by black dots

1.1.4. How to find a Bellman type function?

The key property 3 is not very pleasant to verify. Fortunately we can show that this condition can be replaced by two "infinitesimal" conditions (conditions on derivatives) which are easier to check.

First of all if

$$F=\frac{1}{2}(F_+ + F_-), \qquad f=\frac{1}{2}(f_+ + f_-), \qquad M=\frac{1}{2}(M_+ + M_-) \quad (\text{i.e. } \vartriangle M = 0)\,,$$

and all triples (F, f, M), $(F_\pm, f_\pm, M_\pm)$ are in the domain, then the key property implies concavity of $\mathcal{B}$

$$\mathcal{B}(F,f,M) \geq \frac{1}{2}\Big[\mathcal{B}(F_+,f_+,M_+)+\mathcal{B}(F_-,f_-,M_-)\Big]\,.$$

This means that the second differential $d^2\mathcal{B} \leq 0$, where the second differential can be understood in any reasonable sense.

Put now

$$F_+ = F_- = F, \qquad f_+ = f_- = f, \qquad M_+ = M_- = M - \vartriangle M\,.$$

Then 3) can be rewritten as $\mathcal{B}(F,f,M) - \mathcal{B}(F,f,M-\vartriangle M) \geq \vartriangle M f^2$ which means that $\dfrac{\partial \mathcal{B}}{\partial M} \geq f^2$ (the derivative again can be understood in any reasonable sense).

So the key property 3) implies the following two:

3′) $d^2\mathcal{B} \leq 0$;

3″) $\dfrac{\partial \mathcal{B}}{\partial M} \geq f^2$.

The nice thing is that the condition 3) is equivalent to 3′) and 3″): indeed,

$$\begin{aligned}\vartriangle M f^2 \leq \mathcal{B}(F,f,M) - \mathcal{B}(F,f,M-\vartriangle M) \leq \\ \leq \mathcal{B}(F,f,M) - \frac{1}{2}\big\{\mathcal{B}(F_+,f_+,M_+)-\mathcal{B}(F_-,f_-,M_-)\big\}\,.\end{aligned}$$

The first inequality here follows from $\dfrac{\partial \mathcal{B}}{\partial M} \geq f^2$; the second is simply the concavity of $\mathcal{B}$ (since the domain is convex, we always can derive from infinitesimal conditions 3′) and 3″) the corresponding conditions in finite differences).

1.1.5. Homework for the reader

If the reader wants to understand well what is going on, at this point he should try to find some function $\mathcal{B}$ satisfying conditions 1, 2, 3′, 3″ above by himself. One of the possible answers (but not the best one) can be found in the end of this introduction. The pleasure (or the challenge) to find the true Bellman function (we do not know it) is left completely to the reader.

Here is a hint for finding a Bellman type function. The true Bellman function has some homogeneity: $\mathcal{B}(\alpha^2 F, \alpha f, M) = \alpha^2 \mathcal{B}(F, f, M)$, and it is reasonable to look for a Bellman type function among such functions as well.

The reader is also encouraged to try to prove the L^p case or (and) the weighted version of the (dyadic) Carleson imbedding theorem.

1.1.6. A concluding remark

The above example and Figure 2 we hope give the reader an idea of how one can prove a "dyadic imbedding theorem" by finding an appropriate Bellman type function. By a "dyadic imbedding theorem" we very generally mean an upper estimate for a sum over all dyadic intervals.

Note that in the above example we did not lose anything: if the theorem is true, then the Bellman function exists; and if a Bellman type function exists, then we can run the machine backward and the theorem is true.

That should be the case for any "reasonable dyadic imbedding theorem" if we carefully choose parameters for the Bellman function and investigate the properties of this function.

Now we are going to consider an example of different nature where the idea of using Bellman function still works.

1.2. Example 2: dyadic maximal function

Let

$$(M\varphi)(t) = \sup_{I\,:\,t\in I} \langle |\varphi| \rangle_I$$

be the dyadic maximal function (recall that unless otherwise specified all intervals are supposed to be dyadic). We want to prove the well known result that

$$\int_{\mathbb{R}} (M\varphi)^2 \le C \int_{\mathbb{R}} \varphi^2 .$$

For simplicity we will again suppose that φ is a real valued (and even non-negative) function.

With any interval I and a function φ one can associate the following three quantities:

$$F_I := \langle \varphi^2 \rangle_I , \qquad f_I := \langle \varphi \rangle_I , \qquad L_I := \sup_{J\,:\,J\supset I} \langle \varphi \rangle_J .$$

Again we can define the Bellman function $\mathcal{B}(F, f, L)$ as the supremum

$$\mathcal{B}(F, f, L) := \sup\Big\{ \frac{1}{|I|} \int_I (M\varphi)^2 \,:\, \langle \varphi^2 \rangle_I = F, \langle \varphi \rangle_I = f, L_I = L \Big\} .$$

Note that again $\mathcal{B}$ does not depend on the choice of I.

1.2.1. Properties

1) Domain: $f^2 \le F$; $f \le L$;

2) Range: $L^2 \le \mathcal{B}(F, f, L) \le C(L^2 + F)$.

The inequalities defining the domain are trivial. The inequality $L^2 \le \mathcal{B}(F, f, L)$ comes from the observation that $M\varphi \ge L_I$ at every point of the interval I. The inequality $\mathcal{B}(F, f, L) \le C(L^2 + F)$ is simply our "belief that the theorem is true" formalized:

$$\frac{1}{|I|}\int_I [M(\varphi\chi_I)]^2 \le C\frac{1}{|I|}\int_I \varphi^2 = CF_I\,, \qquad \frac{1}{|I|}\int_I [M(\varphi\chi_{\mathbb{R}\setminus I})]^2 \le L_I^2\,.$$

Looking at how $\int\varphi^2$ and $\int\varphi$ can be shared between I_+ and I_-, we obtain a "pseudo-concavity" for $\mathcal{B}$, namely

3) $\mathcal{B}(F, f, L) \ge \frac{1}{2}\Big[\mathcal{B}(F_+, f_+, \max\{L, f_+\}) + \mathcal{B}(F_-, f_-, \max\{L, f_-\})\Big]$, provided that

$$(f_\pm)^2 \le F_\pm\,, \qquad F = \frac{1}{2}(F_+ + F_-), \qquad f = \frac{1}{2}(f_+ + f_-).$$

1.2.2. Running the machine backwards

Suppose we have *any* function $\mathcal{B}$ satisfying the conditions 1–3 above (again we call such function a Bellman type function). Then the theorem is proved.

Indeed, we can draw the binary tree (see Figure 3). It looks very similar to the Figure 2, but now there is nothing on the vertical edges, and the value at a vertex is at least the sum of values at the two vertices below.

Going n levels down on this tree, we get

$$\sum_{\substack{J\,:\,J\subset I\\ |J|=2^{-n}|I|}} L_J^2|J| \le \sum_{\substack{J\,:\,J\subset I\\ |J|=2^{-n}|I|}} |J|\cdot\mathcal{B}(F_J, f_J, L_J) \le |I|\cdot\mathcal{B}(F_I, f_I, L_I) \le C|I|(F_I + L_I^2)\,.$$

Consider a maximal function M_n, $\big(M_n\varphi\big)(x) = \sup\big\{\langle\varphi\rangle_J \;:\; J \ni x, |J| \ge 2^{-n}|I|\big\}$. Then the above inequality for a function φ supported inside the interval I can be rewritten as

$$\int_I \big(M_n\varphi\big)^2 \le C|I|(F_I + L_I^2) = C\Big[\int_I \varphi^2 + |I|\langle\varphi\rangle_I^2\Big] \le 2C\int_I \varphi^2\,.$$

Since $M_n\varphi \nearrow M\varphi$ as $n \to \infty$, we get for a function φ supported by I

$$\int_I \big(M\varphi\big)^2 \le 2C\int_I \varphi^2\,.$$

Considering intervals $[-2^n, 0]$ and $[0, 2^n]$ with n large enough, we get the inequality

$$\int_{\mathbb{R}} \big(M\varphi\big)^2 \le 2C\int_{\mathbb{R}} \varphi^2$$

for any compactly supported φ, and therefore for any $\varphi \in L^2$.

Again, the reader is encouraged to seek a Bellman type function associated with the "maximal function problem" himself before looking at the answer in the end of the section.

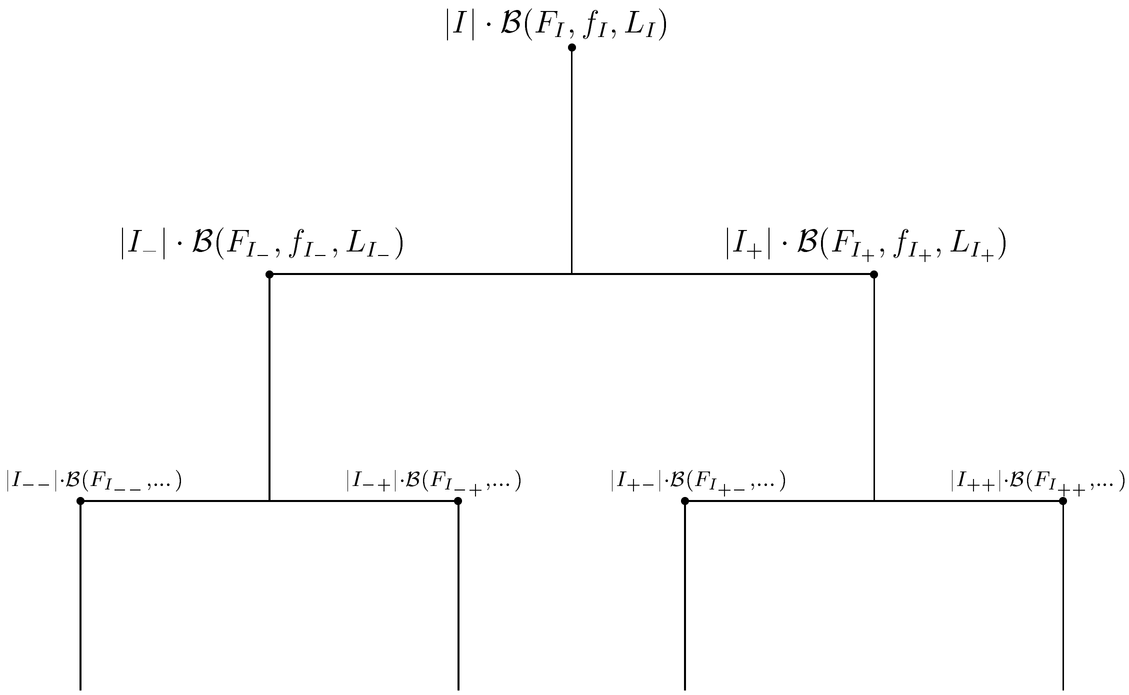

Figure 3: Maximal function theorem from Bellman function. Vertices correspond to dyadic intervals and are denoted by black dots.

1.3. Answers

A possible Bellman type function for the Carleson imbedding theorem:

$$\mathcal{B}(F,f,M) = 4\Big(F - \frac{f^2}{1+M}\Big).$$

A possible function for the maximal theorem:

$$\mathcal{B}(F,f,L) = 4\big(F - fL + \frac{L^2}{2}\big)\,.$$

To check that the functions possess the necessary properties is left to the reader as an easy exercise.

2. Weighted norm inequalities with matrix weights: necessary conditions

Now, we hope, the reader has got some impression of *how* we are going to do things and it is time to tell him *what* we will do. Our main aim will be to investigate the action of the Hilbert transform T defined on "good" functions φ by

$$(T\varphi)(t) = \lim_{\varepsilon\to 0+} \frac{1}{\pi} \int_{|s-t|>\varepsilon} \frac{\varphi(s)}{t-s}\, ds$$

in weighted L^p-spaces with scalar or matrix weights.

2.1. Weighted L^p spaces with matrix weights

The formal definition of a weighted space is the following:

Let E be a finite-dimensional Hilbert space over $\mathbb{C}$. Let $\mathcal{L}_+(E)$ be the cone of (strictly) positively defined self-adjoint operators in E. The space $L^p(W)$ (where $1 < p < +\infty$; $W : \mathbb{R} \to \mathcal{L}_+(E)$) consists of all ($E$-valued) functions φ for which the norm

$$\|\varphi\|_{L^p(W)} := \left\{ \int_{\mathbb{R}} \|W(t)^{1/p}\varphi(t)\|_E^p\, dt \right\}^{1/p}$$

is finite.

There are two cases when the expression for the norm $\|\varphi\|_{L^p(W)}$ gets simpler:

1. $\dim E = 1$ (or just $E = \mathbb{C}$).

   Then W is just a scalar-valued positive function and

$$\|\varphi\|_{L^p(W)} = \left\{ \int_{\mathbb{R}} |\varphi(t)|^p W(t)\, dt \right\}^{1/p}.$$

2. $p = 2$

   Then

$$\|\varphi\|_{L^2(W)} = \left\{ \int_{\mathbb{R}} (W(t)\varphi(t), \varphi(t))_E dt \right\}^{1/2}.$$

As the mapping $L^p(W) \ni \varphi(\cdot) \mapsto W(\cdot)^{1/p}\varphi(\cdot) \in L^p(Id)$ is an isometry, we can easily compute the dual space. The result is

$$L^p(W)^* = L^q(W^{-q/p})$$

with respect to the pairing $(\varphi, \psi) = \int_{\mathbb{R}} (\varphi(t), \psi(t))_E dt$ (as usual q denotes here the conjugate exponent defined by $1/p + 1/q = 1$).

2.2. Hilbert transform in $L^p(W)$

To define the Hilbert transform in $L^p(W)$ we will note first that T can be continued from the set of smooth functions with compact support onto the (non-weighted) space L^2 (scalar or vector-valued, doesn't matter) as an unitary operator. The simplest way to prove this is to use the Fourier transform $\widehat{\varphi}(s) = \int_{\mathbb{R}} \varphi(t) e^{-2\pi ist} dt$ which turns the Hilbert transform T into the multiplication by $-i \operatorname{sign} s$. Since $\|\varphi\|_{L^2} = \|\widehat{\varphi}\|_{L^2}$, we get the desired result immediately.

Note now that $L^2 \cap L^p(W)$ is dense in $L^p(W)$. We will say that the Hilbert transform acts in $L^p(W)$ if for every $\varphi \in L^2 \cap L^p(W)$ the image $T\varphi \in L^p(W)$ and there exists a constant C (not depending on φ) such that $\|T\varphi\|_{L^p(W)} \le C\|\varphi\|_{L^p(W)}$.

2.3. Necessary condition

Let H^2 be the space of E-valued functions $\varphi(z)$ analytic in the upper half-plane $\mathbb{C}_+ = \{z : \operatorname{Im} z > 0\}$ and satisfying

$$\|\varphi\|_{H^2} = \sup_{\sigma > 0} \left\{ \int_{\mathbb{R}} \|\varphi(t + i\sigma)\|_E^2 \right\}^{1/2} < +\infty.$$

Recall that every function $\varphi \in H^2$ has non-tangential boundary values $\varphi(t)$ for a.e. $t \in \mathbb{R}$. Identification of $\varphi(z) \in H^2$ with its boundary value function $\varphi(t) \in L^2$ (possible because $\varphi(z)$ can be recovered from $\varphi(t)$ by the formula $\varphi(z) = (2\pi i)^{-1} \int_{\mathbb{R}} \varphi(t)(z-t)^{-1} dt$) makes H^2 a closed subspace of L^2. It is well-known that the orthogonal projection $P_+ : L^2 \to H^2$ is given by $P_+ = (Id + iT)/2$.

Let now H^2_λ ($\lambda \in \mathbb{C}_+$) be the subspace of the functions $\varphi \in H^2$ for which $\varphi(\lambda) = 0$. The orthogonal projection onto H^2_λ is given by $b_\lambda(\cdot) P_+ \overline{b_\lambda(\cdot)}$ where $b_\lambda(z) = (z - \lambda)(z - \overline{\lambda})^{-1}$ is the corresponding Blaschke factor. The orthogonal complement $H^2 \ominus H^2_\lambda$ is the space $k_\lambda(\cdot)E = \{k_\lambda(\cdot)x : x \in E\}$ where

$$k_\lambda(z) = \sqrt{\frac{\operatorname{Im} \lambda}{\pi}} \cdot \frac{1}{z - \overline{\lambda}}$$

is the normalized (by the condition $\|k_\lambda\|_{H^2} = 1$) reproducing kernel. Thus, the orthogonal projection $\mathcal{P}_\lambda : L^2 \to H^2 \ominus H^2_\lambda$ is given by

$$\mathcal{P}_\lambda \varphi(t) = k_\lambda(t) \int_{\mathbb{R}} \overline{k_\lambda(s)} \varphi(s) ds.$$

On the other hand this projection can be also expressed as $P_+ - b_\lambda(\cdot) P_+ \overline{b_\lambda(\cdot)}$. Since both the identity operator and the multiplication by the unimodular function b_λ (or its conjugate) act isometrically in $L^p(W)$, we conclude that for every $\lambda \in \mathbb{C}_+$ the operator $\mathcal{P}_\lambda$ acts in $L^p(W)$ and what is more, the operator norm

$$\|\mathcal{P}_\lambda\|_{L^p(W)\to L^p(W)} \le 1 + \|T\|_{L^p(W)\to L^p(W)}$$

for every $\lambda \in \mathbb{C}_+$.

Let us now calculate the operator norm of $\mathcal{P}_\lambda$. For this it will be convenient to use a bit different (but, of course, equivalent) definition of the space $L^p(W)$. Namely, instead of saying that we put a (matrix) weight $W(t)$ at every point $t \in \mathbb{R}$, we will say that at every point $t \in \mathbb{R}$ we change the standard norm $\|\cdot\|_E$ of the space E to another norm $\rho_t(x) = \|W(t)^{1/p} x\|_E$ and the space $L^p(W)$ (which from this point of view should be rather denoted by $L^p(\rho_t)$) is defined as the space of all functions φ for which the function $t \mapsto \rho_t(\varphi(t))$ belongs to (scalar) L^p with the norm

$$\|\varphi\|_{L^p(W)} = \|t \mapsto \rho_t(\varphi(t))\|_{L^p} = \left\{ \int_{\mathbb{R}} \rho_t(\varphi(t))^p dt \right\}^{1/p}.$$

Recall that for every norm ρ in E its dual norm ρ^* is defined as $\rho^*(y) = \sup\{|(x,y)_E| : \rho(x) \le 1\}$ and that always $(\rho^*)^* = \rho$.

A simple computation shows that for the norm ρ_t defined above the dual norm $\rho_t^*(y) = \|W(t)^{-1/p} y\|_E$ and therefore the dual space to $L^p(W) = L^p(\rho_t)$ is just $L^q(\rho_t^*)$ (which was denoted $L^q(W^{-q/p})$ when we first introduced it).

Lemma 2.1. *Let $k : \mathbb{R} \to \mathbb{C}$, $k \in L^2$, $\|k\|_{L^2} = 1$. Consider the orthogonal projection $\mathcal{P}_k$ onto the space $k(\cdot)E \subset L^2$ which for (E-valued) functions $\varphi \in L^2$ is given by*

$$(\mathcal{P}\varphi)(t) = k(t) \int_{\mathbb{R}} \overline{k(s)} \varphi(s) ds. \tag{2.1}$$

The operator $\mathcal{P}_k$ acts in $L^p(W)$ if and only if the expressions

$$\langle \rho \rangle (x) = \langle \rho \rangle_{k,p}(x) = \left(\int_{\mathbb{R}} |k(t)|^p \rho_t(x)^p dt \right)^{1/p}$$

and

$$\langle \rho^* \rangle (y) = \langle \rho^* \rangle_{k,q}(y) = \left(\int_{\mathbb{R}} |k(t)|^q \rho_t^*(y)^q dt \right)^{1/q}$$

define finite norms on E (i.e. the integrals converge for every $x, y \in E$). If it is the case, then $\mathcal{P}_k\varphi$ is well-defined (i.e. the corresponding integral in 2.1 converges) for any $\varphi \in L^p(W)$ and the norm of $\mathcal{P}_k$ is

$$\|\mathcal{P}_k\| = \|\mathcal{P}_k\|_{L^p(W)\to L^p(W)} = \sup_{y\in E\setminus\{0\}} \frac{\langle\rho^*\rangle(y)}{\langle\rho\rangle^*(y)}.$$

Proof. By the definition

$$\begin{aligned}
&\|\mathcal{P}_k\|_{L^p(W)\to L^p(W)} = \\
&\sup\Big\{\Big(\int_{\mathbb{R}} \rho_t\Big(k(t)\int_{\mathbb{R}} \overline{k(s)}\varphi(s)ds\Big)^p dt\Big)^{1/p} : \varphi \in L^2\cap L^p(W), \int_{\mathbb{R}} \rho_t(\varphi(t))^p dt \le 1\Big\} = \\
&\sup\Big\{\Big(\int_{\mathbb{R}} |k(t)|^p\rho_t\Big(\int_{\mathbb{R}} \overline{k(s)}\varphi(s)ds\Big)^p dt\Big)^{1/p} : \varphi \in L^2\cap L^p(W), \int_{\mathbb{R}} \rho_t(\varphi(t))^p dt \le 1\Big\} = \\
&\sup\Big\{\langle\rho\rangle\Big(\int_{\mathbb{R}} \overline{k(s)}\varphi(s)ds\Big) : \varphi \in L^2\cap L^p(W), \int_{\mathbb{R}} \rho_t(\varphi(t))^p dt \le 1\Big\}. \quad (2.2)
\end{aligned}$$

Let us show first that $\langle\rho\rangle$ defines a bounded norm on E, i.e. that $\langle\rho\rangle(x)$ is finite for any $x \in E$.

Let $\varepsilon > 0$. Let

$$\begin{aligned}
S_\varepsilon &= \{t \in \mathbb{R} : \varepsilon^p Id \le W(t) \le \varepsilon^{-p} Id\} \cap (-\frac{1}{2\varepsilon}, \frac{1}{2\varepsilon}) \\
&= \{t \in \mathbb{R} : \varepsilon\|x\|_E \le \rho_t(x) \le \varepsilon^{-1}\|x\|_E \text{ for every } x \in E\} \cap (-\frac{1}{2\varepsilon}, \frac{1}{2\varepsilon}).
\end{aligned}$$

Note that $\bigcup_{\varepsilon>0} S_\varepsilon = \mathbb{R}$ and therefore one can find $\varepsilon > 0$ for which $\int_{S_\varepsilon} |k(t)|^2 dt > 0$. Then there exists a scalar function $\psi \in L^\infty$ such that $\operatorname{supp}\psi \subset S_\varepsilon$, $\|\psi\|_{L^\infty} \le \varepsilon^{1+1/p}$ and $\int_{\mathbb{R}} \overline{k(s)}\psi(s)ds = \delta > 0$. For every unit vector $x \in E$ the function $\varphi(t) = \psi(t)x$ belongs to $L^2 \cap L^p(W)$ and

$$\int_{\mathbb{R}} \rho_t(\varphi(t))^p dt \le \varepsilon^{p+1}\int_{S_\varepsilon} \rho_t(x)^p dt \le \varepsilon^{p+1}\int_{(-\frac{1}{2\varepsilon},\frac{1}{2\varepsilon})} \varepsilon^{-p} dt = 1$$

while $\int_{\mathbb{R}} \overline{k(s)}\varphi(s)ds = \delta x$. Thus,

$$\|\mathcal{P}_k\| \ge \delta\sup\{\langle\rho\rangle(x) : \|x\|_E = 1\}.$$

This implies that if $\mathcal{P}_k$ acts in $L^p(W)$, then $\langle\rho\rangle(x)$ is finite for every $x \in E$. Also, if $x \neq 0$ then $\rho_t(x) > 0$ for every $t \in \mathbb{R}$ and, since $|k(t)| > 0$ on a set of positive measure, $\langle\rho\rangle(x) > 0$. Thereby $\langle\rho\rangle$ is a well-defined finite norm on E.

Now we want to replace the set $L^2 \cap L^p(W)$ of test functions in the definition of $\|\mathcal{P}_k\|_{L^p(W)\to L^p(W)}$ by something that is more convenient to work with. Let $D_\varepsilon = \{\varphi \in L^2 \cap L^p(W) : \operatorname{supp}\varphi \subset S_\varepsilon\}$; $D = \bigcup_{\varepsilon>0} D_\varepsilon$. Note that for every $\varphi \in L^2 \cap L^p(W)$ the sequence of functions $\varphi_\varepsilon = \varphi\chi_{S_\varepsilon} \in D$ converges to φ in both L^2 and $L^p(W)$ as $\varepsilon \to 0$. Since $k \in L^2$ and since $\langle\rho\rangle$ is a finite norm on E, this implies

$$\langle\rho\rangle\left(\int_{\mathbb{R}} \overline{k(s)}\varphi_\varepsilon(s)ds\right) \to \langle\rho\rangle\left(\int_{\mathbb{R}} \overline{k(s)}\varphi(s)ds\right) \quad \text{as } \varepsilon \to 0.$$

Thereby to compute the norm $\|\mathcal{P}_k\|_{L^2(W)\to L^2(W)}$ it is enough to take the *supremum* in (2.2) over all $\varphi \in D$.

The main advantage of such a replacement is that the inclusion $\varphi \in D$ (which by the definition means that $\operatorname{supp}\varphi \subset S_\varepsilon$ for some $\varepsilon > 0$, $\{t \mapsto \rho_t(\varphi(t))\} \in L^p$, $\|\varphi(\cdot)\|_E \in L^2$) is equivalent to

$$\operatorname{supp}\varphi \subset S_\varepsilon \text{ for some } \varepsilon > 0, \quad \{t \mapsto \rho_t(\varphi(t))\} \in L^2 \cap L^p$$

i.e. depends on $\rho_t(\varphi(t))$ only. Thus, if we change the direction of the vector $\varphi(t) \in E$ at some points $t \in \mathbb{R}$ without changing $\rho_t(\varphi(t))$, we will stay in D.

Now everything is ready to compute the norm $\|\mathcal{P}_k\|_{L^p(W)\to L^p(W)}$. For every $x \in E$ we have by duality

$$\langle\rho\rangle(x) = \sup_{y\in E\setminus\{0\}} \frac{|(x,y)_E|}{\langle\rho\rangle^*(y)}.$$

Thereby

$$\|\mathcal{P}_k\| = \sup\left\{\frac{1}{\langle\rho\rangle^*(y)}\left|\int_{\mathbb{R}} \big(\varphi(t), k(t)y\big)_E dt\right| : \ \varphi \in D, \ \int_{\mathbb{R}} \rho_t(\varphi(t))^p \le 1, \ y \in E\setminus\{0\}\right\}.$$

The absolute value $\left|\int_{\mathbb{R}}(\dots)_E dt\right|$ can be replaced by $\int_{\mathbb{R}} |(\dots)_E| dt$ because we can always multiply φ by any unimodular scalar function. Note now that when $t \in \mathbb{R}$ and $\rho_t(\varphi(t))$ are fixed, the maximal possible value of $|(\varphi(t), k(t)y)_E|$ is

$$\rho_t(\varphi(t)) \cdot \rho_t^*(k(t)y) = \rho_t(\varphi(t)) \cdot |k(t)|\rho_t^*(y).$$

Keeping $\rho_t(\varphi(t))$ fixed and choosing the optimal direction of the vector $\varphi(t) \in E$ at every point $t \in \mathbb{R}$, we get

$$\|\mathcal{P}_k\| = \sup\left\{\frac{1}{\langle\rho\rangle^*(y)}\int_{\mathbb{R}} \rho(t)\cdot|k(t)|\rho_t^*(y)dt\right\}$$

where the supremum is taken over all scalar functions ρ supported by some S_ε, such that $\int_{\mathbb{R}} \rho(t)^p \le 1$, $\int_{\mathbb{R}} \rho(t)^2 < \infty$, and over all $y \in E\setminus\{0\}$. For a fixed vector $y \in E$ the supremum of the integral $\int_{\mathbb{R}} \rho(t)\cdot|k(t)|\rho_t^*(y)dt$ over all functions $\rho(t)$ is simply L^q-norm of the function $|k(t)|\rho_t^*(y)$, where $1/p + 1/q = 1$. Therefore

$$\|\mathcal{P}_k\| = \sup_{y\in E\setminus\{0\}}\left\{\frac{1}{\langle\rho\rangle^*(y)}\left(\int_{\mathbb{R}} |k(t)|^q\rho_t^*(y)^q dt\right)^{1/q}\right\} = \sup_{y\in E\setminus\{0\}} \frac{\langle\rho^*\rangle(y)}{\langle\rho\rangle^*(y)},$$

which is exactly the formula for the norm to prove.

Note that it follows from the reasoning above that for any $\varphi \in D$ and any $y \in E$

$$\int_{\mathbb{R}} \left|(\overline{k(t)}\varphi(t), y)_E\right| dt \le \langle\rho^*\rangle(y)\cdot\|\varphi\|_{L^p(W)}.$$

Applying standard approximation argument one can get this estimate for any $\varphi \in L^p(W)$, which means that for any such φ the integral $\int \overline{k(t)}\varphi(t)dt$ converges in a weak sense (i.e. the integral can be defined as a weak integral). For functions with values in a (separable) Hilbert space weak and strong integral coincide. In our case $\dim E < \infty$ we can even claim the absolute convergence $\int \|\overline{k(t)}\varphi(t)\|_E dt < \infty$. □

Taking $k = k_\lambda$ in this lemma, we see that if the Hilbert transform T acts in $L^p(W)$, then there exists a constant A (namely, $A = 1 + \|T\|_{L^p(W)\to L^p(W)}$) such that for every $\lambda \in \mathbb{C}_+$

$$\langle \rho^* \rangle_{k_\lambda,q} \le A \langle \rho \rangle^*_{k_\lambda,p}.$$

Now let $I \subset \mathbb{R}$ be an interval. Choosing $\lambda \in \mathbb{C}_+$ such that $\operatorname{Re}\lambda$ is the center of the interval I and $\operatorname{Im}\lambda = |I|/2$, we observe that

$$|k_\lambda(t)| = \sqrt{\frac{\operatorname{Im}\lambda}{\pi}}\frac{1}{|t-\overline{\lambda}|} \ge \frac{1}{\sqrt{2\pi \operatorname{Im}\lambda}} = \frac{1}{\sqrt{\pi|I|}}$$

for every $t \in I$. Therefore

$$\langle \rho \rangle_{k_\lambda,p}(x) = \left(\int_{\mathbb{R}} |k_\lambda(t)|^p \rho_t(x)^p dt\right)^{1/p} \ge \pi^{-1/2}|I|^{\frac{1}{p}-\frac{1}{2}}\langle \rho_t(x)^p\rangle_I^{1/p} = \pi^{-1/2}|I|^{\frac{1}{p}-\frac{1}{2}}\langle \rho \rangle_{I,p}(x)$$

and analogously

$$\langle \rho^* \rangle_{k_\lambda,q}(y) \ge \pi^{-1/2}|I|^{\frac{1}{q}-\frac{1}{2}}\langle \rho^* \rangle_{I,q}(y).$$

Since $1/p + 1/q = 1$, we conclude from here that for every interval $I \subset \mathbb{R}$

$$\langle \rho^* \rangle_{I,q} \le A' \langle \rho \rangle^*_{I,p} \tag{$\mathcal{A}_{p,q}$}$$

with $A' = \pi A$.

2.4. Particular cases of the $(\mathcal{A}_{p,q})$ condition

The condition $(\mathcal{A}_{p,q})$ can be given a more explicit form in two simplest cases $\dim E = 1$ and $p = 2$.

When $\dim E = 1$, we get

$$\langle \rho \rangle_{I,p}(x) = \langle W \rangle_I^{1/p}|x|$$

and therefore

$$\langle \rho \rangle^*_{I,p}(y) = \langle W \rangle_I^{-1/p}|y|.$$

On the other hand,

$$\langle \rho^* \rangle_{I,q}(y) = \langle W^{-q/p} \rangle_I^{1/q}|y|.$$

Thus the $(\mathcal{A}_{p,q})$ condition reduces to

$$\langle W \rangle_I^{1/p}\langle W^{-q/p} \rangle_I^{1/q} \le A'$$

which is the famous Muckenhoupt (A_p) condition.

When $p=2$, we get

$$\langle\rho\rangle_{I,2}(x)=\left\{\frac{1}{|I|}\int_I (W(t)x,x)_E\,dt\right\}^{1/2}=(\langle W\rangle_I x,x)_E^{1/2}=\|\langle W\rangle_I^{1/2}x\|_E$$

and thereby

$$\langle\rho\rangle_{I,2}^*(y)=\|\langle W\rangle_I^{-1/2}y\|_E.$$

On the other hand,

$$\langle\rho^*\rangle_{I,2}(y)=\|\langle W^{-1}\rangle_I^{1/2}y\|_E$$

and therefore $(\mathcal{A}_{2,2})$ reduces to the condition

$$\|\langle W^{-1}\rangle_I^{1/2}\langle W\rangle_I^{1/2}\|_{E\to E}\le A'$$

(which was called "matrix (A_2) condition" in [2]).

Below we will show that $(\mathcal{A}_{p,q})$ is also sufficient, but before we start doing that let us make several more remarks about this condition.

2.5. Properties of $(\mathcal{A}_{p,q})$ weights

1. Since for every $x,y\in E$ we have

$$|(x,y)_E|=\frac{1}{|I|}\int_I |(x,y)_E|dt\le$$
$$\le\frac{1}{|I|}\int_I \rho_t(x)\rho_t^*(y)dt\le\langle\rho_t(x)^p\rangle_I^{1/p}\langle\rho_t^*(y)^q\rangle_I^{1/q}=\langle\rho\rangle_{I,p}(x)\langle\rho^*\rangle_{I,q}(y),$$

   we always have $\langle\rho^*\rangle_{I,q}\ge\langle\rho\rangle_{I,p}^*$.

2. The condition $(\mathcal{A}_{p,q})$ seems asymmetric for the first look but if we just switch to the dual norms on both sides, we will immediately get the "missing" counterpart $\langle\rho\rangle_{I,p}\le A'\langle\rho^*\rangle_{I,q}^*$.

3. In the case $p=2$ one may prefer not to switch from $\langle\rho\rangle_{k_\lambda,2}$ to $\langle\rho\rangle_{I,2}$. The reason is that the square of the normalized reproducing kernel k_λ is exactly the Poisson kernel $P(\lambda,t)$ and therefore we get the so-called "invariant (A_2) condition"

$$\|(W^{-1})(\lambda)^{1/2}W(\lambda)^{1/2}\|_{E\to E}\le A'$$

   where $W(\lambda)=\int_{\mathbb{R}}P(\lambda,t)W(t)dt$ and $(W^{-1})(\lambda)=\int_{\mathbb{R}}P(\lambda,t)W(t)^{-1}dt$ are the harmonic extensions of the weights W and W^{-1} onto the upper half-plane.

   The importance of having a condition in terms of harmonic extensions arises from the fact that the Hilbert transform $T\varphi$ of a function φ is just the harmonic conjugate of φ, which can be used to give an almost trivial proof of sufficiency of $(\mathcal{A}_{2,2})$ (see for instance [3] where it was done for the scalar case).

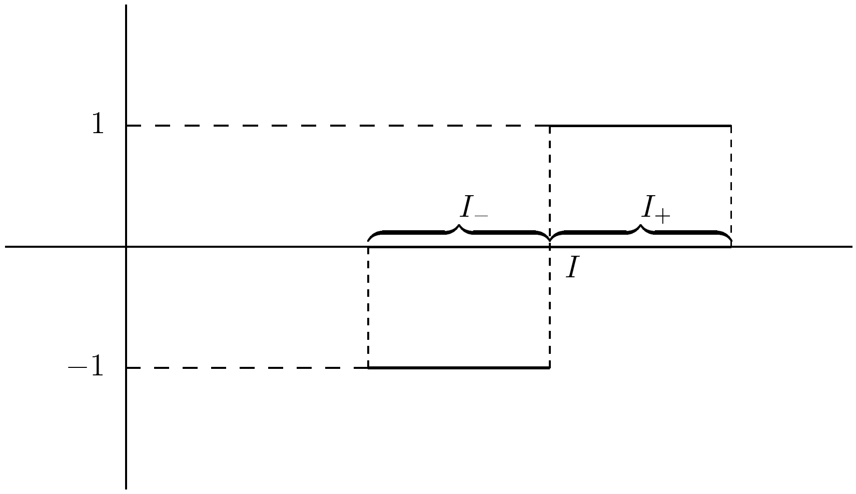

Figure 4: The Haar function h_I

3. A naïve idea

Most part of the rest of the paper will be devoted to the proof of sufficiency of the matrix Muckenhoupt condition $(\mathcal{A}_{p,q})$ for the boundedness of the Hilbert transform in the weighted space $L^p(W)$ (actually we will prove this result for any classical Calderon-Zygmund operator).

The approach we are going to use looks very naïve. How can one hope to prove that an operator is bounded if the "brute force" idea of replacing the kernel by its absolute value does not work? He could try to find a basis in which the matrix of the operator is close to diagonal, so the boundedness of the operator can be verified by using some standard test (say, Schur's test).

In the case of the Hilbert transform the system of exponents $\{z^n\}_{n=-\infty}^{\infty}$ on the circle or its continuous analog $\{\exp(2\pi i s t)\}_{s\in\mathbb{R}}$ on the real line seems to be the best choice: the Hilbert transform gets purely diagonal. Unfortunately this choice is *too* good: to check that this system of exponents is a basis is *exactly the same* as to check that the Hilbert transform is bounded. So, this would not lead us to anything except trivial reformulation of the problem. Such reformulation can either make everything evident (as for non-weighted L^2-space) or not help at all (as it seems to happen in all other cases).

The next famous to the system of exponents orthogonal basis in (scalar) L^2 is the Haar system. Let $\mathcal{D}$ denote the set of all dyadic subintervals of $\mathbb{R}$. Recall that the Haar function h_I associated with a dyadic interval $I \in \mathcal{D}$ is defined by

$$h_I(t) = \begin{cases} 0, & t \notin I\,; \\ -1, & t \in I_-\,; \\ 1, & t \in I_+\,; \end{cases}$$

where as usual I_- and I_+ are the left and the right halves of I respectively (see Figure 4).

Any function φ in (scalar) L^2 can be represented in a unique way as a sum $\sum_I c_I(\varphi) h_I$, where

$$c_I(\varphi) = \frac{1}{\|h_I\|_{L^2}^2}(\varphi, h_I)_{L^2} = \frac{1}{|I|}\int_I \varphi h_I\,.$$

The series converges unconditionally (independently of order) in L^2 and

$$(\varphi,\psi)_{L^2} = \sum_I (c_I(\varphi)h_I, c_I(\psi)h_I)_{L^2} = \sum_I c_I(\varphi)\overline{c_I(\psi)}\cdot|I|\,,$$

so the norm is given by

$$\|\varphi\|_{L^2} = \Big\|\big\{\|c_I(\varphi)h_I\|_{L^2}\big\}_{I\in\mathcal{D}}\Big\|_{\ell^2} = \Big(\sum_I |c_I(\varphi)|^2\|h_I\|_{L^2}^2\Big)^{1/2}.$$

In the E-valued space L^2 the Haar functions generate mutually orthogonal subspaces $\mathcal{E}_I := h_I(\cdot)E = \{h_I(\cdot)\mathbf{e} : \mathbf{e}\in E\}$. The subspaces $\mathcal{E}_I$ form an orthogonal basis in L^2. This means that every $\varphi \in L^2$ can be represented as the sum $\varphi = \sum_I \mathcal{P}_I\varphi$, where $\mathcal{P}_I$ is an orthogonal projection onto $\mathcal{E}_I$.

Again for every φ, $\psi \in L^2$

$$\|\varphi\|_{L^2} = \Big\|\big\{\|\mathcal{P}_I\varphi\|_{L^2}\big\}_{I\in\mathcal{D}}\Big\|_{\ell^2} = \Big(\sum_I \|\mathcal{P}_I\varphi\|_{L^2}^2\Big)^{1/2}$$

and

$$(\varphi,\psi)_{L^2} = \sum_I (\mathcal{P}_I\varphi, \mathcal{P}_I\psi)_{L^2}$$

The projections $\mathcal{P}_I$ can be easily computed:

$$\big(\mathcal{P}_I\varphi\big)(\cdot) = h_I(\cdot)\frac{1}{|I|}\int_I h_I(t)\varphi(t)dt = c_I(\varphi)h_I(\cdot)$$

where

$$c_I(\varphi) = \frac{1}{|I|}\int_I h_I(t)\varphi(t)dt \,\in\, E\,.$$

So, we may also think of the decomposition $\varphi = \sum_I \mathcal{P}_I\varphi$ as of a "linear combination" of Haar functions with vector coefficients. The above formulae for the norm and for the inner product can be then rewritten as

$$\|\varphi\|_{L^2} = \Big(\sum_I \|c_I(\varphi)\|_E^2|I|\Big)^{1/2}$$

and

$$(\varphi,\psi)_{L^2} = \sum_I (c_I(\varphi), c_I(\psi))_E\cdot|I|\,.$$

A nice property of the Haar functions h_I is that they all have mean value 0, so the Haar system should work well with "smooth" singular kernels, like that of the Hilbert transform.

Indeed, the matrix of the Hilbert transform with respect to the Haar system is pretty reasonable, so, as it will be clearly seen in the next sections, the hardest part is to show that the Haar system is a basis.

A reader familiar with wavelets may ask, “Why don’t you use them instead?” It turns out that the wavelets do not have any advantage over the Haar system in this problem. Although the (already good) matrix of the Hilbert transform gets a little bit better in wavelets, when proving that the system is a basis, the Haar system has a clear advantage — its martingale structure.

Our first task will be to show that the Haar system $\{h_I\}$ (or in the vector case the system of the corresponding subspaces $\mathcal{E}_I = h_I(\cdot)E$) is a *strong unconditional basis* in $L^p(W)$. The meaning of this term will be explained in the next section.

4. Bases of subspaces in Banach spaces

In this section we will remind the reader (or acquaint him with) some standard facts from the theory of bases. This entire section was rewritten word by word from a textbook on functional analysis; unfortunately the authors do not agree on from which one, and therefore no reference will be given.

We start with recalling several definition.

4.1. Main definitions

Let $\mathcal{J}$ be some countable index set.

4.1.1. Ideal spaces of sequences

A Banach space Y of sequences $\{c_j\}_{j\in\mathcal{J}}$ of complex numbers is called *ideal* if for any sequence of factors α_j, $|\alpha_j| \le 1$ the sequence $\{\alpha_j c_j\}_{j\in\mathcal{J}} \in Y$ and $\|\{\alpha_j c_j\}_{j\in\mathcal{J}}\|_Y \le \|\{c_j\}_{j\in\mathcal{J}}\|_Y$

4.1.2. Unconditional convergence

Let X be a Banach space, $\{x_j\}_{j\in\mathcal{J}}$ be a sequence of elements of X. We say that the series $\sum_{j\in\mathcal{J}} x_j$ *converges unconditionally* to $x \in X$ if it converges to x in the norm of X for any ordering of the index set $\mathcal{J}$. An equivalent definition is that for any $\varepsilon > 0$ there exists a finite set of indices $\mathcal{J}_0 \subset \mathcal{J}$ such that for any finite set $\mathcal{K} \supset \mathcal{J}_0$

$$\Big\|\sum_{j\in\mathcal{K}} x_j - x\Big\|_X < \varepsilon\,.$$

If $\dim X < \infty$, the unconditional convergence of the series $\sum_{j\in\mathcal{J}} x_j$ is equivalent to *absolute convergence* (i.e. to $\sum_{j\in\mathcal{J}} \|x_j\|_X < \infty$). In general absolute convergence implies unconditional, but not vice versa.

4.1.3. Strong unconditional bases

Let $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ be a family of non-trivial ($\mathcal{E}_j \neq \{0\}$) closed subspaces of X. We say that $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a *strong unconditional basis* in X if

1. The linear span $\mathcal{L}\{\mathcal{E}_j : j \in \mathcal{J}\}$ is dense in X;
2. There exists an ideal Banach space Y of sequences $\{c_j\}_{j\in\mathcal{J}}$ and a constant $A > 0$ such that for any sequence $\{x_j\}_{j\in\mathcal{J}}$, $x_j \in \mathcal{E}_j$ with finitely many non-zero elements
$$\frac{1}{A}\Big\|\sum_{j\in\mathcal{J}} x_j\Big\|_X \leq \Big\|\{\|x_j\|\}_{j\in\mathcal{J}}\Big\|_Y \leq A\Big\|\sum_{j\in\mathcal{J}} x_j\Big\|_X .$$

Note that it follows from the norm equivalence property 2 that the subspaces $\mathcal{E}_j$ are linearly independent (which means that among all finite linear combinations $\sum_j x_j$, $x_j \in \mathcal{E}_j$ only trivial one (all $x_j = 0$) can be zero).

4.1.4. Coordinate projections $\mathcal{P}_j$

Let $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ be a strong unconditional basis in (a Banach space) X. Since any $x \in \mathcal{E} := \mathcal{L}\{\mathcal{E}_j : j \in \mathcal{J}\}$ can be represented in a unique way as a finite sum $x = \sum_{j\in\mathcal{J}} x_j$, $x_j \in \mathcal{E}_j$ (recall that $\mathcal{E}_j$ are linearly independent), we can define on $\mathcal{E}$ coordinate projections $\mathcal{P}_j$, $\mathcal{P}_j x = x_j$.

Clearly $\mathcal{P}_j\mathcal{E} = \mathcal{E}_j$, $\mathcal{P}_j^2 = \mathcal{P}_j$ and at last $\mathcal{P}_{j'}\mathcal{P}_{j''} = 0$ if $j' \neq j''$.

From the norm equivalence property 2 it follows that all $\mathcal{P}_j$ are uniformly bounded, namely $\|\mathcal{P}_j x\|_X \leq A^2\|x\|_X$ for every $x \in \mathcal{E}$. Since $\mathcal{E}$ is dense in X, all the projections $\mathcal{P}_j$ can be in a unique way continuously extended to the whole space X (with all their properties preserved).

Proposition 4.1. *If $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a strong unconditional basis in X, then any $x \in X$ can be in a unique way represented as a sum of unconditionally convergent series $x = \sum_{j\in\mathcal{J}} x_j$, $x_j \in \mathcal{E}_j$, and, moreover, $x_j = \mathcal{P}_j x$.*

Proof. Assume that $x = \sum_{j\in\mathcal{J}} x_j$, $x_j \in \mathcal{E}_j$ and the series converges unconditionally. Then for every $i \in \mathcal{J}$ we have
$$\mathcal{P}_i x = \mathcal{P}_i \sum_{j\in\mathcal{J}} x_j = \sum_{j\in\mathcal{J}} \mathcal{P}_i x_j = x_i,$$
proving the uniqueness.

It remains to prove that the series $\sum_{j\in\mathcal{J}} \mathcal{P}_j x$ converges unconditionally to x for every $x \in X$. Let $\mathcal{J}_k$, $k = 1, 2, \dots$ be an increasing ($\mathcal{J}_k \subset \mathcal{J}_{k+1}$) sequence of finite subsets of $\mathcal{J}$ such that $\cup_k \mathcal{J}_k = \mathcal{J}$. Define projections $\mathcal{P}^{(k)} = \sum_{j\in\mathcal{J}_k} \mathcal{P}_j$. The norm equivalence property implies that the norms of all projections $\mathcal{P}^{(k)}$ are uniformly bounded (by A^2) on the dense set $\mathcal{E} = \mathcal{L}\{\mathcal{E}_j : j \in \mathcal{J}\}$ (and therefore on the whole X). Clearly, for any $x \in \mathcal{E}$ we have $\|\mathcal{P}^{(k)}x - x\|_X \to 0$ as $k \to \infty$. Since $\mathcal{P}^{(k)}$ are uniformly bounded and $\mathcal{E}$ is dense, the convergence takes place for any $x \in X$. Since the sequence $\mathcal{J}_k$ was arbitrary, we get the unconditional convergence desired. $\square$

Using standard approximation argument one can show that the norm equivalence property 2 holds for all sequences, not only for finite ones.

Corollary 4.2. *If $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a strong unconditional basis in X, and $x = \sum_{j\in\mathcal{J}} \mathcal{P}_j x$ is the corresponding decomposition of x, then the sequence $\{\|\mathcal{P}_j x\|_X\}_{j\in\mathcal{J}} \in Y$ and*

$$\frac{1}{A}\|x\|_X \le \left\| \{\|\mathcal{P}_j x\|_X\}_{j\in\mathcal{J}} \right\|_Y \le A\|x\|_X .$$

Proof. Let again $\mathcal{J}_k$, $k = 1, 2, \dots$ be an increasing sequence of finite sets such that $\cup_k \mathcal{J}_k = \mathcal{J}$. Since the partial sums $x^k = \sum_{j\in\mathcal{J}_k} \mathcal{P}_j x \in \mathcal{E}$ form a Cauchy sequence in X, the norm equivalence property implies that the corresponding sequences $\{\|\mathcal{P}_j x^k\|_X\}_{j\in\mathcal{J}}$, $k = 1, 2, \dots$ form a Cauchy sequence in Y. Therefore they converge to some limit $y = \{y_j\}_{j\in\mathcal{J}} \in Y$ in the norm of Y.

On the other hand $\mathcal{P}_j x^k = \mathcal{P}_j x$ whenever $\mathcal{J}_k \ni j$. Thus for a fixed j we have $\|\mathcal{P}_j x^k\|_X \to \|\mathcal{P}_j x\|_X$ as $k \to \infty$, and therefore $y_j = \|\mathcal{P}_j x\|_X$ for all $j \in \mathcal{J}$.

The norm equivalence can be easily obtained by taking a limit as $k \to \infty$ in

$$\frac{1}{A}\left\|x^k\right\|_X \le \|\{\|\mathcal{P}_j x^k\|\}_{j\in\mathcal{J}}\|_Y \le A\left\|x^k\right\|_X .$$

□

4.1.5. Unconditional bases vs strong unconditional bases

Probably the notion of strong unconditional basis is new for the reader, and he is more familiar with the notion of *unconditional basis.*

Let us recall that the system of subspaces $\mathcal{E}_j$ is called an *unconditional basis* if any vector $x \in X$ admits a unique representation

$$x = \sum_{j\in\mathcal{J}} x_j , \qquad x_j \in \mathcal{E}_j ,$$

where the series converges unconditionally (in the norm of X). Proposition 4.1 implies that a strong unconditional basis is an unconditional basis as well.

The converse is true if X is a Hilbert space, or if $\dim \mathcal{E}_j = 1$ for all j, but is not true in general. Indeed, let $X = \ell^2 \oplus \ell^3$. Let $\{e_j\}_{j\in\mathbb{N}}$ be the standard basis in ℓ^2 (and ℓ^3 as well), namely $e_j = (0, 0, \dots, 0, \underset{j}{1}, 0, 0, 0, \dots)$. Define two-dimensional subspaces $\mathcal{E}_j$ as $\mathcal{E}_j = \mathcal{L}\{e_j \oplus 0, 0 \oplus e_j\}$. Clearly, $\{\mathcal{E}_j\}_{j\in\mathbb{N}}$ is an unconditional basis in $\ell^2 \oplus \ell^3$, but it is not a strong unconditional basis.

4.2. Strong unconditional bases in reflexive spaces

Throughout the rest of this section we will assume that X is a *reflexive* Banach space.

4.2.1. Unconditional convergence of series $\sum_{j\in\mathcal{J}} x_j$, $x_j \in \mathcal{E}_j$

Lemma 4.3. *Let $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ be a strong unconditional basis in a reflexive Banach space X, and let an ideal sequence space Y be its coefficient space. Let $x_j \in \mathcal{E}_j$ and $\left\{\|x_j\|_X\right\}_{j\in\mathcal{J}} \in Y$. Then the series*

$$\sum_{j\in\mathcal{J}} x_j$$

converges unconditionally to some $x \in X$, and, of course, $x_j = \mathcal{P}_j x$.

Proof. Let $\mathcal{J}_k$ be an increasing ($\mathcal{J}_k \subset \mathcal{J}_{k+1}$) sequence of finite subsets of $\mathcal{J}$, $\cup_k \mathcal{J}_k = \mathcal{J}$. The assumption $\left\{\|x_j\|_X\right\}_{j\in\mathcal{J}} \in Y$ implies that the norms of vectors $x^k := \sum_{j\in\mathcal{J}_k} x_j$ are uniformly bounded. Since the unit ball in X is weakly compact, the sequence x^k, $k = 1, 2, \ldots$ has a weak cluster point, say $x \in X$.

So, for a fixed $j \in \mathcal{J}$ vector $\mathcal{P}_j x$ is a weak cluster point of $\mathcal{P}_j x^k$, $k = 1, 2, \ldots$. But $\mathcal{P}_j x^k = x_j$ whenever $j \in \mathcal{J}_k$, and therefore $x_j = \mathcal{P}_j x$. Applying Proposition 4.1, we get the conclusion of the lemma. □

Remark 4.4. Lemma 4.3 together with Corollary 4.2 immediately imply that a sequence $\{\alpha_j\}_{j\in\mathcal{J}}$ belongs to the coefficient space Y if and only if there exists a vector $x \in X$ such that $|\alpha_j| = \|\mathcal{P}_j x\|_X$ for every $j \in \mathcal{J}$, and moreover, for every such x

$$\frac{1}{A}\|x\|_X \le \left\|\{\alpha_j\}_{j\in\mathcal{J}}\right\|_Y \le A\|x\|_X\,.$$

This means that the coefficient space Y is unique up to replacement of the norm $\|\cdot\|_Y$ by an equivalent one.

4.2.2. Dual basis

Proposition 4.5. *Let $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ be a strong unconditional basis in a reflexive Banach space X, a sequence space Y be its coefficient space and let $\mathcal{P}_j$, $j \in \mathcal{J}$ be the corresponding coordinate projections. Then the system of subspaces $\mathcal{E}_j^* := \mathcal{P}_j^* X^*$, $j \in \mathcal{J}$ is a strong unconditional basis in X^*. Moreover the coefficient space for the basis $\{\mathcal{E}_j^*\}_{j\in\mathcal{J}}$ is exactly Y^*, that means Y^* is an ideal space of sequences and the norms $\|x^*\|_{X^*}$ and $\left\|\{\|\mathcal{P}_j^* x^*\|_{X^*}\}_{j\in\mathcal{J}}\right\|_{Y^*}$ are equivalent.*

Proof. First of all note that $\left(\mathcal{P}_j^*\right)^2 = \left(\mathcal{P}_j^2\right)^* = \mathcal{P}_j^*$, so $\mathcal{P}_j^*$ are bounded projections. This immediately implies that $\mathcal{E}_j^*$ are non-trivial closed subspaces of X^*.

Let us now show that the system $\{\mathcal{E}_j^*\}_{j\in\mathcal{J}}$ is total in X^*, i.e. that its linear span is dense in X^*. Suppose that a vector $x \in X^{**} = X$ satisfies

$$(x, x^*) = 0\,, \qquad \forall x^* \in \mathcal{E}_j^*\,,\ \forall j \in \mathcal{J}\,.$$

Then for all $x^* \in X^*$ and for the above vector x we have $(\mathcal{P}_j x, x^*) = (x, \mathcal{P}_j^* x^*) = 0$. This means $\mathcal{P}_j x = 0$ for all j. Since by Proposition 4.1 the series $\sum_{j\in\mathcal{J}} \mathcal{P}_j x$ converges unconditionally to x, one can conclude that $x = 0$.

Consider the so called *formal dual* to Y space $Y^\sharp$, i.e. the set of all numerical sequences $\{\beta_j\}_{j\in\mathcal{J}}$ for which the norm

$$\|\{\beta_j\}_{j\in\mathcal{J}}\|_{Y^\sharp} = \sup\Big\{\sum_{j\in\mathcal{J}} |\alpha_j|\cdot|\beta_j| \ :\ \{\alpha_j\}_{j\in\mathcal{J}} \in Y\,,\ \big\|\{\alpha_j\}_{j\in\mathcal{J}}\big\|_Y \le 1\Big\}$$

is finite. Clearly $Y^\sharp$ is an ideal Banach space of sequences. Note that every sequence $\{\beta_j\}_{j\in\mathcal{J}} \in Y^\sharp$ defines a bounded linear functional $\{\alpha_j\}_{j\in\mathcal{J}} \mapsto \sum_{j\in\mathcal{J}}\alpha_j\beta_j$ on Y and $\|\{\beta_j\}_{j\in\mathcal{J}}\|_{Y^*} = \|\{\beta_j\}_{j\in\mathcal{J}}\|_{Y^\sharp}$.

Let us now check the norm equivalence property for $Y^\sharp$. The reader might expect that it is just a dual reformulation of the norm equivalence property for X and Y. It is interesting that it is not exactly the case and the true (sharp) dual reformulation involves the norms different from $\|\mathcal{P}_j^* x^*\|_{X^*}$.

Define

$$\|x^*\|_{(\mathcal{E}_j)^*} := \sup\Big\{|(x,x^*)| \ :\ x\in\mathcal{E}_j\,,\ \|x\|_X \le 1\Big\} \tag{4.1}$$

The notation comes from the fact that the supremum is exactly the norm of x^* considered as a functional on the space $\mathcal{E}_j$, i.e. as an element of the dual to $\mathcal{E}_j$ space $(\mathcal{E}_j)^*$ (note that $\mathcal{E}_j^*$ and $(\mathcal{E}_j)^*$ are *different*)

The exact dual reformulation of the norm equivalence property is the following lemma

Lemma 4.6. *Under the assumptions of Proposition 4.5*

$$\frac{1}{A}\|x^*\|_{X^*} \le \Big\|\big\{\|x^*\|_{(\mathcal{E}_j)^*}\big\}\Big\|_{Y^\sharp} \le A\|x^*\|_{X^*}\,.$$

Proof. Take any $x^*\in X^*$. Then for every $x\in X$

$$\big|(x,x^*)\big| = \big|(\sum_{j\in\mathcal{J}}\mathcal{P}_j x, x^*)\big| \le \sum_{j\in\mathcal{J}} |(\mathcal{P}_j x, x^*)| \le$$
$$\le \sum_{j\in\mathcal{J}} \|\mathcal{P}_j x\|_X\cdot\|x^*\|_{(\mathcal{E}_j)^*} \le \big\|\{\|\mathcal{P}_j x\|_X\}_{j\in\mathcal{J}}\big\|_Y \cdot \big\|\{\|x^*\|_{(\mathcal{E}_j)^*}\}_{j\in\mathcal{J}}\big\|_{Y^\sharp}\,.$$

Taking supremum over all $x\in X$, $\|x\|_X \le 1$ and recalling that by Corollary 4.2 $\big\|\{\|\mathcal{P}_j x\|_X\}_{j\in\mathcal{J}}\big\|_Y \le A\|x\|_X$, one can conclude that $\|x^*\|_{X^*} \le A\cdot\big\|\{\|x^*\|_{(\mathcal{E}_j)^*}\}_{j\in\mathcal{J}}\big\|_{Y^\sharp}$.

To get the opposite estimate consider a sequence $\{\alpha_j\}_{j\in\mathcal{J}}\in Y$, $\big\|\{\alpha_j\}_{j\in\mathcal{J}}\big\|_Y = 1$. Take $r\in(0,1)$ close to 1. For each j there exists a vector $x_j\in\mathcal{E}_j$ such that $\|x_j\|_X = |\alpha_j|$ and $(x_j\,,x^*) \ge r|\alpha_j|\cdot\|x^*\|_{(\mathcal{E}_j)^*}$. By Lemma 4.3 and Corollary 4.2 the vector $x=\sum_{j\in\mathcal{J}}x_j$ belongs to X and $\|x\|_X\le A$. Therefore

$$r\sum_{j\in\mathcal{J}}|\alpha_j|\cdot\|x^*\|_{(\mathcal{E}_j)^*} \le \sum_{j\in\mathcal{J}}(x_j\,,x^*) = \big(\sum_{j\in\mathcal{J}}x_j\,,x^*\big) =$$
$$= (x,x^*) \le \|x\|_X\cdot\|x^*\|_{X^*} \le A\|x^*\|_{X^*}\,.$$

As the sequence $\{\alpha_j\}_{j\in\mathcal{J}}$ was arbitrary, $\big\|\{\|x^*\|_{(\mathcal{E}_j)^*}\}_{j\in\mathcal{J}}\big\|_{Y^\sharp} \le Ar^{-1}\|x^*\|_{X^*}$. Since $r<1$ was arbitrary as well, we get $\big\|\{\|x^*\|_{(\mathcal{E}_j)^*}\}_{j\in\mathcal{J}}\big\|_{Y^\sharp} \le A\|x^*\|_{X^*}$. $\square$

We are now going to complete the proof of Proposition 4.5. Though the sequence $\{\|x^*\|_{(\mathcal{E}_j)^*}\}_{j\in\mathcal{J}}$ does not coincide with the desired sequence $\{\|\mathcal{P}_j^* x^*\|_{X^*}\}_{j\in\mathcal{J}}$, they do not differ too much.

Indeed, since for every $x \in \mathcal{E}_j$ we have

$$|(x, x^*)| = |(x, \mathcal{P}_j^* x^*)| \le \|\mathcal{P}_j^* x^*\|_{X^*} \cdot \|x\|_X ,$$

we can conclude that always $\|x^*\|_{(\mathcal{E}_j)^*} \le \|\mathcal{P}_j^* x^*\|_{X^*}$.

On the other hand, for every $x \in X$ we have $\mathcal{P}_j x \in \mathcal{E}_j$ and thereby

$$|(x, \mathcal{P}_j^* x^*)| = |(\mathcal{P}_j x, x^*)| \le \|\mathcal{P}_j x\|_X \cdot \|x^*\|_{(\mathcal{E}_j)^*} \le \|\mathcal{P}_j\| \cdot \|x\|_X \cdot \|x^*\|_{(\mathcal{E}_j)^*} .$$

Thus $\|\mathcal{P}_j^* x^*\|_{X^*} \le \|\mathcal{P}_j\| \cdot \|x^*\|_{(\mathcal{E}_j)^*} \le A^2 \|x^*\|_{(\mathcal{E}_j)^*}$ (recall that (see Sec. 4.1.4) $\|\mathcal{P}_j\| \le A^2$). This results in

$$\Big\|\{\|x^*\|_{(\mathcal{E}_j)^*}\}\Big\|_{Y^\sharp} \le \Big\|\{\|\mathcal{P}_j x^*\|_{X^*}\}\Big\|_{Y^\sharp} \le A^2 \Big\|\{\|x^*\|_{(\mathcal{E}_j)^*}\}\Big\|_{Y^\sharp}$$

and finally in

$$\frac{1}{A}\|x^*\|_{X^*} \le \Big\|\{\|\mathcal{P}_j x^*\|_{X^*}\}_{j\in\mathcal{J}}\Big\|_{Y^\sharp} \le A^3 \|x^*\|_{X^*} .$$

It remains to show that the formal dual $Y^\sharp$ coincides with the true dual space Y^* (i.e. that Y^* contains nothing like Banach limits in ℓ^∞). Take any $y^* \in Y^*$. Fix some vectors $x_j \in \mathcal{E}_j$, $\|x_j\| = 1$ and consider the mapping $\tau : Y \to X$,

$$\tau\{\alpha_j\}_{j\in\mathcal{J}} = \sum_{j\in\mathcal{J}} \alpha_j x_j .$$

Lemma 4.3 implies that the mapping τ is well defined (the series converges for all sequences $\{\alpha_j\}_{j\in\mathcal{J}} \in Y$) and for every sequence $\alpha = \{\alpha_j\}_{j\in\mathcal{J}} \in Y$

$$\frac{1}{A}\|\alpha\|_Y \le \|\tau\alpha\|_X \le A\|\alpha\|_Y .$$

Therefore the functional $x_0^* : \tau Y \ni x \mapsto (\tau^{-1}x, y^*)$ is well defined and has the norm at most $A\|y^*\|_{Y^*}$. By the Hahn–Banach theorem x_0^* can be extended to a bounded linear functional x^* on the entire X. Therefore for every sequence $\alpha = \{\alpha_j\}_{j\in\mathcal{J}} \in Y$ we get

$$(\alpha, y^*) = (\tau\alpha, x^*) = \Big(\sum_{j\in\mathcal{J}} \alpha_j x_j\, ,\, x^*\Big) = \sum_{j\in\mathcal{J}} \alpha_j (x_j, x^*) .$$

It remains to note that the sequence $\big\{(x_j, x^*)\big\}_{j\in\mathcal{J}}$ is majorized by the sequence $\big\{\|x^*\|_{(\mathcal{E}_j)^*}\big\}_{j\in\mathcal{J}} \in Y^\sharp$ and therefore belongs to $Y^\sharp$ as well.

$\square$

4.2.3. Reflexivity of the coefficient space.

If we start from the basis $\{\mathcal{E}_j^*\}_{j\in\mathcal{J}}$ in the dual space X^* and construct a dual to it basis $\{\mathcal{E}_j^{**}\}_{j\in\mathcal{J}}$ in $X^{**}=X$, we just will get the system $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ again, but the coefficient space will be now $Y^{**}=Y^{\sharp\sharp}$ instead of Y.

Since the coefficient space is unique up to the norm equivalence (see Remark 4.4), we conclude that spaces Y and $Y^{**}=Y^{\sharp\sharp}$ coincide as sets. Since for every Banach space Y the norms $\|\cdot\|_Y$ and $\|\cdot\|_{Y^{**}}$ coincide on Y, we observe that $Y=Y^{**}=Y^{\sharp\sharp}$, i.e. the norms coincide as well. This means that the space Y is reflexive and formally reflexive.

In particular it means that a sequence $\{\alpha_j\}_{j\in\mathcal{J}}$ belongs to Y if and only if

$$\sup\Big\{\sum_{j\in\mathcal{J}}|\alpha_j|\cdot|\beta_j|\ :\ \{\beta_j\}_{j\in\mathcal{J}}\in Y^*,\ \big\|\{\beta_j\}_{j\in\mathcal{J}}\big\|_{Y^*}\le 1\Big\}<\infty\,,$$

and moreover, the norm $\big\|\{\alpha_j\}_{j\in\mathcal{J}}\big\|_Y$ is equal to this supremum. Recalling (see Remark 4.4) that $\{\beta_j\}_{j\in\mathcal{J}}\in Y^*$ if and only if $|\beta_j|=\|\mathcal{P}_j^*x^*\|_{X^*}$ for some $x^*\in X^*$ and that the norms $\big\|\{\|\mathcal{P}_j^*x^*\|_{X^*}\big\|_{Y^\sharp}$ and $\|x^*\|_{X^*}$ are equivalent, we can reformulate the last property as follows: *a sequence $\{\alpha_j\}_{j\in\mathcal{J}}$ belongs to Y if and only if*

$$\sup\Big\{\sum_{j\in\mathcal{J}}|\alpha_j|\cdot\|\mathcal{P}_j^*x^*\|_{X^*}\ :\ \|x^*\|_{X^*}\le 1\Big\}\ <\infty$$

and the norm $\big\|\{\alpha_j\}_{j\in\mathcal{J}}\big\|_Y$ is equivalent to this supremum.

4.2.4. "Parseval identity" and "Bessel inequality"

Lemma 4.5 immediately implies the following generalization of a well known Hilbert space result:

Lemma 4.7. *If X is a reflexive Banach space, $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a strong unconditional basis in X, and $\mathcal{P}_j$ are the corresponding coordinate projections, then for any $x\in X$, $x^*\in X^*$*

$$(x\,,\,x^*)=\sum_{j\in\mathcal{J}}(\mathcal{P}_jx\,,\,\mathcal{P}_j^*x^*)\tag{4.2}$$

(the" Parseval identity"), where the series converges in a very strong sense, namely

$$\sum_{j\in\mathcal{J}}\|\mathcal{P}_jx\|_X\cdot\|\mathcal{P}_j^*x^*\|_{X^*}\le C\|x\|_X\cdot\|x^*\|_{X^*}\tag{4.3}$$

(the "Bessel inequality").

4.3. Criteria for a strong unconditional basis

Now we will give several simple necessary and sufficient conditions that guarantee that a given system of subspaces $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a strong unconditional basis. Of course, our choice was dictated by the main example we are going to deal with: the space $X = L^p(W)$. As it usually happens, it is convenient to work simultaneously with the space X and with its dual X^*.

So, from the beginning we assume that we are given a family of bounded projections $\mathcal{P}_j$, $j \in \mathcal{J}$ in a reflexive Banach space X (projection means $\mathcal{P}_j^2 = \mathcal{P}_j$) satisfying $\mathcal{P}_{j'}\mathcal{P}_{j''} = 0$ for $j' \neq j''$ and the corresponding families of non-trivial closed subspaces $\mathcal{E}_j = \mathcal{P}_j X$, and $\mathcal{E}_j^* = \mathcal{P}_j^* X^*$.

Note that we do not *a priori* assume that the system $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is total (spans all X).

Lemma 4.8 (Criterion 1). *Under the above assumptions the system $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a strong unconditional basis in X if and only if for any $x \in X$ and for any $x^* \in X^*$ the "Bessel inequality"* (4.3) *and the "Parseval identity"* (4.2) *hold.*

Proof. The "only if" part is just Lemma 4.7. Let us prove the "if" part.

Let Y be the space of sequences $\{\alpha_j\}_{j\in\mathcal{J}}$ for which

$$\left\|\{\alpha_j\}_{j\in\mathcal{J}}\right\|_Y := \sup\Big\{\sum_{j\in\mathcal{J}} |\alpha_j| \cdot \|\mathcal{P}_j^* x^*\|_{X^*} \ : \ x^* \in X^*,\ \|x^*\| \le 1\Big\}$$

is finite. Clearly Y is an ideal Banach space. The Bessel inequality implies that for every $x \in X$

$$\left\{\|\mathcal{P}_j x\|_X\right\}_{j\in\mathcal{J}} \in Y \qquad \text{and} \qquad \left\|\left\{\|\mathcal{P}_j x\|_X\right\}_{j\in\mathcal{J}}\right\|_Y \le C\|x\|_X\,.$$

To get the inverse inequality, choose a functional $x^* \in X^*$ such that $\|x^*\|_{X^*} = 1$ and $(x, x^*) = \|x\|_X$. Then

$$\|x\|_X = (x, x^*) = \sum_{j\in\mathcal{J}} (\mathcal{P}_j x, \mathcal{P}_j^* x^*) \le \sum_{j\in\mathcal{J}} \|\mathcal{P}_j x\|_X \cdot \|\mathcal{P}_j^* x^*\|_{X^*}$$

and therefore

$$\|x\|_X \le \left\|\left\{\|\mathcal{P}_j x\|_X\right\}_{j\in\mathcal{J}}\right\|_Y\,.$$

It remains only to check that the linear span of $\mathcal{E}_j$ is dense in X. If $x^* \in X^*$ vanishes on all $\mathcal{E}_j$, $j \in \mathcal{J}$, then $\mathcal{P}_j^* x^* = 0$ for every $j \in \mathcal{J}$, and therefore for any $x \in X$

$$(x, x^*) = \sum_{j\in\mathcal{J}} (\mathcal{P}_j x\,,\, \mathcal{P}_j^* x^*) = 0,$$

i.e. $x^* = 0$. □

The following (almost trivial) observation will help us a lot.

Remark 4.9. Note that in the previous lemma instead of checking the "Parseval identity" for *all* pairs $x \in X$, $x^* \in X^*$ it is enough to check it only for x and x^* in some dense subsets $\widetilde{X} \subset X$ and $\widetilde{X^*} \subset X^*$ respectively.

Indeed, the "Bessel inequality" implies that the series $\sum_{j\in\mathcal{J}}(\mathcal{P}_j x, \mathcal{P}_j^* x^*)$ converges absolutely and defines a continuous bilinear form on $X \times X^*$. On the dense set $\widetilde{X} \times \widetilde{X^*}$ this form coincides with (x, x^*), so by continuity they coincide on the entire $X \times X^*$, i.e the "Parseval identity" holds for all $x \in X$, $x^* \in X^*$.

Remark 4.10. Clearly, the "Bessel inequality" may also be checked on dense sets only. However, in what follows we will not use this fact.

4.3.1. Another criterion

The next criterion will allow us to prove that $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a basis with a given coefficient space. Let $\mathcal{E}_j$, $\mathcal{E}_j^*$, $\mathcal{P}_j$, $\mathcal{P}_j^*$ be as above. Let Y be an ideal Banach space of sequences and $Y^\sharp$ be its formal dual.

Lemma 4.11. *The system $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a strong unconditional basis in X with the coefficient space Y if and only if the "Parseval identity"*

$$(x\,,\,x^*) = \sum_j (\mathcal{P}_j x\,,\,\mathcal{P}_j^* x^*)$$

holds for all x and x^* in some dense sets in X and X^* respectively, and the following two "imbedding theorems" hold

1. *for every $x \in X$ the sequence $\left\{\|\mathcal{P}_j x\|_X\right\}_{j\in\mathcal{J}} \in Y$ and*

$$\left\|\left\{\|\mathcal{P}_j x\|_X\right\}_{j\in\mathcal{J}}\right\|_Y \le C\|x\|_X\,;$$

2. *for every $x^* \in X^*$ the sequence $\left\{\|\mathcal{P}_j^* x^*\|_{X^*}\right\}_{j\in\mathcal{J}} \in Y^\sharp$ and*

$$\left\|\left\{\|\mathcal{P}_j^* x^*\|_{X^*}\right\}_{j\in\mathcal{J}}\right\|_{Y^\sharp} \le C\|x\|_{X^*}\,.$$

Proof of Lemma 4.11. The only if part is trivial, because the 'imbedding theorems" are just "halves" of the norm equivalence property for X and X^*. Let us prove the "if" part.

The "imbedding theorems" 1 and 2 together immediately imply the "Bessel inequality", so $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a strong unconditional basis. It remains to check that Y serves as the coefficient space.

The "imbedding theorem" 1 is exactly one half of the norm equivalence property. To get another half (the opposite estimate) we can use duality and the "Parseval identity". Let us remind that because of the "Bessel inequality" it holds now for all $x \in X$, $x^* \in X^*$. Fix

$x \in X$ and take $x^* \in X^*$ such that $\|x^*\|_{X^*} = 1$ and $(x, x^*) = \|x\|_X$. Using the imbedding 2 we get

$$\|x\|_X = (x, x^*) = \sum_{j\in\mathcal{J}} (\mathcal{P}_j x\,,\, \mathcal{P}_j^* x^*) \le \sum_{j\in\mathcal{J}} \|\mathcal{P}_j x\|_X \cdot \|\mathcal{P}_j^* x^*\|_{X^*} \le$$
$$\le \left\| \left\{ \|\mathcal{P}_j x\|_X \right\}_{j\in\mathcal{J}} \right\|_Y \cdot \left\| \left\{ \|\mathcal{P}_j^* x^*\|_{X^*} \right\}_{j\in\mathcal{J}} \right\|_{Y^\sharp} \le$$
$$\le C\|x^*\|_{X^*} \cdot \left\| \left\{ \|\mathcal{P}_j x\|_X \right\}_{j\in\mathcal{J}} \right\|_Y = C \left\| \left\{ \|\mathcal{P}_j x\|_X \right\}_{j\in\mathcal{J}} \right\|_Y .$$

□

Remark. If the assumptions of the lemma are satisfied, then automatically (see Proposition 4.5) the formal dual $Y^\sharp$ coincides with the true dual Y^*.

We will need another form of the basis criterion, now in terms of norms $\|x^*\|_{(\mathcal{E}_j)^*}$ and $\|x\|_{(\mathcal{E}_j^*)^*}$. Recall (see the definition(4.1) before Lemma 4.6) that the the norm $\|x^*\|_{(\mathcal{E}_j)^*}$ is the norm of x^* treated as a linear functional on $\mathcal{E}_j$. Similarly, $\|x\|_{(\mathcal{E}_j^*)^*}$ is the norm of x treated as a linear functional on $\mathcal{E}_j^*$.

Lemma 4.12. *The system $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a strong unconditional basis in X with the coefficient space Y if and only if*

$$\sup_j \|\mathcal{P}_j\| < \infty\,;$$

the "Parseval identity"

$$(x\,,\, x^*) = \sum_j (\mathcal{P}_j x\,,\, \mathcal{P}_j^* x^*)$$

holds for all x and x^* in some dense sets in X and X^* respectively; and the following two "imbedding theorems" hold:

1. *for every $x^* \in X^*$ the sequence $\left\{ \|x^*\|_{(\mathcal{E}_j)^*} \right\}_{j\in\mathcal{J}} \in Y^\sharp$ and*
$$\left\| \left\{ \|x^*\|_{(\mathcal{E}_j)^*} \right\}_{j\in\mathcal{J}} \right\|_{Y^\sharp} \le\; C\|x^*\|_{X^*}\,;$$
2. *for every $x \in X$ the sequence $\left\{ \|x\|_{(\mathcal{E}_j^*)^*} \right\}_{j\in\mathcal{J}} \in Y$ and*
$$\left\| \left\{ \|x\|_{(\mathcal{E}_j^*)^*} \right\}_{j\in\mathcal{J}} \right\|_Y \le\; C\|x\|_X\,.$$

Proof. As we saw above (when proving Lemma 4.5),

$$\|x^*\|_{(\mathcal{E}_j)^*} \le \|\mathcal{P}_j^* x^*\|_{X^*} \le \|\mathcal{P}_j\|\|x^*\|_{(\mathcal{E}_j)^*}\,.$$

Similar reasoning yields

$$\|x\|_{(\mathcal{E}_j^*)^*} \le \|\mathcal{P}_j x\|_X \le \|\mathcal{P}_j\|\|x\|_{(\mathcal{E}_j^*)^*}\,.$$

Thus, if we are *a priori* given that $\sup_j \|\mathcal{P}_j\| < \infty$, the criterion to prove coincides with the previous one. It remains to notice that the condition $\sup_j \|\mathcal{P}_j\| < \infty$ is necessary for the system $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ be a strong unconditional basis in X.

□

Remark 4.13. Note that (without any assumptions about subspaces $\mathcal{E}_j$ or projectors $\mathcal{P}_j$) the "imbedding theorem" 1 from the above lemma is equivalent to the following inequality for X:

$$\Big\|\sum_{j\in\mathcal{J}} x_j\Big\|_X \le C\Big\|\big\{\|x_j\|_X\big\}_{j\in\mathcal{J}}\Big\|_Y \tag{4.4}$$

for all finite sums $\sum_{j\in\mathcal{J}} x_j$, $x_j \in \mathcal{E}_j$.

Indeed, suppose that the "imbedding theorem" 1 holds. Take any $x^* \in X^*$. Then

$$\Big|\big(\sum_{j\in\mathcal{J}} x_j, x^*\big)\Big| \le \sum_{j\in\mathcal{J}} |(x_j, x^*)| \le \sum_{j\in\mathcal{J}} \|x_j\|_X \cdot \|x^*\|_{(\mathcal{E}_j)^*} \le$$
$$\le \big\|\{\|x_j\|_X\}_{j\in\mathcal{J}}\big\|_Y \cdot \big\|\{\|x^*\|_{(\mathcal{E}_j)^*}\}_{j\in\mathcal{J}}\big\|_{Y^\sharp} \le C\cdot \big\|\{\|x_j\|_X\}_{j\in\mathcal{J}}\big\|_Y \cdot \|x^*\|_{X^*},$$

so (4.4) holds.

Suppose now that (4.4) holds. Consider a sequence $\{\alpha_j\}_{j\in\mathcal{J}} \in Y$, $\big\|\{\alpha_j\}_{j\in\mathcal{J}}\big\|_Y = 1$ with finitely many non-zero terms. Take $r \in (0,1)$ close to 1. For each j there exists a vector $x_j \in \mathcal{E}_j$ such that $\|x_j\|_X = |\alpha_j|$ and $(x_j\,,\, x^*) \ge r|\alpha_j|\cdot\|x^*\|_{(\mathcal{E}_j)^*}$. Therefore

$$r\sum_{j\in\mathcal{J}} |\alpha_j|\cdot\|x^*\|_{(\mathcal{E}_j)^*} \le \sum_{j\in\mathcal{J}} (x_j\,,\, x^*) = \Big(\sum_{j\in\mathcal{J}} x_j\,,\, x^*\Big) \le \Big\|\sum_{j\in\mathcal{J}} x_j\Big\|_X \cdot \|x^*\|_{X^*} \le$$
$$\le C\Big\|\big\{\|x_j\|_X\big\}_{j\in\mathcal{J}}\Big\|_Y \cdot \|x^*\|_{X^*} = C\big\|\{\alpha_j\}_{j\in\mathcal{J}}\big\|_Y \cdot \|x^*\|_{X^*} = C\|x^*\|_{X^*}.$$

As the sequence $\{\alpha_j\}_{j\in\mathcal{J}}$ was arbitrary, $\big\|\{\|x^*\|_{(\mathcal{E}_j)^*}\}_{j\in\mathcal{J}}\big\|_{Y^\sharp} \le Cr^{-1}\|x^*\|_{X^*}$. Since $r<1$ was arbitrary as well, we get $\big\|\{\|x^*\|_{(\mathcal{E}_j)^*}\}_{j\in\mathcal{J}}\big\|_{Y^\sharp} \le C\|x^*\|_{X^*}$.

Remark 4.14. Note that in the case of a Hilbert space the norm $\|x^*\|_{(\mathcal{E}_j)^*}$ is simply the norm of the orthogonal projection of x^* onto $\mathcal{E}_j$.

5. Haar system as a strong unconditional basis in weighted L^p spaces: reduction to the imbedding theorem.

Now let us turn from an abstract Banach space X and a general system of subspaces $\mathcal{E}_j$ and projections $\mathcal{P}_j$ to our particular case $X = L^p(W) = L^p(\rho_t)$, $\mathcal{E}_I = h_I(\cdot)E = \{h_I(\cdot)\mathbf{e} : \mathbf{e} \in E\}$. The corresponding projections $\mathcal{P}_I$, $I \in \mathcal{D}$ are given by

$$\big(\mathcal{P}_I\varphi\big)(\cdot) = h_I(\cdot)\frac{1}{|I|}\int_I h_I(t)\varphi(t)dt = c_I(\varphi)h_I(\cdot).$$

In general the projections $\mathcal{P}_I$ can be defined only on a dense subset of $L^p(W)$ (say, $L^2 \cap L^p(W)$). But if the weight W satisfies the (dyadic) $(\mathcal{A}_{p,q})$ condition $(1/p + 1/q = 1)$

$$\langle \rho^* \rangle_{I,q} \le C \langle \rho \rangle^*_{I,p} \qquad \text{for all } I \in \mathcal{D}\,, \tag{$\mathcal{A}^{\mathrm{dy}}_{p,q}$}$$

then Lemma 2.1 shows that the definition makes sense for every $\varphi \in L^p(W)$. Moreover, it implies that the operators $\mathcal{P}_I$ are uniformly bounded,

$$\left\| \mathcal{P}_I \right\|_{L^p(W) \to L^p(W)} = \sup_{\mathbf{e} \in E \setminus \{0\}} \frac{\langle \rho^* \rangle_{I,q}(\mathbf{e})}{\langle \rho \rangle^*_{I,p}(\mathbf{e})} \le C < \infty\,.$$

We are going to show that the dyadic Muckenhoupt $(\mathcal{A}^{\mathrm{dy}}_{p,q})$ condition implies that the system $\{\mathcal{E}_I\}_{I \in \mathcal{D}}$ is a *strong unconditional basis* in $L^p(W)$.

The dual space to X is $X^* = L^q(W^{-p/q}) = L^q(\rho^*_t)$, $(1/p + 1/q = 1)$ and the duality is given by the formula

$$(\varphi, \psi) = \int_{\mathbb{R}} (\varphi(t), \psi(t))_E\, dt\,, \qquad \varphi \in L^p(W),\ \psi \in L^q(W^{-q/p})\,.$$

In this notation the dual projections $\mathcal{P}^*_I$ are just the same $\mathcal{P}_I$, but considered now as operators in the dual space $X^* = L^q(W^{-q/p})$. Also the dual subspaces $\mathcal{E}^*_I = \mathcal{E}_I$ (in the same sense).

We have no trouble with checking the "Parseval identity"

$$(\varphi, \psi) = \sum_{I \in \mathcal{D}} (\mathcal{P}_I \varphi, \mathcal{P}_I \psi)$$

for $\varphi \in L^2 \cap L^p(W)$, $\psi \in L^2 \cap L^q(W^{-q/p})$, because it is true for *any* φ, ψ in (non-weighted!) L^2.

Note that $L^2 \cap L^p(W)$ is dense in $L^p(W)$ and $L^2 \cap L^q(W^{-q/p})$ is dense in $L^q(W^{-q/p})$.

Since we already know that the projections $\mathcal{P}_I$ are uniformly bounded, it seems a good idea to use Criterion 3 (Lemma 4.12). Another, much deeper reason for this choice will be given later in Section 6.1.

Of course to check the imbeddings (the only thing that remains to do), we have to guess somehow what the coefficient space Y should be.

We will discuss below where the right choice of Y may come from. Now let us carry out the work we should do anyhow once we decided to apply Lemma 4.12: let us compute the norms $\|\varphi\|_{(\mathcal{E}^*_I)^*}$ and $\|\psi\|_{(\mathcal{E}_I)^*}$ $(\varphi \in L^p(W),\ \psi \in L^q(W^{-q/p}))$.

5.1. Computation of $\|\psi\|_{(\mathcal{E}_I)^*}$

Let $\psi \in L^q(W^{-q/p})$. For every $\varphi = h_I \mathbf{e} \in \mathcal{E}_I \subset L^p(W)$ $(\mathbf{e} \in E)$ we have

$$\|\varphi\|_{L^p(W)} = \left\{ \int_{\mathbb{R}} \left[\rho_t(h_I(t)\mathbf{e})\right]^p dt \right\}^{1/p} = \left\{ \int_I \left[\rho_t(\mathbf{e})\right]^p dt \right\}^{1/p} = |I|^{1/p} \langle \rho \rangle_{I,p}(\mathbf{e})\,.$$

Also

$$(\varphi, \psi) = \int_{\mathbb{R}} (h_I(t)\mathbf{e}, \psi(t))_E dt = |I| \cdot (\mathbf{e}, c_I(\psi))_E .$$

Thus we get

$$\|\psi\|_{(\mathcal{E}_I)^*} = |I|^{1/q} \sup_{\mathbf{e}\in E\setminus\{0\}} \frac{|(\mathbf{e}, c_I(\psi))_E|}{\langle\rho\rangle_{I,p}(\mathbf{e})} = |I|^{1/q} \langle\rho\rangle^*_{I,p}(c_I(\psi)) .$$

Similarly

$$\|\varphi\|_{(\mathcal{E}^*_I)^*} = |I|^{1/p} \langle\rho^*\rangle^*_{I,q}(c_I(\varphi)) .$$

5.2. About the coefficient space Y in the case of Hilbert space

Note that in the case of a Hilbert space there is no question about the coefficient space Y. Indeed, if a system $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a strong unconditional basis in a *Hilbert* space X (we return for a second to our abstract setting), then for every combination $\sum_j \varepsilon_j x_j$ ($\varepsilon_j = \pm 1$, $x_j \in \mathcal{E}_j$) the norms $\|\sum_j \varepsilon_j x_j\|_X$ and $\big\|\{\|x_j\|_X\}_{j\in\mathcal{J}}\big\|_Y$ should be equivalent.

Taking the average of $\|\sum_j \varepsilon_j x_j\|^2_X$ over all possible $\varepsilon_j = \pm 1$, we get

$$\sum_{j\in\mathcal{J}} \|x_j\|^2_X .$$

Since the norm $\big\|\{\|x_j\|_X\}_{j\in\mathcal{J}}\big\|_Y$ does not depend on the choice of ε_j, one can conclude that the coefficient space is isomorphic to ℓ^2.

6. Haar system as an unconditional basis in a weighted L^2 space: the scalar case.

Consider now the simplest case of the scalar weighted L^2 space. To distinguish with the vector case we will use the letter w (instead of W) for a weight. Recall that (see Section 2.4) in this case the condition $(\mathcal{A}^{\mathrm{dy}}_{2,2})$ can be rewritten as

$$\sup_{I\,:\,I\in\mathcal{D}} \langle w\rangle_I \langle w^{-1}\rangle_I = A < \infty . \tag{A^{dy}_2}$$

Multiplying and dividing by $\exp\{\langle\log w\rangle_I\}$, we can rewrite the (A^{dy}_2) condition as

$$\sup_{I\,:\,I\in\mathcal{D}} \Big[\langle w\rangle_I \exp\{-\langle\log w\rangle_I\}\Big] \Big[\langle w^{-1}\rangle_I \exp\{-\langle\log w^{-1}\rangle_I\}\Big] < \infty .$$

The inequality between arithmetic and geometric means implies that the expressions in brackets are at least 1, so the condition (A^{dy}_2) can be splitted into two.

6.1. Imbedding theorem and (A_∞) condition

We will say that a weight w satisfies the (dyadic) (A_∞) condition if for every dyadic interval I

$$\langle w \rangle_I \le C \exp\{\langle \log w \rangle_I\} \qquad (A_\infty^{\mathrm{dy}})$$

(the inverse inequality $\langle w \rangle_I \ge \exp\{\langle \log w \rangle_I\}$ is true for *any* weight w). The best constant C is called the (A_∞^{dy}) norm of w.

Note that $w \in A_2^{\mathrm{dy}} \between w^{-1} \in A_2^{\mathrm{dy}} \Longleftrightarrow w \in A_\infty^{\mathrm{dy}} \,\&\, w^{-1} \in A_\infty^{\mathrm{dy}}$.

Let us write down the imbeddings from Lemma 4.12. For any $\varphi \in X^* = L^2(w^{-1})$ (*sic!*) the norm $\|\varphi\|_{(\mathcal{E}_I)^*}$ can be easily computed (see Section 5.1):

$$\|\varphi\|_{(\mathcal{E}_I)^*} = |I|^{1/2} \langle w \rangle_I^{-1/2} |c_I(\varphi)| \,,$$

and therefore the imbedding 1 from Lemma 4.12 can be rewritten as

$$\sum_{I \in \mathcal{D}} \langle w \rangle_I^{-1} |c_I(\varphi)|^2 |I| \le C \|\varphi\|_{L^2(w^{-1})} \,. \tag{6.1}$$

Theorem 6.1. *Let a (scalar) weight w satisfy the (A_∞^{dy}) condition. Then the imbedding* (6.1) *holds.*

This theorem immediately implies that if the weight w satisfies (A_2^{dy}) condition, then the Haar system is a strong unconditional basis (or, equivalently, unconditional basis) in $L^2(w)$. Indeed the imbedding 2 from Lemma 4.12 is simply (6.1) with roles of w and w^{-1} exchanged. Since (A_2^{dy}) implies (A_∞^{dy}) for both w and w^{-1}, the imbedding 2 also holds. And it has already been shown above that we only need to prove these two imbeddings.

Now we think the reader should understand why we use Lemma 4.12 to check that the system is a basis — in this case each imbedding needs only a "half" of the Muckenhoupt condition.

6.2. The Bellman function

The rest of Section 6 is devoted to the proof of Theorem 6.1. Since the inequality (6.1) is a typical "dyadic imbedding theorem", we are going to prove it using the Bellman function approach. Clearly it is enough to prove the imbedding theorem for real-valued functions (consider $\operatorname{Re}\varphi$ and $\operatorname{Im}\varphi$ separately).

Again we fix all averages involved, which are

$$F_I := \langle \varphi^2 w^{-1} \rangle_I \,, \qquad u_I := \langle w \rangle_I \,, \qquad \omega_I := \langle \log w \rangle_I \,, \qquad f_I := \langle \varphi \rangle_I \,.$$

Fix a bound C for the (A_∞^{dy}) norm of w and define the Bellman function $\mathcal{B} = \mathcal{B}_C$ corresponding to the problem as

$$\mathcal{B}(F, f, u, \omega) = \frac{1}{|I|} \sup \Big\{ \sum_{J \subset I} \frac{c_J(\varphi)^2}{\langle w \rangle_J} |J| \ : \ \varphi, w \text{ satisfy } \|w\|_{A_\infty^{\mathrm{dy}}} \le C \text{ and} \\ \langle \varphi^2 w^{-1} \rangle_I = F \,, \langle \varphi \rangle_I = f \,, \langle w \rangle_I = u \,, \langle \log w \rangle_I = \omega \ \Big\} \,.$$

6.2.1. Properties of the Bellman function.

1) Domain $D(C)$: $f^2 \le Fu$, $e^\omega \le u \le Ce^\omega$ (Cauchy inequality and the bound for (A_∞^{dy}) norm);
2) Range: $0 \le \mathcal{B}(F, f, u, \omega) \le C'F$ (our belief that the result is true);
3) Key property: for any three quadruplets (F, f, u, ω), $(F_\pm, f_\pm, u_\pm, \omega_\pm) \in D(C)$ satisfying
$$(F, f, u, \omega) = \frac{1}{2}\{(F_+, f_+, u_+, \omega_+) + (F_-, f_-, u_-, \omega_-)\}$$
the inequality
$$\mathcal{B}(F, f, u, \omega) \ge \frac{(f_+ - f_-)^2}{4u} + \frac{1}{2}\Big\{\mathcal{B}(F_+, f_+, u_+, \omega_+) + \mathcal{B}(F_-, f_-, u_-, \omega_-)\Big\}$$
holds; recall that $c_I(\varphi) = \dfrac{1}{|I|}\displaystyle\int_I \varphi h_I = \dfrac{\langle\varphi\rangle_{I_+} - \langle\varphi\rangle_{I_-}}{2} = \dfrac{f_+ - f_-}{2}$.

We leave the proof of all these properties as an exercise for the reader. The reasoning is very similar to that we used when proving the Carleson Imbedding Theorem.

Note that formally we even do not have to prove them. After we somehow get (guess) these properties, we just have to check that it is possible "to run the machine backward".

6.2.2. Running the machine backward.

Suppose we found *any* function $\mathcal{B}$ satisfying properties 1–3. Then we are done. Indeed, the key property 3 can be rewritten as
$$\frac{c_I(\varphi)^2}{\langle w\rangle_I} = \frac{(f_{I_+} - f_{I_-})^2}{4u_I} \le \mathcal{B}(F_I, f_I, u_I, \omega_I) - \frac{1}{2}\Big\{\mathcal{B}(F_{I_+}, f_{I_+}, u_{I_+}, \omega_{I_+}) + \mathcal{B}(F_{I_-}, f_{I_-}, u_{I_-}, \omega_{I_-})\Big\}.$$

We can draw a tree like that in Section 1.1.3, (see Figure 5)

Again, as on Fig. 2, the quantity on each vertical edge is at most the difference between the value at the upper vertex and the sum of values at two lower vertices.

Going n levels down on this picture, we get the inequality
$$\sum_{\substack{J:J\subset I\\ |J|\ge 2^{-n+1}|I|}} c_J(\varphi)^2\langle w\rangle_J^{-1}|J| \le |I|\cdot\mathcal{B}(F_I, f_I, u_I, \omega_I) - \sum_{\substack{J:J\subset I\\ |J|=2^{-n}|I|}} |J|\cdot\mathcal{B}(F_J, f_J, u_J, \omega_J).$$

Since $\mathcal{B}(F_J, f_J, u_J, \omega_J) \ge 0$ for every J, this gives
$$\sum_{J\subset I} c_J(\varphi)^2\langle w\rangle_J^{-1}|J| \le |I|\cdot\mathcal{B}(F_I, f_I, u_I, \omega_I) \le C'|I|F_I = C'\int_I \varphi^2 w^{-1}.$$

Applying the above estimate to intervals $[-2^n, 0]$ and $[0, 2^n]$ and taking the limit as $n \to \infty$, one can immediately obtain the desired imbedding theorem. $\square$

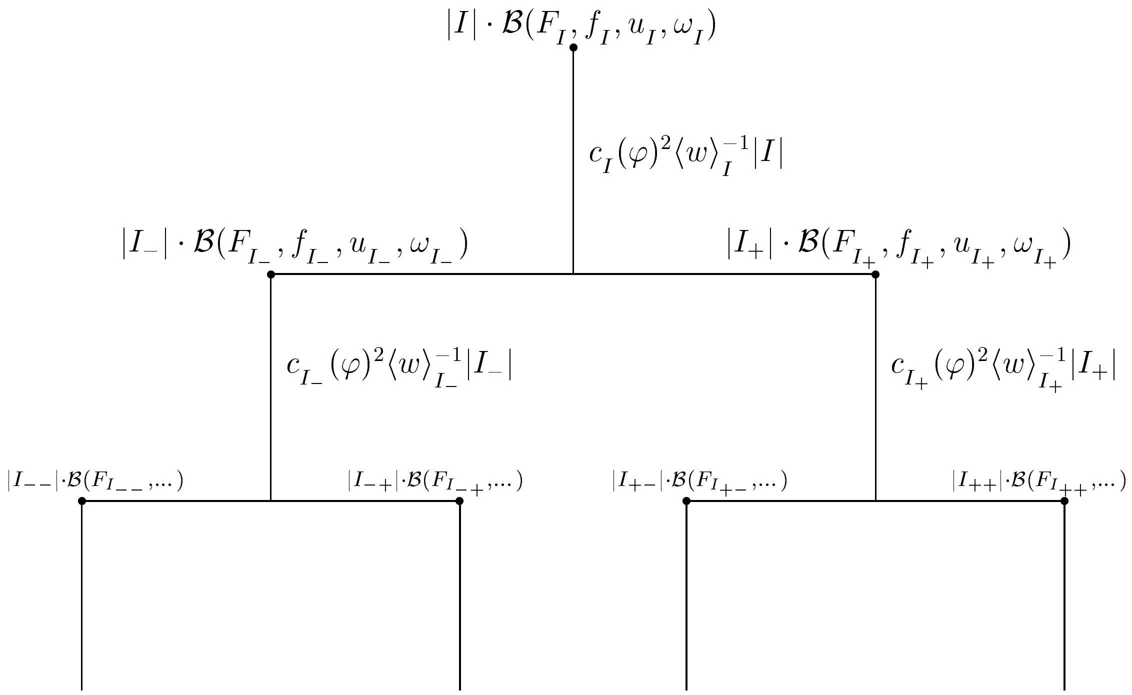

Figure 5: Imbedding theorem for Haar system. Vertices correspond to dyadic intervals and are denoted by black dots.

6.3. The hunt for a Bellman (type) function: an infinitesimal version of the key property

We will try to kill everything by one function.

We will construct a function $\mathcal{B}(F, f, u, \omega)$ defined in a whole region $D = \cup_C D(C) = \{(F, f, u, \omega) : f^2 \le Fu,\ e^\omega \le u\}$ (which is convex) satisfying

1) $0 \le \mathcal{B}(F, f, u, \omega) \le F$;

2) For any quadruplets (F, f, u, ω), $(F_\pm, f_\pm, u_\pm, \omega_\pm) \in D$ such that

$$(F, f, u, \omega) = \frac{1}{2}\{(F_+, f_+, u_+, \omega_+) + (F_-, f_-, u_-, \omega_-)\},$$

the following inequality holds:

$$\mathcal{B}(F, f, u, \omega) \ge \Xi(x)\frac{(f_+ - f_-)^2}{4u} + \frac{1}{2}\Big\{\mathcal{B}(F_+, f_+, u_+, \omega_+) + \mathcal{B}(F_-, f_-, u_-, \omega_-)\Big\};$$

here $x = ue^{-\omega}$, and Ξ is some positive decreasing function on $[1, \infty)$.

If we have such "universal" Bellman function, we can just put $\mathcal{B}_C := \Xi(C)^{-1}\mathcal{B}$ to get an "individual" one.

Again it seems to be no pleasure to check the key property in finite differences, and we would like to switch to its infinitesimal version

2′) $d^2\mathcal{B} \le -2\Xi(x)\dfrac{(df)^2}{u}$.

Let us check that the infinitesimal version 2′) of the key property implies 2) (with a little worse Ξ).

Indeed, assume that 2′) holds everywhere in D. Let (F, f, u, ω), $(F_\pm, f_\pm, u_\pm, \omega_\pm) \in D$ satisfy

$$(F, f, u, \omega) = \frac{1}{2}\{(F_+, f_+, u_+, \omega_+) + (F_-, f_-, u_-, \omega_-)\}.$$

Put

$$(F_\tau, f_\tau, u_\tau, \omega_\tau) = \frac{1+\tau}{2}(F_+, f_+, u_+, \omega_+) + \frac{1-\tau}{2}(F_-, f_-, u_-, \omega_-).$$

Since D is convex, $(F_\tau, f_\tau, u_\tau, \omega_\tau) \in D$ for every $\tau \in [-1, 1]$. Put $\Phi(\tau) = \mathcal{B}(F_\tau, f_\tau, u_\tau, \omega_\tau)$. Then the Taylor formula gives

$$\mathcal{B}(F, f, u, \omega) - \frac{1}{2}\Big\{\mathcal{B}(F_+, f_+, u_+, \omega_+) + \mathcal{B}(F_-, f_-, u_-, \omega_-)\Big\} = \Phi(0) - \frac{1}{2}\big(\Phi(-1) + \Phi(1)\big) =$$
$$-\frac{1}{2}\int_{-1}^{1}(1 - |\tau|)\Phi''(\tau)d\tau \ge \left(\int_{-1}^{1}(1 - |\tau|)\Xi(x_\tau)\frac{d\tau}{u_\tau}\right)\left(\frac{f_+ - f_-}{2}\right)^2, \quad (6.2)$$

where, as usual, $x_\tau = u_\tau \exp(-\omega_\tau)$.

Since $u_\tau \ge 0$ for every $\tau \in [-1,1]$ and $u_\tau + u_{-\tau} = 2u_0 = 2u$, we conclude that $u_\tau \le 2u$ for every τ and therefore

$$\mathcal{B}(F,\dots) - \frac{1}{2}\Big\{\mathcal{B}(F_+,\dots) + \mathcal{B}(F_-,\dots)\Big\} \ge \Big(\int_{-1}^{1}(1-|\tau|)\Xi(x_\tau)d\tau\Big)\,\frac{1}{2u}\Big(\frac{f_+-f_-}{2}\Big)^2 .$$

So it remains only to estimate the integral from below. The key point is that the inequality $x = ue^{-\omega} \ge C$ determines a convex region for every $C > 0$. Indeed we have

$$\frac{u'+u''}{2}\exp\big(-\frac{\omega'+\omega''}{2}\big) \ge \sqrt{(u'e^{-\omega'})(u''e^{-\omega''})} \ge C$$

if both $u'e^{-\omega'} \ge C$ and $u''e^{-\omega''} \ge C$.

Therefore on at least one of the intervals $[-1,0]$ and $[0,1]$ the quantity x_τ never exceeds $x_0 = x$. Assuming for the definiteness that $x_\tau \le x$ for every $\tau \in [0,1]$ and recalling that Ξ is decreasing, we get

$$\int_{-1}^{1}(1-|\tau|)\Xi(x_\tau)d\tau \ge \int_{0}^{1}(1-|\tau|)\Xi(x)d\tau = \frac{1}{2}\Xi(x).$$

This means that $2'$) implies 2) with the factor $\Xi(x)/4$ instead of $\Xi(x)$.

6.3.1. Remark

Of course, *a priori* we have no guarantee that the universal Bellman function exists. We only know that (if the theorem is true) for any region $D(C)$ an *individual* Bellman function satisfying the key property 2) (in finite difference form) exists.

The question arises whether for such individual Bellman function an infinitesimal version of the key property implies the finite difference one. Since the domain $D(C)$ is *not convex* the above reasoning for the universal Bellman function formally fails.

Nevertheless, let us note that if all three quadruplets (F, f, u, w), $(F_\pm, f_\pm, u_\pm, w_\pm) \in D(C)$, then $(F_\tau, f_\tau, u_\tau, w_\tau) \in D(C^2)$. Indeed, all we need is to check that if $u_\tau e^{-\omega_\tau} \le C$ for $\tau = 0, 1, -1$, then $u_\tau e^{-\omega_\tau} \le C^2$ for all $\tau \in [-1,1]$.

Define $g(\tau) = \log u_\tau - \omega_\tau$. Note that $g(\tau)$ is concave ($g'' \le 0$) non-negative function on $[-1,1]$ and $g(0) \le \log C$. Therefore, the whole graph of g lies within the “butterfly” on Figure 6, and we get $g(\tau) \le 2g(0) \le 2\log C$, which means $u_\tau e^{-\omega_\tau} \le C^2$.

So, if we find a Bellman function satisfying the infinitesimal version of the key property in $D(C^2)$, we can “run the machine backward” in $D(C)$. If we are not after the best constant, we lose nothing.

6.4. The hunt for a Bellman (type) function: how to find it

We will seek our Bellman type function $\mathcal{B}$ satisfying 1–2 above in the form

$$\mathcal{B}(F, f, u, \omega) = F - \frac{f^2}{u}\xi(ue^{-\omega}) \tag{6.3}$$

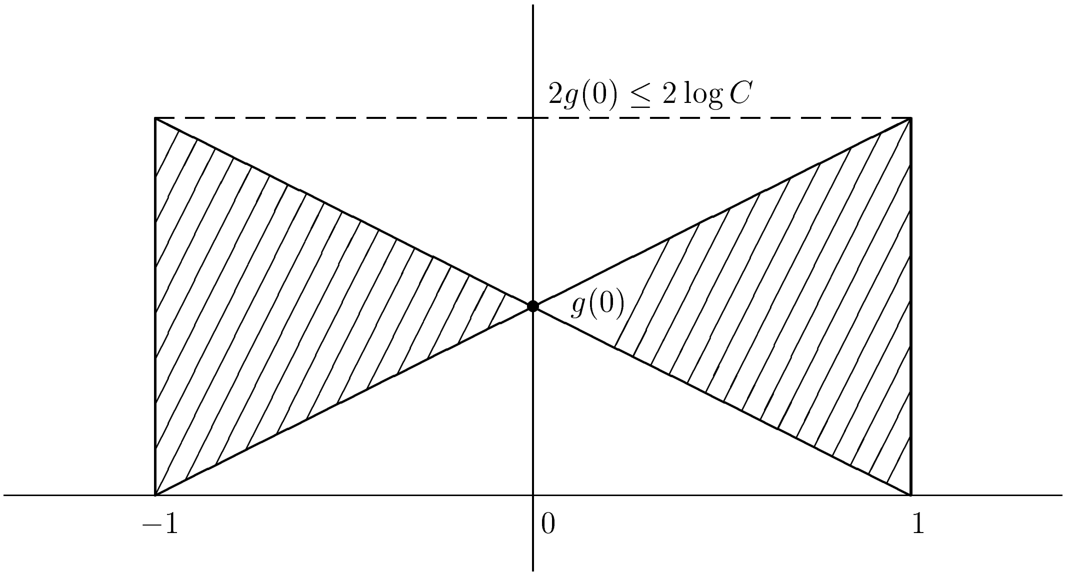

Figure 6: The graph of $g(\tau) = \log u_\tau - \omega_\tau$ has to be inside the "butterfly".

where $\xi : [1, \infty) \to [0, 1]$.

The guess is quite natural, because when looking for a function $\mathcal{B}$ satisfying $d^2\mathcal{B} \leq -2\Xi(x)(df)^2/u$, one is tempted to take just $\mathcal{B} = -f^2/u$ (the second differential of which just begins with the desired term $-2(df)^2/u$). This function has two drawbacks: first, it is not positive, and second,

$$d^2\Big(-\frac{f^2}{u}\Big) = -2\frac{f^2}{u}\Big(\frac{df}{f} - \frac{du}{u}\Big)^2$$

degenerates.

The first drawback is easy to fix: we may just consider $\mathcal{B} = F - f^2u^{-1}$ instead (recall that $f^2 \leq uF$ in the domain of $\mathcal{B}$). As for the second drawback, we may hope that changing the term f^2u^{-1} to $(f^2u^{-1})\xi(ue^{-\omega})$, we will be able to add some convexity into it. The idea that ξ should depend on $x = ue^{-\omega}$ comes from the demand that Ξ should depend on x only.

Thus we need to find a bounded non-negative function ξ such that

$$d^2\Big(\frac{f^2}{u}\xi(ue^{-\omega})\Big) \geq 2\Xi(x)\frac{(df)^2}{u}\,.$$

The direct computation gives

$$d^2\Big(\frac{f^2}{u}\xi(ue^{-\omega})\Big) =$$

$$= \frac{f^2}{u}\cdot\begin{pmatrix} \frac{df}{f} \\ \frac{du}{u} \\ d\omega \end{pmatrix}^T \begin{pmatrix} 2\xi & -2\xi+2x\xi' & -2x\xi' \\ -2\xi+2x\xi' & 2\xi-2x\xi'+x^2\xi'' & -x^2\xi'' \\ -2x\xi' & -x^2\xi'' & x^2\xi''+x\xi' \end{pmatrix}\begin{pmatrix} \frac{df}{f} \\ \frac{du}{u} \\ d\omega \end{pmatrix} =$$

$$= \frac{f^2}{u}\cdot\begin{pmatrix} \frac{df}{f}-\frac{du}{u} \\ \frac{du}{u} \\ d\omega \end{pmatrix}^T \begin{pmatrix} 2\xi & 2x\xi' & -2x\xi' \\ 2x\xi' & 2x\xi'+x^2\xi'' & -x^2\xi''-2x\xi' \\ -2x\xi' & -x^2\xi''-2x\xi' & x^2\xi''+x\xi' \end{pmatrix}\begin{pmatrix} \frac{df}{f}-\frac{du}{u} \\ \frac{du}{u} \\ d\omega \end{pmatrix} =$$

$$= \frac{f^2}{u}\cdot\begin{pmatrix} \frac{df}{f}-\frac{du}{u} \\ \frac{du}{u}-d\omega \\ d\omega \end{pmatrix}^T \begin{pmatrix} 2\xi & 2x\xi' & 0 \\ 2x\xi' & 2x\xi'+x^2\xi'' & 0 \\ 0 & 0 & -x\xi' \end{pmatrix}\begin{pmatrix} \frac{df}{f}-\frac{du}{u} \\ \frac{du}{u}-d\omega \\ d\omega \end{pmatrix}$$

If our task were just to find a function ξ on a finite interval $[1, C]$, then to make the second differential strictly positive, it would suffice to satisfy the inequalities $\xi'(x) < 0$ and $2x\xi' + x^2\xi'' > 0$. Then, adding a large constant to ξ (which raises the upper left corner of the matrix but does not change other entries) we would get the strict convexity desired. Of course, then we always can renormalize ξ to make it bounded above by 1.

The solution of the equation $2x\xi' + x^2\xi'' = 0$ is (up to a multiplicative constant) $\xi(x) = A + x^{-1}$. So if we take $\xi(x) = A + x^{-\alpha}$, $\alpha > 1$, then $2x\xi' + x^2\xi'' > 0$.

The differential $d^2\Big(\frac{f^2}{u}\xi(ue^{-\omega})\Big)$ for this choice of ξ is

$$\frac{f^2}{u}\cdot\begin{pmatrix} \frac{df}{f}-\frac{du}{u} \\ \frac{du}{u}-d\omega \\ d\omega \end{pmatrix}^T \begin{pmatrix} 2A+2x^{-\alpha} & -2\alpha x^{-\alpha} & 0 \\ -2\alpha x^{-\alpha} & \alpha(\alpha-1)x^{-\alpha} & 0 \\ 0 & 0 & \alpha x^{-\alpha} \end{pmatrix}\begin{pmatrix} \frac{df}{f}-\frac{du}{u} \\ \frac{du}{u}-d\omega \\ d\omega \end{pmatrix}$$

If $A \geq 4\alpha/(\alpha-1)$, then

$$2A\frac{\alpha(\alpha-1)}{2}x^{-\alpha} \geq (2\alpha)^2x^{-\alpha} \geq (2\alpha x^{-\alpha})^2 \qquad \text{for every } x \geq 1\,,$$

and therefore

$$\begin{pmatrix} 2A+2x^{-\alpha} & -2\alpha x^{-\alpha} & 0 \\ -2\alpha x^{-\alpha} & \alpha(\alpha-1)x^{-\alpha} & 0 \\ 0 & 0 & \alpha x^{-\alpha} \end{pmatrix} \geq \begin{pmatrix} 2x^{-\alpha} & 0 & 0 \\ 0 & \frac{1}{2}\alpha(\alpha-1)x^{-\alpha} & 0 \\ 0 & 0 & \alpha x^{-\alpha} \end{pmatrix} > 0\,.$$

It follows immediately from here that

$$\begin{aligned} d^2\Big(\frac{f^2}{u}\xi(ue^{-\omega})\Big) &\geq \\ &\geq x^{-\alpha}\frac{f^2}{u}\Big[2\Big(\frac{df}{f}-\frac{du}{u}\Big)^2+\frac{\alpha(\alpha-1)}{2}\Big(\frac{du}{u}-d\omega\Big)^2+\alpha(d\omega)^2\Big] \geq \\ \geq \Big(\frac{1}{2}+\frac{2}{\alpha(\alpha-1)}+\frac{1}{\alpha}\Big)^{-1} & x^{-\alpha}\frac{f^2}{u}\Big[\Big(\frac{df}{f}-\frac{du}{u}\Big)+\Big(\frac{du}{u}-d\omega\Big)+(d\omega)\Big]^2 = \gamma x^{-\alpha}\frac{(df)^2}{u}\,, \end{aligned}$$

i.e. we got what we wanted.

Just for the technical convenience we will renormalize $\xi(x)$ to $(1+\delta x^{-\alpha})/2$, where $\alpha>1$, $\delta\leq(\alpha-1)(4\alpha)^{-1}$. The Bellman type function we obtain is

$$\mathcal{B}(F,f,u,\omega)=F-\frac{f^2}{2u}\Big(1+\delta\big(ue^{-\omega}\big)^{-\alpha}\Big)\,,$$

with $\Xi(x)=\gamma' x^{-\alpha}$, where $\gamma'=\gamma\delta/4$ for the infinitesimal version, or $\gamma'=\gamma\delta/16$ for the finite difference one.

The expression for $\mathcal{B}$ may look a bit cumbersome, but the reader should take into account that this single line constitutes the *complete* proof of the imbedding theorem (up to setting the notation and *checking* the strict convexity property). As the imbedding theorem implies that the Haar system is an unconditional basis in $L^2(w)$ if $w\in(A_2^{\mathrm{dy}})$, and this result is just one step from the proof of the famous Hunt–Muckenhoupt–Wheeden theorem, the expression cannot be *too simple*.

6.5. Vector-valued problem with scalar weight

If we substitute $\|\mathbf{f}\|_E$ ($\mathbf{f}\in E$) instead of f in the expression for $\mathcal{B}(F,f,u,\omega)$, we will get a Bellman function corresponding to the *vector-valued problem with scalar weight* (i.e. to the case of E-valued L^2 with $W(t)=w(t)Id$, where w is a positive scalar function).

Indeed, let us compute the second differential of $\mathcal{B}(F,\|\mathbf{f}\|_E,u,\omega)$ in $\mathbb{R}\times E\times\mathbb{R}\times\mathbb{R}$. We have

$$\begin{aligned} d^2\Big[\mathcal{B}(F,\|\mathbf{f}\|_E,u,\omega)\Big] = (d^2\mathcal{B})\,\Big|_{F,\|\mathbf{f}\|_E,u,\omega}(dF,d\|\mathbf{f}\|_E,du,d\omega)+\frac{\partial\mathcal{B}}{\partial f}d^2\|\mathbf{f}\|_E \leq \\ \leq -2\Xi(x)\frac{(d\|\mathbf{f}\|_E)^2}{u}-\frac{2\|\mathbf{f}\|_E}{u}d^2\|\mathbf{f}\|_E\,. \end{aligned}$$

The differentials $d\|\mathbf{f}\|_E$ and $d^2\|\mathbf{f}\|_E$ exist and are easy to compute when $\|\mathbf{f}\|_E \neq 0$:

$$d\|\mathbf{f}\|_E = (d\mathbf{f}, e)_{E,\mathbb{R}}\,;$$
$$d^2\|\mathbf{f}\|_E = \|\mathbf{f}\|_E^{-1}\|Qd\mathbf{f}\|_E^2\,,$$

where $e = \|\mathbf{f}\|_E^{-1}\mathbf{f}$, $(\cdot,\cdot)_{E,R} = \mathrm{Re}(\cdot,\cdot)_E$ is the scalar product in E considered as a Hilbert space over $\mathbb{R}$, and Q is the (real linear) orthogonal projection onto the orthogonal complement of the one-dimensional space $\mathbb{R}e = \mathbb{R}\mathbf{f}$,

$$Qx = x - (x, e)_{E,\mathbb{R}}\,e\,.$$

Since $(d\mathbf{f}, e)_{E,\mathbb{R}}^2 + \|Qd\mathbf{f}\|_E^2 = \|d\mathbf{f}\|_E^2$, we observe that

$$d^2\Big[\mathcal{B}(F, \|\mathbf{f}\|_E, u, \omega)\Big] \le -2\min\{\Xi(x), 1\}\frac{\|d\mathbf{f}\|_E^2}{u} = -2\Xi(x)\frac{\|d\mathbf{f}\|_E^2}{u}\,,$$

and we get the infinitesimal version of the key property at every point where $\mathbf{f} \neq 0$.

For $\mathbf{f} = 0$ the function $\mathcal{B}(F, \|\mathbf{f}\|_E, u, \omega)$ remains C^1-smooth (actually, even C^∞-smooth here, but it will not be so for $p \neq 2$). Therefore we can use the same reasoning as above to derive the key property in its finite difference form (we leave to the reader to justify that the Taylor formula we used can be applied to C^1-functions which have negative second derivative everywhere except finite number of exceptional points).

7. Imbedding theorem for the matrix L^2 case

In this section we will prove the inequality

$$\sum_{I\in\mathcal{D}}\big\|\langle W\rangle_I^{-1/2}c_I(\varphi)\big\|_E^2\cdot|I| \le C\cdot\int_{\mathbb{R}}\big\|W(t)^{-1/2}\varphi(t)\big\|_E^2 dt \tag{7.1}$$

(the imbedding 1 from Lemma 4.12) and thereby will show that the Haar system is a strong unconditional basis in $L^2(W)$, because to prove the second imbedding we merely have to replace W by W^{-1}.

Again, this imbedding follows from a condition weaker that $(\mathcal{A}_{2,2})$.

7.1. $(\mathcal{A}_{2,0})$ condition

The matrix analogue of the (A_∞) condition is the inequality

$$\det\langle W\rangle_I \le C\exp\{\langle\log\det W\rangle_I\} \qquad \text{for every interval } I \subset \mathbb{R}. \tag{$\mathcal{A}_{2,0}$}$$

The reasons for the name $(\mathcal{A}_{2,0})$ will be explained in Section 11.4. Note that if $\dim E = 1$, this condition exactly coincides with (A_∞). We will say that a weight W satisfies a dyadic $(\mathcal{A}_{2,0})$ condition (and write $W \in (\mathcal{A}_{2,0}^{\mathrm{dy}})$) if the above property holds for every *dyadic* interval I.

Let us show first of all that again, like it was in the scalar case,

$$W \in (\mathcal{A}_{2,2}) \iff W \in (\mathcal{A}_{2,0})\ \&\ W^{-1} \in (\mathcal{A}_{2,0})$$

(and the same for dyadic versions).

7.2. Jensen inequality for matrices

Recall that for a linear operator $W : E \to E$, $\dim E = n < \infty$ its trace $\operatorname{Tr} W$ is defined as $\sum_{j=1}^{n} (We_j, e_j)_E$ where $\{e_j\}_{j=1}^n$ is any orthonormal basis in E. The trace actually does not depend on the choice of the basis $\{e_j\}_{j=1}^n$, and can be also defined by $\operatorname{Tr} W = n \int_S (We, e) dm(e)$, where $S = \{e \in E : \|e\|_E = 1\}$ is the unit sphere in E and m is the Lebesgue measure on S normalized by $m(S) = 1$.

Another formula for trace is $\operatorname{Tr} W = \sum_{j=1}^n \lambda_j$, where $\{\lambda_j\}_{j=1}^n$ are the eigenvalues of W (counted with multiplicity).

Suppose now that $W \in \mathcal{L}_+(E)$ (i.e. $W = W^* > 0$). Since $\det W = \prod_{j=1}^n \lambda_j$ and all the eigenvalues $\lambda_j > 0$, we immediately get the inequality $n^{-1} \operatorname{Tr} W \geq (\det W)^{1/n}$.

Let A be a linear operator in E (not necessarily self-adjoint) such that $|\det A| = 1$. Then for every $W \in \mathcal{L}_+(E)$ the operator $A^*WA \in \mathcal{L}_+(E)$ as well, and we can conclude that

$$\frac{1}{n} \operatorname{Tr}(A^*WA) \geq \big(\det(A^*WA)\big)^{1/n} = \big(\det W\big)^{1/n}.$$

When $A = \big(\det W\big)^{1/(2n)} \cdot W^{-1/2}$, we have

$$A^*WA = \big(\det W\big)^{1/n} \cdot Id\,,$$

and the equality is attained. So we have proved the following lemma

Lemma 7.1 (Formula for determinant). *If $W \in \mathcal{L}_+(E)$ and $\dim E = n < \infty$, then*

$$\big(\det W\big)^{1/n} = \inf\Big\{\frac{1}{n} \operatorname{Tr}(A^*WA) \,:\, |\det A| = 1\Big\}.$$

Now we can easily prove the following

Lemma 7.2 (Jensen inequality for matrix functions). *For a matrix weight W*

$$\det\big\langle W\big\rangle_I \geq \exp\big\{\big\langle \log\det W\big\rangle_I\big\}.$$

Proof. By Lemma 7.1

$$\begin{aligned}\big(\det W\big)^{1/n} &= \inf_{A : |\det A| = 1} \frac{1}{n} \operatorname{Tr}\big(A^*\langle W\rangle_I A\big) = \inf_{A : |\det A| = 1} \frac{1}{n}\big\langle \operatorname{Tr} A^*WA\big\rangle_I \geq \\ &\geq \Big\langle \inf_{A : |\det A| = 1} \frac{1}{n} \operatorname{Tr} A^*WA\Big\rangle_I = \big\langle (\det W)^{1/n}\big\rangle_I \geq \exp\Big\{\frac{1}{n}\big\langle \log\det W\big\rangle_I\Big\}.\end{aligned}$$

Taking the nth power of this inequality, we get the conclusion of the lemma. □

The way we chose to prove the Jensen inequality is surely not the shortest one. Nevertheless it demonstrates one important idea we will frequently use when working with matrices: whenever any quantity is (or, at least, seems) not easy to handle (like the determinant of a sum of matrices in the above case), we will try to represent it as a solution of some extremal problem.

The main advantage of this trick is that it will allow us almost to forget about the greatest nightmare of the matrix case: non-commutativity of the multiplication. The main drawback is an essential loss of control over constants for large $n = \dim E$. Still, since we can not make the constants independent of n (i.e. prove the result for operator-valued weights), there seems to be no reason to care much about *how exactly we can not* do it.

Lemma 7.3. *Let W be a matrix weight. Then*

$$W \in (\mathcal{A}_{2,2}) \iff W \in (\mathcal{A}_{2,0}) \ \&\ W^{-1} \in (\mathcal{A}_{2,0})$$

(and the same for dyadic versions).

Proof. Since $\left\|\langle W^{-1}\rangle_I^{1/2}\langle W\rangle_I^{1/2}\right\| \le C$, we can conclude that

$$\begin{aligned} C^{2n} &\ge \det\langle W\rangle_I \det\langle W^{-1}\rangle_I = \\ &= \Big[\det\langle W\rangle_I \exp\Big\{-\langle \log\det W\rangle_I\Big\}\Big]\cdot\Big[\det\langle W^{-1}\rangle_I \exp\Big\{-\langle \log\det W^{-1}\rangle_I\Big\}\Big]. \end{aligned}$$

By the Jensen inequality expressions in brackets are at least 1, so $(\mathcal{A}_{2,2})$ for W implies $(\mathcal{A}_{2,0})$ for W and W^{-1}.

To get the inverse implication we need the following standard fact from linear algebra.

Lemma 7.4. *Let B be a bounded operator in a finite-dimensional Hilbert space E such that $\|B\| \le 1$ and $|\det B| \ge \delta > 0$. Then $\|B^{-1}\| \le 1/\delta$.*

Proof. Exercise for the reader. □

Corollary 7.5. *Let B be a bounded operator in a finite-dimensional Hilbert space E such that $\|Bx\|_E \ge \|x\|_E$ for all $x \in E$ and $|\det B| \le C < \infty$. Then $\|B\| \le C$.*

Proof. Apply the previous lemma to B^{-1}. □

Note that the condition $W, W^{-1} \in (\mathcal{A}_{2,0})$ implies that

$$\left|\det\left(\langle W^{-1}\rangle_I^{1/2}\langle W\rangle_I^{1/2}\right)\right| \le C'.$$

Recalling that $\left\|\langle W\rangle_I^{-1/2}x\right\|_E = \langle\rho\rangle_{I,2}^*(x)$, $\left\|\langle W^{-1}\rangle_I^{1/2}x\right\|_E = \langle\rho^*\rangle_{I,2}(x)$ and that always (see properties of $(\mathcal{A}_{p,q})$ weights in Section 2.5) $\langle\rho^*\rangle_{I,2}(x) \ge \langle\rho\rangle_{I,2}^*(x)$, we get for every $x \in E$

$$\left\|\langle W^{-1}\rangle_I^{1/2}x\right\|_E \ge \left\|\langle W\rangle_I^{-1/2}x\right\|_E.$$

Therefore for all $x \in E$

$$\left\|\langle W^{-1}\rangle_I^{1/2}\langle W\rangle_I^{1/2}x\right\|_E \ge \|x\|_E,$$

and applying the above Corollary to the operator $B = \langle W^{-1}\rangle_I^{1/2}\langle W\rangle_I^{1/2}$ we get the desired result. □

We will also need an analogue of the scalar Cauchy inequality $\langle\varphi\rangle_I^2 \le \langle\varphi^2 w^{-1}\rangle_I\langle w\rangle_I$.

Lemma 7.6 (Matrix Cauchy inequality). *For a matrix weight W*

$$\left\|\langle W\rangle_I^{-1/2}\langle\varphi\rangle_I\right\|_E \le \left\langle \left\|W^{-1/2}\varphi\right\|_E^2\right\rangle_I^{1/2}.$$

This is a particular case of the following "Hölder inequality " that holds for an arbitrary family of norms ρ_t.

Lemma 7.7 (Matrix Hölder inequality). *Let ρ_t, $t\in\mathbb{R}$ be a family of norms on E. Then*

$$\langle\rho\rangle_{I,p}^*(\langle\varphi\rangle_I) \le \left\langle \left[\rho_t^*(\varphi(t))\right]^q\right\rangle_I^{1/q}.$$

Proof. For any $e\in E$

$$\left|(\langle\varphi\rangle_I,e)_E\right| = \left|\left\langle(\varphi,e)_E\right\rangle_I\right| \le \left\langle\left|(\varphi,e)_E\right|\right\rangle_I \le \frac{1}{|I|}\int_I \rho_t^*(\varphi(t))\cdot\rho_t(e)dt \le$$
$$\le \left\{\frac{1}{|I|}\int_I\left[\rho_t^*(\varphi(t))\right]^q dt\right\}^{1/q}\cdot\left\{\frac{1}{|I|}\int_I\left[\rho_t(e)\right]^p dt\right\}^{1/p} = \left\{\frac{1}{|I|}\int_I\left[\rho_t^*(\varphi(t))\right]^q dt\right\}^{1/q}\cdot\langle\rho\rangle_{I,p}(e),$$

and the conclusion of the lemma follows immediately from the definition of the dual norm. □

7.3. Proof of the imbedding theorem

Now everything is ready to check the imbedding theorem. We are going to prove the following

Theorem 7.8. *If a matrix weight W satisfies the Muckenhoupt condition $(\mathcal{A}_{2,0})$, then the imbedding*

$$\sum_{I\in\mathcal{D}}\left\|\langle W\rangle_I^{-1/2}c_I(\varphi)\right\|_E^2\cdot|I| \le C\cdot\int_{\mathbb{R}}\left\|W(t)^{-1/2}\varphi(t)\right\|_E^2 dt$$

holds for all $\varphi\in L^2(W^{-1})$.

To prove the theorem we will again look for a Bellman function corresponding to the problem. The averages to deal with are now

$$U_I = \langle W\rangle_I \in \mathcal{L}_+(E)\,, \qquad \omega_I = \frac{1}{n}\left\langle\log\det W\right\rangle_I \in \mathbb{R},$$
$$f_I = \langle\varphi\rangle_I \in E\,, \qquad F_I = \left\langle\left\|W^{-1/2}\varphi\right\|_E^2\right\rangle_I \in \mathbb{R}_+\,.$$

(the normalizing factor $1/n$ in the definition of ω is added just for technical convenience).

Again we will try to find an universal Bellman function, i.e. the function $\mathcal{B}(F,f,U,\omega)$ defined in the whole region

$$\left\{(F,f,U,\omega)\in\mathbb{R}\times E\times\mathcal{L}_+(E)\times\mathbb{R}\,:\,\|U^{-1/2}f\|_E^2\le F, e^\omega\le(\det U)^{1/n}\right\}$$

and satisfying two properties:

1) (Range) $0 \le \mathcal{B}(F, f, U, \omega) \le F$;

2) (the key property)

$$\mathcal{B}(F, f, U, \omega) \ge \Xi(x) \left\| U^{-1/2} \frac{f_+ - f_-}{2} \right\|_E^2 + \frac{1}{2}\Big\{\mathcal{B}(F_-, f_-, U_-, \omega_-) + \mathcal{B}(F_+, f_+, U_+, \omega_+)\Big\}$$

($x = (\det U)^{1/n} e^{-\omega}$) whenever all three quadruplets (F, f, U, ω), $(F_\pm, f_\pm, U_\pm, \omega_\pm)$ belong to the domain and satisfy

$$(F, f, U, \omega) = \frac{1}{2}\Big\{(F_-, f_-, U_-, \omega_-) + (F_+, f_+, U_+, \omega_+)\Big\}.$$

How to find the Bellman function? Note that if one has it for the matrix case we are considering, he can immediately get the Bellman function for the scalar problem as a particular case corresponding to $\dim E = 1$. Since we have already found the latter, it seems reasonable to use induction instead of deduction and just to try to plug matrix and vector arguments into our "scalar" expression. We have already seen (see Remark in Section 6.5) that there is no problem with replacing f by $\|f\|_E$. Going all the way in this direction, we get something like

$$\mathcal{B}(F, f, U, \omega) = F - \frac{1}{2}\|U^{-1/2} f\|_E^2 \Big(1 + \delta\big[(\det U)^{1/n} e^{-\omega}\big]^{-\alpha}\Big).$$

Is this a Bellman type function we need? Probably it is, but we do not feel at ease with differentiating functions of matrix arguments (try to find the second differential of $\det U$ or even the *first* one of $U^{-1/2}$ and you will understand why). So, we are going to use our idea to *get the quantities which seem hard to handle directly as solutions of extremal problems.* For that we should decide first of all what kind of invariance we have in our "imbedding theorem". Note that if we pick up an invertible linear operator $A : E \to E$ and replace $\varphi(t)$ by $A^*\varphi(t)$ and $W(t)$ by $A^*W(t)A$, the inequality to prove will not change. The averages F, f, U, ω will change correspondently to F, A^*f, A^*UA, $\omega + 2\log|\det A|$. The identity

$$n(\det U)^{1/n} = \inf_{\det A \ne 0} \frac{\operatorname{Tr}(A^*UA)}{|\det A|^{2/n}}$$

proved above in a slightly different form shows that we may avoid explicit differentiation of the determinant $\det U$ if we replace $(\det U)^{1/n}$ in our conjectured expression for the Bellman function by $\operatorname{Tr}(A^*UA)|\det A|^{-2/n}$ (note that unlike the determinant $\det U$ the trace $\operatorname{Tr}(A^*UA)$ is very easy to handle: it is just a *linear* function of $U \in \mathcal{L}_+(E)$).

Also, though in principle it is possible to differentiate $\|U^{-1/2}f\|_E^2$ (if we write it as $(U^{-1}f, f)_E$ first), we do not want to deal with this term as well: too much trouble with non-commutative multiplication which results in weird expressions like $d(W^{-1}) = W^{-1}(dW)W^{-1}$ instead of "normal" $\frac{dW}{W^2}$. What extremal problem could be posed here? Since we always have

$$\|U^{-1/2} f\|_E^2 \ge \frac{\|f\|_E^2}{\|U^{1/2}\|^2} = \frac{\|f\|_E^2}{\|U\|},$$

it seems reasonable to try $\sup\limits_{\det A\neq 0}\dfrac{\|A^*f\|_E^2}{\|A^*UA\|}$. Changing notation A to $U^{-1/2}A$, we get

$$\sup_{\det A\neq 0}\frac{\|A^*f\|_E^2}{\|A^*UA\|}=\sup_{\det A\neq 0}\frac{\|A^*U^{-1/2}f\|_E^2}{\|A^*A\|}=\sup_{\det A\neq 0}\frac{\|A^*U^{-1/2}f\|_E^2}{\|A^*\|^2}=\|U^{-1/2}f\|_E^2\ ,$$

so we really have what we want.

The operator norm $\|A^*UA\|$ in the denominator is a bad thing: first, it is unpleasant to handle, but what is the most important, it is a *convex* function of U, which would be good if we had had it in the numerator (then it only helps the Bellman function $\mathcal{B}$ to be concave), but which can completely destroy the key property when standing in the denominator. Since for any $U\in\mathcal{L}_+(E)$ we have

$$\|U\|=\max_{1\le j\le n}\lambda_j\le \operatorname{Tr}U=\sum_{1\le j\le n}\lambda_j\le n\|U\|=n\max_{1\le j\le n}\lambda_j$$

(as before λ_j ($j=1,\dots,n$) are the eigenvalues of U counted with multiplicity), we may try to replace $\|A^*UA\|$ by $\operatorname{Tr}(A^*UA)$ as well and thereby we come to the expression

$$\mathcal{B}_A(F,f,U,\omega)=\mathcal{B}_{\text{scalar}}(F,\|A^*f\|_E,\operatorname{Tr}(A^*UA),\omega+\frac{2}{n}\log|\det A|)=F-\frac{\|A^*f\|_E^2}{2\operatorname{Tr}(A^*UA)}(1+\delta x_A^{-\alpha})$$

where $x_A=\dfrac{\operatorname{Tr}(A^*UA)}{e^\omega|\det A|^{2/n}}$. Note that, as we have already seen above,

$$\frac{\|A^*f\|_E^2}{\operatorname{Tr}(A^*UA)}\le\frac{\|A^*f\|_E^2}{\|A^*UA\|}\le\|U^{-1/2}f\|_E^2\le F\ ,$$

and

$$x_A\ge n\frac{\{\det(A^*UA)\}^{1/n}}{e^\omega|\det A|^{2/n}}=n\frac{(\det U)^{1/n}}{e^\omega}\ge n\ge 1.$$

Thus for every quadruplet (F,f,U,ω) in the domain the corresponding quadruplet

$$(F,\|A^*f\|_E,\operatorname{Tr}(A^*UA),\omega+\frac{2}{n}\log|\det A|)$$

lies in the domain of $\mathcal{B}_{\text{scalar}}$ and from the key property of $\mathcal{B}_{\text{scalar}}$ and Remark from Section 6.5 it follows immediately that

$$\begin{aligned}&\mathcal{B}_A(F,f,U,\omega)\\&\ge\Xi(x_A)\,[\operatorname{Tr}(A^*UA)]^{-1}\left\|A^*\frac{f_+-f_-}{2}\right\|_E^2+\frac{1}{2}\Big\{\mathcal{B}_A(F_-,f_-,U_-,\omega_-)+\mathcal{B}_A(F_+,f_+,U_+,\omega_+)\Big\}\ .\end{aligned}$$

Let now $\mathcal{B}(F,f,U,\omega) := \inf_{\det A\neq 0} \mathcal{B}_A(F,f,U,\omega)$. Then, if for a quadruplet (F,f,U,ω) the infimum is attained at some operator $A_{\rm opt}$ ("opt" is just an abbreviation of "optimal", of course), we get

$$\mathcal{B}(F,f,U,\omega) = \mathcal{B}_{A_{\rm opt}}(F,f,U,\omega)$$
$$\geq \Xi(x_{A_{\rm opt}})\left[\mathrm{Tr}(A^*_{\rm opt}UA_{\rm opt})\right]^{-1}\left\|A^*_{\rm opt}\frac{f_+-f_-}{2}\right\|^2_E + \frac{1}{2}\Big\{\mathcal{B}_{A_{\rm opt}}(F_-,\dots)+\mathcal{B}_{A_{\rm opt}}(F_+,\dots)\Big\} \geq$$
$$\Xi(x_{A_{\rm opt}})\left[\mathrm{Tr}(A^*_{\rm opt}UA_{\rm opt})\right]^{-1}\left\|A^*_{\rm opt}\frac{f_+-f_-}{2}\right\|^2_E + \frac{1}{2}\Big\{\mathcal{B}(F_-,\dots)+\mathcal{B}(F_+,\dots)\Big\}\,.$$

If we contrive to show that $A_{\rm opt}$ exists and doesn't differ much from $U^{-1/2}$, then we will have

$$x_{A_{\rm opt}} \approx \frac{\mathrm{Tr}\big[(U^{-1/2})U(U^{-1/2})\big]}{e^\omega\big(\det(U^{-1/2})\big)^{2/n}} = n(\det U)^{1/n}e^{-\omega} = nx$$

and

$$\left[\mathrm{Tr}(A^*_{\rm opt}UA_{\rm opt})\right]^{-1}\left\|A^*_{\rm opt}\frac{f_+-f_-}{2}\right\|^2_E \approx \frac{1}{n}\left\|U^{-1/2}\frac{f_+-f_-}{2}\right\|^2_E\,,$$

i.e. we shall get the key property desired (or, at least, something close to it).

Forgetting about the irrelevant (i.e. not depending on A) term F and the non-essential constant $1/2$, we see that our task is to show that the solution of the extremal problem

$$\frac{\|A^*f\|^2_E}{\mathrm{Tr}(A^*UA)}\left(1+\delta\Big\{\frac{e^\omega|\det A|^{2/n}}{\mathrm{Tr}(A^*UA)}\Big\}^\alpha\right) \longrightarrow \max$$

in the class of all invertible linear operators $A : E \to E$ exists and is close to $U^{-1/2}$.

Representing $A = U^{-1/2}B$, we get the extremal problem

$$\Phi(B) := \frac{\|B^*U^{-1/2}f\|^2_E}{\mathrm{Tr}(B^*B)}\left(1+\delta\Big\{\frac{(\det U)^{1/n}}{e^\omega}\Big\}^{-\alpha}\Big\{\frac{|\det B|^{2/n}}{\mathrm{Tr}(B^*B)}\Big\}^\alpha\right) \longrightarrow \max$$

to investigate (now we need to show that the solution $B_{\rm opt}$ is close to Id).

Recalling our standard notation $x = (\det U)^{1/n}e^{-\omega}$ and denoting $U^{-1/2}f$ by $\widetilde{f}$, say, we get

$$\Phi(B) = \frac{\|B^*\widetilde{f}\|^2_E}{\mathrm{Tr}(B^*B)}\left(1+\delta x^{-\alpha}\Big\{\frac{|\det B|^{2/n}}{\mathrm{Tr}(B^*B)}\Big\}^\alpha\right)$$

to maximize.

Note that multiplication of the operator B by any scalar constant does not change $\Phi(B)$ and therefore we may look for the solution in the class of operators B satisfying $\|B\| = 1$. If we forget about the invertibility demand, we will have a maximum problem for a continuous function on a compact set and the solution $B_{\rm opt}$ will exist for sure (though, maybe, degenerated).

Note that the maximal value $\|\widetilde{f}\|^2_E$ of the ratio $\dfrac{\|B^*\widetilde{f}\|^2_E}{\operatorname{Tr}(B^*B)}$ is attained at the orthogonal projection B_0 onto the direction of the vector $\widetilde{f}\in E$ (i.e. on the (degenerated if $n>1$) operator $B_0x=\|\widetilde{f}\|^{-2}(x,\widetilde{f})_E\widetilde{f}$).

The reader may ask at this point "what if $\widetilde{f}=0$?" The answer is that in this case $\Phi(B)\equiv 0$ and the "maximum" value 0 is attained at *any* operator B_{opt}. In particular, we may take $B_{\text{opt}}=Id$ and we are done.

If $\widetilde{f}\neq 0$ (i.e. if we have a nontrivial extremal problem to solve), consider the family of operators $B_\varepsilon:=(1-\varepsilon)B_0+\varepsilon Id$ $(0\le\varepsilon\le 1)$. Note that

$$B_\varepsilon^*\widetilde{f}=\widetilde{f},\qquad \operatorname{Tr}(B_\varepsilon^*B_\varepsilon)=1+(n-1)\varepsilon^2,\qquad \det B_\varepsilon=\varepsilon^{n-1},$$

and, at last,

$$\Phi(B_\varepsilon)=\frac{\|\widetilde{f}\|^2_E}{1+(n-1)\varepsilon^2}\Big(1+\delta x^{-\alpha}\frac{\varepsilon^{2\alpha(n-1)/n}}{1+(n-1)\varepsilon^2}\Big).$$

The key point is that if $2\alpha(n-1)/n<2$, i.e. if $\alpha<n/(n-1)$ (recall that the only restriction on α we had so far was $\alpha>1$), then the factor

$$\frac{1}{1+(n-1)\varepsilon^2}\Big(1+\delta x^{-\alpha}\frac{\varepsilon^{2\alpha(n-1)/n}}{1+(n-1)\varepsilon^2}\Big)$$

rises above 1 for some $\varepsilon=\varepsilon(x)$. Let us denote the value of this factor at $\varepsilon(x)$ by $1+\delta x^{-\alpha}\eta(x)^\alpha$ (note that $\eta(x)>0$ and depends on x only). Then for the solution B_{opt} of the extremal problem $\|B\|=1,\quad \Phi(B)\to\max$ we should have $\Phi(B_{\text{opt}})\ge\Phi(B_{\varepsilon(x)})$ and thereby

$$\frac{|\det B_{\text{opt}}|^{2/n}}{\operatorname{Tr}(B^*_{\text{opt}}B_{\text{opt}})}\ge\eta(x).$$

Since $\operatorname{Tr}(B^*_{\text{opt}}B_{\text{opt}})\ge\|B_{\text{opt}}\|^2=1$, this means that $|\det B_{\text{opt}}|\ge\eta(x)^{n/2}$ and, according to Lemma 7.4, $\|B^{-1}_{\text{opt}}\|\le\eta(x)^{-n/2}$. So, B_{opt} is really close to identity, which allows to replace the heuristic "approximate equalities"

$$x_{A_{\text{opt}}}\approx nx\qquad\text{and}\qquad\big[\operatorname{Tr}(A^*_{\text{opt}}UA_{\text{opt}})\big]^{-1}\Big\|A^*_{\text{opt}}\frac{f_+-f_-}{2}\Big\|^2_E\approx\frac1n\Big\|U^{-1/2}\frac{f_+-f_-}{2}\Big\|^2_E$$

above by the rigorous inequalities

$$x_{A_{\text{opt}}}=\frac{\operatorname{Tr}(B^*_{\text{opt}}B_{\text{opt}})}{e^\omega\big|\det(U^{-1/2}B_{\text{opt}})\big|^{2/n}}=\frac{\operatorname{Tr}(B^*_{\text{opt}}B_{\text{opt}})}{|\det B_{\text{opt}}|^{2/n}}(\det U)^{1/n}e^{-\omega}\le\eta(x)^{-1}x$$

and

$$\begin{aligned}\big[\operatorname{Tr}(A^*_{\text{opt}}UA_{\text{opt}})\big]^{-1}\Big\|A^*_{\text{opt}}\frac{f_+-f_-}{2}\Big\|^2_E&=\big[\operatorname{Tr}(B^*_{\text{opt}}B_{\text{opt}})\big]^{-1}\Big\|B^*_{\text{opt}}U^{-1/2}\frac{f_+-f_-}{2}\Big\|^2_E\\ &\ge\frac{1}{n\|B_{\text{opt}}\|^2}\frac{1}{\|B^{-1}_{\text{opt}}\|^2}\Big\|U^{-1/2}\frac{f_+-f_-}{2}\Big\|^2_E\ge n^{-1}\eta(x)^n\Big\|U^{-1/2}\frac{f_+-f_-}{2}\Big\|^2_E,\end{aligned}$$

proving the key property with the factor $\Xi'(x)=n^{-1}\eta(x)^n\Xi(\eta(x)^{-1}x)$ instead of $\Xi(x)$.

8. The Haar system as an unconditional basis in scalar non-weighted L^p.

The reader who has been able to remain with us up to this point may be proud of himself: his training is now over and we invite him to a real hunt. The beast to catch is the Bellman function responsible for the Haar system's being a basis in $L^p(W)$ for any $p \in (1, +\infty)$. Note that for $p \neq 2$ even the scalar problem without any weight is nontrivial. So, let us start our hunt with this simplest case.

We are going to use Criterion 1 (Lemma 4.8) now. Again, the Parseval identity holds for every $\varphi \in L^2 \cap L^p$, $\psi \in L^2 \cap L^q$ $\quad (1/p + 1/q = 1)$ and therefore it is enough to check the Bessel inequality

$$\sum_{I \in \mathcal{D}} \|\mathcal{P}_I \varphi\|_{L^p} \|\mathcal{P}_I \psi\|_{L^q} \leq C \|\varphi\|_{L^p} \|\psi\|_{L^q}.$$

Note that $\|\mathcal{P}_I \varphi\|_{L^p} = |I|^{1/p} |c_I(\varphi)|$, $\|\mathcal{P}_I \psi\|_{L^q} = |I|^{1/q} |c_I(\psi)|$ and therefore we need to check that

$$\sum_{I \in \mathcal{D}} |c_I(\varphi)||c_I(\psi)||I| \leq C \|\varphi\|_{L^p} \|\psi\|_{L^q}$$

for every $\varphi \in L^p$, $\psi \in L^q$. Again it is enough to prove this inequality for real-valued functions φ and ψ.

Fix the averages

$$F_I := \langle |\varphi|^p \rangle_I, \quad G_I := \langle |\psi|^q \rangle_I, \quad f_I := \langle \varphi \rangle_I, \quad g_I := \langle \psi \rangle_I$$

and define the Bellman function $\mathcal{B}(F, G, f, g)$ corresponding to the problem as

$$\frac{1}{|I|} \sup \Big\{ \sum_{J \subset I} |c_J(\varphi)||c_J(\psi)||J| \,:\, \langle |\varphi|^p \rangle_I = F,\ \langle |\psi|^q \rangle_I = G,\ \langle \varphi \rangle_I = f,\ \langle \psi \rangle_I = g \ \Big\}.$$

The properties of the Bellman function $\mathcal{B}(F, G, f, g)$ are easy to write down:

1) Domain: $|f|^p \leq F$, $|g|^q \leq G$;

2) Range: $0 \leq \mathcal{B}(F, G, f, g) \leq C F^{1/p} G^{1/q}$;

3) Key property:

$$\mathcal{B}(F, G, f, g) - \frac{1}{2}\big\{\mathcal{B}(F_-, G_-, f_-, g_-) + \mathcal{B}(F_+, G_+, f_+, g_+)\big\} \geq \left|\frac{f_+ - f_-}{2}\right| \cdot \left|\frac{g_+ - g_-}{2}\right|$$

whenever all three quadruplets (F, G, f, g), $(F_\pm, G_\pm, f_\pm, g_\pm)$ lie in the domain and

$$(F, G, f, g) = \frac{1}{2}\big\{(F_-, G_-, f_-, g_-) + (F_+, G_+, f_+, g_+)\big\}.$$

It is easy to check (using the Taylor formula) that 3) is equivalent to its infinitesimal version

3′) $d^2\mathcal{B} \le -2|df||dg|$.

How to find the Bellman (type) function $\mathcal{B}$? Seems, shouldn't be very hard: we may try to find the Bellman function for the simplest case $p = 2$ first and then, probably, it will remain only to guess the general expression from that for this particular case. Let us carry this program out and see whether it really works. For $p = 2$ the definition of the Bellman function gets

$$\mathcal{B}(F, G, f, g) = \frac{1}{|I|}\sup\Big\{\sum_{J\subset I}|c_J(\varphi)||c_J(\psi)||J| \,:\, \langle\varphi^2\rangle_I = F,\ \langle\psi^2\rangle_I = G,\ \langle\varphi\rangle_I = f,\ \langle\psi\rangle_I = g\Big\}.$$

Since replacement of any of the coefficients $c_J(\varphi)$ ($J \subset I$) by $-c_J(\varphi)$ changes neither $\langle\varphi^2\rangle_I$, nor $\langle\varphi\rangle_I$, we may omit the absolute value sign everywhere and get

$$\begin{aligned}\mathcal{B}(F, G, f, g) &= \frac{1}{|I|}\sup\Big\{\sum_{J\subset I}c_J(\varphi)c_J(\psi)|J| \,:\, \langle\varphi^2\rangle_I = F, \dots\Big\}\\ &= \sup\Big\{\big\langle(\varphi - \langle\varphi\rangle_I)(\psi - \langle\psi\rangle_I)\big\rangle_I \,:\, \langle\varphi^2\rangle_I = F, \dots\Big\}.\end{aligned}$$

Note that $\varphi - \langle\varphi\rangle_I$ is just an arbitrary function $\widetilde{\varphi}$ satisfying $\langle\widetilde{\varphi}\rangle_I = 0$ and $\langle\widetilde{\varphi}^2\rangle_I = F - f^2$. Analogously, $\psi - \langle\psi\rangle_I$ is an arbitrary function $\widetilde{\psi}$ satisfying $\langle\widetilde{\psi}\rangle_I = 0$ and $\langle\widetilde{\psi}^2\rangle_I = G - g^2$. Thus

$$\begin{aligned}\mathcal{B}(F, G, f, g) &= \sup\Big\{\langle\widetilde{\varphi}\widetilde{\psi}\rangle_I \,:\, \langle\widetilde{\varphi}^2\rangle_I = F - f^2,\ \langle\widetilde{\varphi}\rangle_I = 0,\ \langle\widetilde{\psi}^2\rangle_I = G - g^2,\ \langle\widetilde{\psi}^2\rangle_I = 0\Big\}\\ &= \big(F - f^2\big)^{1/2}\big(G - g^2\big)^{1/2}.\end{aligned}$$

This is what the general expression should reduce to when $p = q = 2$. So, what can it be? The most natural suggestion is

$$\mathcal{B}(F, G, f, g) = (F - |f|^p)^{1/p}(G - |g|^q)^{1/q}$$

(seems pretty reasonable at any rate...). But after looking for several minutes at this formula, one begins to understand that something is desperately wrong with it: if, say, $p > 2$, then we have $\dfrac{\partial^2\mathcal{B}}{\partial f^2} = 0$ whenever $f = 0$, and therefore the second differential of $\mathcal{B}$ has no chance to satisfy $d^2\mathcal{B} \le -2|df||dg|$ on that hyperplane. Maybe we were just a little bit hasty in our conclusions? We see now that both $\dfrac{\partial^2\mathcal{B}}{\partial f^2}$ and $\dfrac{\partial^2\mathcal{B}}{\partial g^2}$ should be negative everywhere, so, maybe, we merely need to write down an expression including f^2 and g^2 rather than $|f|^p$ and $|g|^q$, i.e. something like

$$\mathcal{B}(F, G, f, g) = \big(F^{2/p} - f^2\big)^{1/2}\big(G^{2/q} - g^2\big)^{1/2}.$$

Now everything is all right with the second derivatives with respect to f and g, but, to our great surprise, it turns out that the whole guy is *no longer concave*!

Several more attempts like this and, if the reader is not much cleverer (or much luckier) than we are, he will make an unpleasant discover:

The beast can hide !!!

We have a clear footprint (the *true(!)* Bellman function for $p = 2$) and still it doesn't tell us much about in what direction to move. Of course, we might consider more and more sophisticated conjectures: for instance, we might add two above formulae to see whether the concavity in the first is enough to kill the lack of concavity in the second (by the way, it *is not*). But we should never forget that to catch the Bellman function for the *scalar non-weighted* L^p is not the end of the story: we will then have also to guess what "1" in the formula comes from the weight, what the role of the $(\mathcal{A}_{p,q})$ condition is etc. If our expression for the simplest case turns out to be very complicated, we will risk to get completely lost. The moral is that we have to find a way to simplify the situation.

8.1. Hölder vs Young.

Let us return to the inequality

$$\sum_{I\in\mathcal{D}} |c_I(\varphi)||c_I(\psi)||I| \le C\|\varphi\|_{L^p}\|\psi\|_{L^q}$$

we have to prove. If we omit the absolute value sign everywhere, we will get

$$\sum_{I\in\mathcal{D}} c_I(\varphi)c_I(\psi)|I| = \int_{\mathbb{R}} \varphi(t)\psi(t)\,dt \le C\|\varphi\|_{L^p}\|\psi\|_{L^q}$$

which is just the Holder inequality (when $C = 1$). The reader probably remembers that the *direct* proof of the Holder inequality is quite technical and that it is much easier to check the Young inequality

$$\int_{\mathbb{R}} \varphi(t)\psi(t)\,dt \le \frac{1}{p}\int_{\mathbb{R}} |\varphi(t)|^p dt + \frac{1}{q}\int_{\mathbb{R}} |\psi(t)|^q dt$$

(which can be checked *pointwise*) first, and then to use the invariance of the left hand part under the transformations $\varphi \to \lambda\varphi$, $\psi \to \lambda^{-1}\psi$ $(\lambda > 0)$ to get

$$\int_{\mathbb{R}} \varphi(t)\psi(t)\,dt \le \inf_{\lambda>0}\Big\{\frac{\lambda^p}{p}\int_{\mathbb{R}} |\varphi(t)|^p dt + \frac{\lambda^{-q}}{q}\int_{\mathbb{R}} |\psi(t)|^q dt\Big\} = \|\varphi\|_{L^p}\|\psi\|_{L^q}.$$

Maybe we will gain the simplification desired if we follow the same way in our original problem, i.e. if we try to prove the inequality

$$\sum_{I\in\mathcal{D}} |c_I(\varphi)||c_I(\psi)||I| \le C\Big(\int_{\mathbb{R}} |\varphi(t)|^p dt + \int_{\mathbb{R}} |\psi(t)|^q dt\Big) \quad ?$$

The range for the corresponding Bellman function will be now $0 \le \mathcal{B}(F, G, f, g) \le C(F+G)$. Also for technical simplicity (just not to write absolute value signs every time) we will assume that $\varphi, \psi > 0$. This will result in an additional restriction $f, g > 0$ for the domain. The key property remains, of course, unchanged.

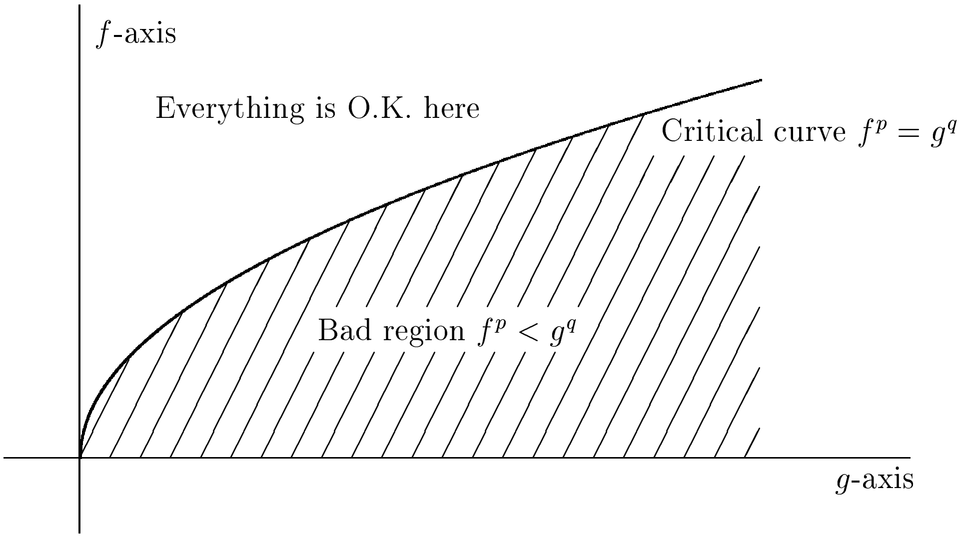

Figure 7: Bad region for the naïve guess $\mathcal{B}_0 = F + G - f^p - g^q$.

For $p = 2$ we may then just take $\mathcal{B}(F, G, f, g) = F + G - f^2 - g^2$ (which is really simpler than the expression with square roots because now all the variables have got *separated*). Our first naive guess will then become

$$\mathcal{B}(F, G, f, g) = F + G - f^p - g^q.$$

Of course, this function still does not satisfy the key property, but the expression now is much simpler and *we can see clearly where and how much we lack.*

Computing the second differential of $\mathcal{B}$, we get

$$d^2\mathcal{B} = -p(p-1)f^{p-2}(df)^2 - q(q-1)g^{q-2}(dg)^2 \leq -2\sqrt{f^{p-2}g^{q-2}}\,|df||dg|$$

for $pq(p-1)(q-1) = pq > 1$. Thereby, if $f^{p-2}g^{q-2} \geq 1$, the key property is satisfied.

Assume for definiteness that $p > 2 > q$. Note that

$$f^{p-2}g^{q-2} \geq 1 \iff (f^p)^{1-\frac{2}{p}} \geq (g^q)^{\frac{2}{q}-1} \iff f^p \geq g^q$$

because $1/p + 1/q = 1$, $p > 2$, and thereby $1 - \frac{2}{p} = \frac{2}{q} - 1 > 0$. So, we lack "strict convexity" only in the region $\{f^p < g^q\}$. Our investigation of the "naive guess" $\mathcal{B}_0 := F + G - f^p - g^q$ is summarized on Figure 7.

8.2. The hidden term.

We would like to add some term $\widetilde{B}(f, g)$ to the *main part* $\mathcal{B}_0 = F + G - f^p - g^q$ to gain strict convexity below the critical curve $f^p = g^q$. More precisely, we want to find a function $0 \leq \widetilde{\mathcal{B}}(f, g) \leq f^p + g^q$ such that the expression

$$\mathcal{B}(F, G, f, g) = 2(F + G) - f^p - g^q - \widetilde{\mathcal{B}}(f, g)$$

satisfies $d^2\mathcal{B} \le -\gamma|df||dg|$ with some $\gamma > 0$ everywhere.

Looking at how we got $|df||dg|$ above the critical curve, we see that probably we should find something which has the term $g^{2-q}(df)^2$ in its second differential. Then this term together with $g^{q-2}(dg)^2$ coming from the "main" part will give us $|df||dg|$ desired. The simplest suitable expression is f^2g^{2-q}. Note that due to the Young inequality

$$f^2g^{2-q} \le \frac{2}{p}f^p + \big(\frac{2}{q} - 1\big)g^q \le f^p + g^q,$$

and the equality sign is attained exactly on the critical curve $f^p = g^q$.

The whole second differential of f^2g^{2-q} is

$$d^2\big(f^2g^{2-q}\big) = f^2g^{2-q}\begin{pmatrix} \dfrac{df}{f} \\ \dfrac{dg}{g} \end{pmatrix}^T \begin{pmatrix} 2 & 2(2-q) \\ 2(2-q) & -(2-q)(q-1) \end{pmatrix}\begin{pmatrix} \dfrac{df}{f} \\ \dfrac{dg}{g} \end{pmatrix}.$$

So, f^2g^{2-q} is not convex (the lower right corner of the matrix is negative). But the term $2f^2g^{2-q}(df/f)^2$ is exactly what we would like to add to $-d^2\mathcal{B}_0$ to get the strict convexity we desire. The unpleasant (negative) diagonal term $-(2-q)(q-1)f^2g^{2-q}\,(dg/g)^2$ can be majorized below the critical curve by a scalar multiple of the term $q(q-1)g^q\,(dg/g)^2$ from $-d^2\mathcal{B}_0$ (recall that below the critical curve we have $f^p < g^q \iff f^2g^{2-q} < g^q$). So for sufficiently small δ the quadratic form $d^2\big[f^p + g^q + \delta f^2g^{2-q}\big]$ has below the critical curve the "diagonal terms" (i.e. terms with $(df)^2$ and $(dg)^2$) exactly as we want. Note that off-diagonal terms (terms with $df\,dg$) are irrelevant, because when computing the determinant, the product of the diagonal terms gives the order δ and the product of the "off-diagonal" ones gives the order δ^2.

This idea doesn't work above the critical curve. But above the critical curve we do not need any additional convexity, so we may replace f^2g^{2-q} by $(2/p)f^p + \big[(2/q) - 1\big]g^q$ which is convex (this is essentially the main part).

Thus we get an impression that the function $\mathcal{B}(F, G, f, g) = 2(F+G) - f^p - g^q - \delta\widetilde{\mathcal{B}}(f, g)$ where

$$\widetilde{\mathcal{B}}(f,g) = \begin{cases} \dfrac{2}{p}f^p + \big(\dfrac{2}{q} - 1\big)g^q\,, & f^p \ge g^q \\ f^2g^{2-q}\,, & f^p \le g^q \end{cases}$$

will satisfy the key property if $\delta > 0$ is sufficiently small.

Note that since the upper line in the definition of $\widetilde{\mathcal{B}}$ is always not less than the lower one, and since both lines define C^∞-smooth in the whole region $\{f > 0, g > 0\}$ functions which coincide on the critical curve, the entire function $\widetilde{\mathcal{B}}$ is also at least C^1-smooth on the critical curve (C^∞-smooth everywhere except the critical curve).

The above heuristic reasoning gives in fact a perfectly correct formal proof (after cleaning some details), so the reader should be convinced at this point that we found a Bellman function. For a suspicious reader who did not believe our "hand waving" we present below all computations:

Computing the second differential $d^2\mathcal{B}$ below the critical curve, we get that $-d^2\mathcal{B}$ equals

$$\begin{pmatrix} \frac{df}{f} \\ \frac{dg}{g} \end{pmatrix}^T \begin{pmatrix} p(p-1)f^p + 2\delta f^2 g^{2-q} & 2(2-q)\delta f^2 g^{2-q} \\ 2(2-q)\delta f^2 g^{2-q} & q(q-1)g^q - (2-q)(q-1)\delta f^2 g^{2-q} \end{pmatrix} \begin{pmatrix} \frac{df}{f} \\ \frac{dg}{g} \end{pmatrix}$$

We want to show that the quadratic form $-d^2\mathcal{B}$ is greater or equal than

$$\delta f^2 g^{2-q}\Big(\frac{df}{f}\Big)^2 + \frac{q(q-1)}{2} g^q \Big(\frac{dg}{g}\Big)^2 \geq \sqrt{2\delta q(q-1)}\ |df|\cdot|dg|\,.$$

That is surely the case if the reminder is non-negatively defined. Checking the determinant (the upper left corner of the matrix of the reminder is always positive) one can easily see that for the non-negativity of the reminder it is enough to have the inequality

$$\delta\Big(\frac{q(q-1)}{2} - (2-q)(q-1)\delta\Big) \geq 4(2-q)^2\delta^2 \qquad \Longleftrightarrow \qquad \delta \leq \frac{q(q-1)}{2(2-q)(7-3q)}$$

(recall that $f^2g^{2-q} < g^q$ below the critical curve).

Since $q > 1$ and thereby $2-q < 1$, and $7-3q < 4$, we observe that $\delta = q(q-1)/8$ would fit and we finally get

$$d^2\mathcal{B} \leq -\frac{q(q-1)}{2}|df|\cdot|dg|.$$

8.3. Remarks

Remark 1. Note how well the term f^2g^{2-q} was hidden in the expression $F+G-f^2-g^2$ for $p=q=2$: in this case it gets just f^2 and therefore is completely indistinguishable from the "trivial" term f^p.

Remark 2. To get the Bellman type function satisfying $\mathcal{B} \leq C'F^{1/p}G^{1/q}$ instead of $C(F+G)$, we may just write

$$\mathcal{B}(F,G,f,g) = \inf_{\lambda>0} \mathcal{B}_1(\lambda^p F, \lambda^{-q}G, \lambda f, \lambda^{-1}g)$$

where $\mathcal{B}_1$ is the function we found above multiplied by $4q^{-1}(q-1)^{-1}$ (to satisfy the key property without any constant). Is the reader able to give an explicit formula? (in principle, it is possible now; of course, the formula is going to be *equivalent* to the infimum, not to be exactly equal to it: the latter is hardly possible and has no sense because it still was only a Bellman *type* function, that we found). If he is, how high (or low) does he estimate the chance of guessing that explicit formula *without* switching to the Young inequality (i.e. right from the expression $(F-f^2)^{1/2}(G-g^2)^{1/2}$)?

Remark 3. The most amazing thing is that the true Bellman function *is known*! This fantastic achievement belongs to Burkholder (see [4]). We didn't use his result here because our goal was *not* to find the best constant for the scalar non-weighted case, but to get an expression *as simple as possible* to be able to proceed beyond this case. In this respect the function found by Burkholder is much harder to work with than that resulting from our simple construction.

9. The coefficient space in L^p case.

We are going to use the same idea as before: to put random signs before the coefficients and to average the norm. It is worth mentioning that this trick is universal. Namely, if we have *any* reflexive Banach space X and a strong unconditional basis of subspaces $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ in it with the coefficient space Y, then we may choose an arbitrary sequence of unit vectors $x_j \in \mathcal{E}_j$ and consider the average norm

$$\|\{\alpha_j\}_{j\in\mathcal{J}}\|_{\mathrm{av}} = \mathbb{E}\Big\|\sum_j \varepsilon_j\alpha_j x_j\Big\|_X$$

where $\alpha_j \in \mathbb{C}$ and ε_j are independent random variables uniformly distributed over the unit circumference $\{z \in \mathbb{C} : |z| = 1\}$ each. The average norm is well-defined for, at least, all sequences $\{\alpha_j\}_{j\in\mathcal{J}}$ containing only finitely many non-zero elements and is an "ideal norm", i.e. $\|\{\alpha'_j\}_{j\in\mathcal{J}}\|_{\mathrm{av}} \le \|\{\alpha_j\}_{j\in\mathcal{J}}\|_{\mathrm{av}}$ whenever $|\alpha'_j| \le |\alpha_j|$ for every $j \in \mathcal{J}$. To prove the last statement, just note that if $|\alpha'_j|/|\alpha_j| = \cos\theta_j$, then

$$\begin{aligned}\mathbb{E}\Big\|\sum_j \varepsilon_j\alpha'_j x_j\Big\|_X &= \mathbb{E}\Big\|\sum_j \varepsilon_j|\alpha'_j| x_j\Big\|_X = \mathbb{E}\Big\|\sum_j \varepsilon_j\Big(\frac{e^{i\theta_j}+e^{-i\theta_j}}{2}\Big)|\alpha_j| x_j\Big\|_X \\ &\le \frac{1}{2}\Big\{\mathbb{E}\Big\|\sum_j \varepsilon_j e^{i\theta_j}|\alpha_j| x_j\Big\|_X + \mathbb{E}\Big\|\sum_j \varepsilon_j e^{-i\theta_j}|\alpha_j| x_j\Big\|_X\Big\} = \mathbb{E}\Big\|\sum_j \varepsilon_j\alpha_j x_j\Big\|_X .\end{aligned}$$

Also, since $\{\mathcal{E}_j\}_{j\in\mathcal{J}}$ is a strong unconditional basis, we should have the average norm $\|\{\alpha_j\}_{j\in\mathcal{J}}\|_{\mathrm{av}}$ equivalent to $\|\{\alpha_j\}_{j\in\mathcal{J}}\|_Y$. Therefore we can always replace the original norm in the coefficient space Y by it. When $\dim\mathcal{E}_j = 1$ for every $j \in \mathcal{J}$, the average norm $\|\cdot\|_{\mathrm{av}}$ can be considered as a canonical norm in the coefficient space. When $\dim\mathcal{E}_j > 1$, different choices of the unit vectors $x_j \in \mathcal{E}_j$ may result in different norms $\|\cdot\|_{\mathrm{av}}$, but, what is important (and what actually distinguishes *strong* unconditional bases from just unconditional ones) is that all the norms corresponding to different possible choices of x_j are equivalent.

Remark 9.1. Note that instead of simple averaging we can average the pth power of the norm, i.e. define the norm $\|\cdot\|_{\mathrm{av},p}$ by

$$\|\{\alpha_j\}_{j\in\mathcal{J}}\|^p_{\mathrm{av},p} = \mathbb{E}\Big\|\sum_j \varepsilon_j\alpha_j x_j\Big\|^p_X .$$

Since we know that the norm in the coefficient space is uniquely defined up to norm equivalence (cf Remark 4.4), we can take as a canonical norm any of the norms $\|\cdot\|_{\mathrm{av},p}$.

To compute the average norms we need the following classical result:

Lemma 9.2 (Khinchin lemma).

$$\Big(\mathbb{E}\Big|\sum_j a_j\varepsilon_j\Big|^p\Big)^{1/p} \asymp \Big(\sum_j |a_j|^2\Big)^{1/2}$$

for every $p \in (0,+\infty)$ and any sequence $\{a_j\}$ of vectors in a (finite-dimensional) Hilbert space.

This lemma for numerical sequences can be found in any probability textbook. To get the result for finite-dimensional Hilbert spaces, it is enough to apply the scalar result to each coordinate; of course, in this approach the constants will blow up as $n \to \infty$.

Note, that the lemma holds for an arbitrary Hilbert space and (at least in the finite-dimensional case) for any $p > -1$. These results are much less known and their proofs are more complicated. We only need the case $\dim E < \infty$, $p > 1$.

9.1. Triebel–Lizorkin space $\mathcal{TL}_p$

9.1.1. The coefficient space Y

Now we are ready to describe the coefficient space Y for the Haar system in scalar, non-weighted L^p. Choose any sequence $\{\alpha_I\}_{I\in\mathcal{D}}$ of complex numbers (with finitely many nonzero elements, say) and consider a random linear combination $\sum \varepsilon_I \alpha_I H_I(t)$ of normalized (in L^p) Haar functions $H_I(t) = |I|^{-1/p} h_I(t)$. Here again ε_I are independent random variables uniformly distributed over the unit circle.

We have

$$\|\{\alpha_I\}_{I\in\mathcal{D}}\|_{\mathrm{av},p} = \Big\{\mathbb{E}\Big\|\sum_{I\in\mathcal{D}} \varepsilon_I \alpha_I H_I(\cdot)\Big\|_{L^p}^p\Big\}^{1/p} = \Big\{\mathbb{E}\int_{\mathbb{R}} \Big|\sum_{I\in\mathcal{D}} \varepsilon_I \alpha_I H_I(t)\Big|^p dt\Big\}^{1/p} =$$
$$= \Big\{\int_{\mathbb{R}} \mathbb{E}\Big(\Big|\sum_{I\in\mathcal{D}} \varepsilon_I \alpha_I H_I(t)\Big|^p\Big) dt\Big\}^{1/p} \asymp \Big\{\int_{\mathbb{R}} \Big(\sum_{I\in\mathcal{D}} |\alpha_I|^2 H_I(t)^2\Big)^{p/2} dt\Big\}^{1/p} =$$
$$= \Big\{\int_{\mathbb{R}} \Big(\sum_{I\in\mathcal{D}} \big(|\alpha_I| \cdot |I|^{-1/p}\big)^2 \chi_I(t)\Big)^{p/2} dt\Big\}^{1/p}.$$

It is easy to check that the space Y of sequences $\{\alpha_I\}_{I\in\mathcal{D}}$ satisfying

$$\|\{\alpha_I\}_{I\in\mathcal{D}}\|_Y := \Big\{\int_{\mathbb{R}} \Big(\sum_I \big(|\alpha_I| \cdot |I|^{-1/p}\big)^2 \chi_I(t)\Big)^{p/2} dt\Big\}^{1/p} < \infty$$

is an ideal Banach space and thereby is the coefficient space for L^p with respect to the Haar system. For every $\varphi = \sum_{I\in\mathcal{D}} c_I(\varphi) h_I \in L^p$ we have

$$\|\varphi\|_{L^p} \asymp \Big\|\big\{\|\mathcal{P}_I\varphi\|_{L^p}\big\}_{I\in\mathcal{D}}\Big\|_Y = \Big\{\int_{\mathbb{R}} \Big(\sum_{I\in\mathcal{D}} \big(\|\mathcal{P}_I\varphi\|_{L^p} \cdot |I|^{-1/p}\big)^2 \chi_I(t)\Big)^{p/2} dt\Big\}^{1/p}.$$

9.1.2. The space $\mathcal{TL}_p$

The expression for the norm $\|\cdot\|_Y$ in the equation above gets a bit simpler in terms of the Haar coefficients $c_I(\varphi)$. Namely, recalling that $\|\mathcal{P}_I\|_{L^p} = |I|^{1/p}|c_I(\varphi)|$, we get

$$\|\varphi\|_{L^p} \asymp \Big\{\int_{\mathbb{R}} \Big(\sum_{I\in\mathcal{D}} |c_I(\varphi)|^2 \chi_I(t)\Big)^{p/2} dt\Big\}^{1/p} =: \big\|\{c_I(\varphi)\}_{I\in\mathcal{D}}\big\|_{\mathcal{TL}_p}. \tag{9.1}$$

Symbol $\|\cdot\|_{\mathcal{TL}_p}$ denotes a norm in the so-called Triebel–Lizorkin space $\mathcal{TL}_p$. [1] As one can guess from the above equality, this space consists of all sequences $\{\alpha_I\}_{I\in\mathcal{D}}$ for which

$$\|\{\alpha_I\}_{I\in\mathcal{D}}\|_{\mathcal{TL}_p} := \left\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}|\alpha_I|^2\chi_I(t)\Big)^{p/2}dt\right\}^{1/p} < \infty$$

The coefficient space Y is then simply a weighted Triebel–Lizorkin space.

The function $S_\varphi(t) = \Big(\sum_{I\in\mathcal{D}}|c_I(\varphi)|^2\chi_I(t)\Big)^{1/2}$ is called the (dyadic) *square function.* The norm equivalence (9.1) above means that $\|\varphi\|_{L^p} \asymp \|S_\varphi\|_{L^p}$ for every $\varphi\in L^p$.

9.1.3. Dual of Triebel–Lizorkin space $\mathcal{TL}_p$

The dual space $\mathcal{TL}_p^*$ is isomorphic (up to the norm equivalence) to $\mathcal{TL}_q$ where $1/p+1/q=1$, with the duality given by

$$\big(\{c_I\}_{I\in\mathcal{D}}\,,\,\{c_I^*\}_{I\in\mathcal{D}}\big) = \sum_{I\in\mathcal{D}} c_I\overline{c_I^*}|I|$$

(it is convenient for us to consider "conjugate linear" duality here).

Indeed, the mapping $\{c_I\}_{I\in\mathcal{D}} \mapsto \{|I|^{1/p}\cdot c_I\}_{I\in\mathcal{D}}$ is an isometric isomorphism from $\mathcal{TL}_p$ to the coefficient space Y (with respect to the Haar system) for L^p. The dual space Y^* is (up to the norm equivalence) the coefficient space for L^q, $(1/p+1/q=1)$, so the mapping $\{c_I\}_{I\in\mathcal{D}} \mapsto \{|I|^{1/q}\cdot c_I\}_{I\in\mathcal{D}}$ is an isomorphism (not necessarily isometric) from $\mathcal{TL}_q$ to Y^*. Writing the duality for Y and Y^*, we get the duality for $\mathcal{TL}_p$ and $\mathcal{TL}_q$.

9.1.4. An L^p representation of $\mathcal{TL}_p$

The space $\mathcal{TL}_p$ appears quite complicated, so it seems that when switching from the investigation of the Hilbert transform in L^p to the investigation of its matrix in $\mathcal{TL}_p$, we just switch from "a bad operator in a good space" to "a good operator in a bad space," i.e. we gain nothing. But let us take a bit closer look at $\mathcal{TL}_p$ to understand whether it is really "bad."

Let $\{c_I\}_{I\in\mathcal{D}}\in\mathcal{TL}_p$. Consider the function f on $\mathbb{R}\times\mathbb{Z}$,

$$f(t,k) = \sum_{I\in\mathcal{D},\,|I|=2^k} c_I\chi_I(t)\,, \qquad t\in\mathbb{R},\ k\in\mathbb{Z}\,.$$

Note that

$$\Big\|\{c_I\}_{I\in\mathcal{D}}\Big\|_{\mathcal{TL}_p} = \left\{\int_{\mathbb{R}}\Big(\big(\sum_k|f(t,k)|^2\big)^{1/2}\Big)^p dt\right\}^{1/p}.$$

[1] The reader familiar with wavelets probably has already met Triebel–Lizorkin spaces: a whole scale of spaces $\dot{\mathbf{f}}_p^{\alpha,q}$ was introduced and studied by M. Frazier and B. Jawerth. Our space $\mathcal{TL}_p$ is (isometrically) isomorphic to the space $\dot{\mathbf{f}}_p^{0,2}$, the mapping

$$\mathcal{TL}_p \ni \{\alpha_I\}_{I\in\mathcal{D}} \mapsto \{|I|^{1/2}\alpha_I\}_{I\in\mathcal{D}} \in \dot{\mathbf{f}}_p^{0,2}$$

giving the canonical isomorphism.

So we can treat f as an element of the ℓ^2-valued L^p space $L^p_{\ell^2}$ (value of f at a point $t \in \mathbb{R}$ is the sequence $\{f(t,k)\}_{k\in\mathbb{Z}}$). Keeping this in mind we can see that the above equality simply means that the norm of the sequence $\{c_I\}_{I\in\mathcal{D}}$ in $\mathcal{TL}_p$ equals the norm of the corresponding f in $L^p_{\ell^2}$ space. Therefore $\mathcal{TL}_p$ can be identified with the subspace of $L^p_{\ell^2}$, which consists of all f for which the functions $f(\cdot,k)$ are constant on every dyadic interval $I \in \mathcal{D}$ of length 2^k, $k \in \mathbb{Z}$.

Note that in this "function representation" the above duality between $\mathcal{TL}_p$ and $\mathcal{TL}_q$ is simply the standard L^p–L^q duality (for ℓ^2-valued L^p):

$$(f,g) = \int_{\mathbb{R}} \sum_{k=-\infty}^{\infty} f(t,k)\overline{g}(t,k)dt = \sum_{k=-\infty}^{\infty} \int_{\mathbb{R}} f(t,k)\overline{g}(t,k)dt$$

(all integrals and sums converge absolutely).

9.1.5. Natural projection from $L^p_{\ell^2}$ onto $\mathcal{TL}_p$.

The natural projection $\mathcal{P}$ from ℓ^2-valued L^p onto $\mathcal{TL}_p$,

$$(\mathcal{P}f)(t,k) = \sum_{I\in\mathcal{D},\,|I|=2^k} \Big(\frac{1}{|I|}\int_I f(s,k)ds\Big)\cdot \chi(t)_I\,,$$

is bounded.

Indeed, consider the natural imbedding $\mathbf{j}$ of $\mathcal{TL}_q$ into the ℓ^2-valued L^q. This operator is clearly bounded, and its dual (which acts from L^p to $\mathcal{TL}_q^* = \mathcal{TL}_p$) is exactly the projection $\mathcal{P}$.

The above properties show that a "matrix" operator $M = \{M_{I,J}\}_{I,J\in\mathcal{D}}$ acts in $\mathcal{TL}_p$ if and only if the corresponding "integral" operator $M\mathcal{P}$ acts in the ℓ^2-valued L^p. So, we are in L^p again! Of course, now it is ℓ^2-valued L^p, but let us remind that we were going to start with an E-valued L^p. So, seems we lose not much (if anything) in the "goodness" of the space, while, as it will get clear a bit later our gain in the "goodness" of the operator will be enormous!

10. The coefficient space for $L^p(W)$

10.1. The coefficient space Y is the same for all $L^p(W)$

We are going to show that, if our assumption that the spaces $\mathcal{E}_I = h_I(\cdot)E$ form a strong unconditional basis in $L^p(W)$ for *any matrix weight* $W \in (\mathcal{A}^{\mathrm{dy}}_{p,q})$ is true, then the corresponding coefficient space *does not* depend on W. Indeed, assume that we have two finite-dimensional Hilbert spaces E_1 and E_2 and two weights $W_{1,2} : \mathbb{R} \to \mathcal{L}_+(E_{1,2})$ satisfying (dyadic) $(\mathcal{A}_{p,q})$ condition. Then we can consider the orthogonal sum $E = E_1 \oplus E_2$ of the spaces E_1 and E_2 and the weight $W : \mathbb{R} \to \mathcal{L}_+(E)$ defined by $W(t)(e_1 \oplus e_2) = (W_1(t)e_1) \oplus (W_2(t)e_2)$. It is easy to check that $W \in (\mathcal{A}^{\mathrm{dy}}_{p,q})$ and therefore the spaces $\mathcal{E}_I = h_I(\cdot)E = h_I(\cdot)(E_1 \oplus E_2)$ should form a strong unconditional basis in $L^p(W)$.

What can the coefficient space Y for this basis be? We remember that to evaluate the norm $\|\{\alpha_I\}_{I\in\mathcal{D}}\|_Y$ of a sequence $\{\alpha_I\}_{I\in\mathcal{D}}$ of complex numbers containing finitely many non-zero elements, one should choose some unit vectors $\varphi_I \in \mathcal{E}_I$ and to find the norm $\|\sum_{I\in\mathcal{D}}\alpha_I\varphi_I\|_{L^p(W)}$. This norm should be equivalent to the norm $\|\{\alpha_I\}_{I\in\mathcal{D}}\|_Y$ for any possible choice of φ_I. Choosing $\varphi_I = \varphi_I^{(1)}\oplus 0$ where $\varphi_I^{(1)} \in \mathcal{E}_I^{(1)} = h_I(\cdot)E_1 \subset L^p(W_1)$, $\|\varphi_I^{(1)}\|_{L^p(W_1)} = 1$, we observe that

$$\|\{\alpha_I\}_{I\in\mathcal{D}}\|_Y \asymp \Big\|\sum_{I\in\mathcal{D}}\alpha_I\varphi_I\Big\|_{L^p(W)} = \Big\|\sum_{I\in\mathcal{D}}\alpha_I\varphi_I^{(1)}\Big\|_{L^p(W_1)} \asymp \|\{\alpha_I\}_{I\in\mathcal{D}}\|_{Y_1}$$

where Y_1 is the coefficient space for $L^p(W_1)$ with respect to the basis $\{\mathcal{E}_I^{(1)}\}$. The choice $\varphi_I = 0\oplus\varphi_I^{(2)}$ yields $\|\{\alpha_I\}_{I\in\mathcal{D}}\|_Y \asymp \|\{\alpha_I\}_{I\in\mathcal{D}}\|_{Y_2}$ and therefore $Y_1 = Y_2 = Y$ up to norm equivalence.

Since we already know that the coefficient space Y for $E=\mathbb{C}$ and $W(t)\equiv 1$ is the weighted Triebel-Lizorkin space, we conclude that it should be *the same weighted* $\mathcal{TL}_p$ for any matrix weight $W\in(\mathcal{A}_{p,q})$.

10.2. The Haar system as a strong unconditional basis in $L^p(W)$

We are going to return to Criterion 3 (Lemma 4.12)again. Recall that the only task remained is to check the imbeddings

$$\Big\|\{\|\varphi\|_{(\mathcal{E}_I^*)^*}\}_{I\in\mathcal{D}}\Big\|_Y \le C\|\varphi\|_{L^p(W)}$$

and

$$\Big\|\{\|\psi\|_{(\mathcal{E}_I)^*}\}_{I\in\mathcal{D}}\Big\|_{Y^*} \le C\|\psi\|_{L^q(W^{-q/p})}.$$

The norms $\|\varphi\|_{(\mathcal{E}_I^*)^*} = |I|^{1/p}\langle\rho^*\rangle_{I,q}^*(c_I(\varphi))$ and $\|\psi\|_{(\mathcal{E}_I)^*} = |I|^{1/q}\langle\rho\rangle_{I,p}^*(c_I(\psi))$ were computed in Section 5.1 and therefore we have the inequalities

$$\Big\|\{|I|^{1/p}\langle\rho^*\rangle_{I,q}^*(c_I(\varphi))\}_{I\in\mathcal{D}}\Big\|_Y \le C\|\varphi\|_{L^p(W)}$$

and

$$\Big\|\{|I|^{1/q}\langle\rho\rangle_{I,p}^*(c_I(\psi))\}_{I\in\mathcal{D}}\Big\|_{Y^*} \le C\|\psi\|_{L^q(W^{-q/p})}$$

to prove. Since everything (including the $(\mathcal{A}_{p,q})$ condition) is symmetric with respect to exchange of p, W and $q, W^{-q/p}$, it is enough to prove one of these inequalities, say, the second that.

10.3. Reduction to "dyadic imbedding theorem"

Though we know the explicit description of the spaces Y and Y^* (which are weighted $\mathcal{TL}_p$ and $\mathcal{TL}_q$ respectively), it doesn't help much because it results in the inequality

$$\Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\langle\rho\rangle_{I,p}^*(c_I(\psi))^2\chi_I(t)\Big)^{q/2}dt\Big\}^{1/q} \le \|\psi\|_{L^q(W^{-q/p})},$$

which (even after raising both parts to the q th power) hardly can be called a "typical dyadic imbedding theorem".

Recall nevertheless (see Section 4.2.3) that, since Y^* is the coefficient space for the scalar non-weighted L^q, a sequence $\{\alpha_I\}_{I\in\mathcal{D}}$ belongs to Y^* if and only if

$$\sup\Big\{\sum_{I\in\mathcal{D}}\|\mathcal{P}_I\varphi\|_{L^p}|\alpha_I| : \varphi\in(\text{scalar non-weighted})\ L^p;\ \|\varphi\|_{L^p}=1\Big\}<+\infty$$

and the norm $\|\{\alpha_I\}_{I\in\mathcal{D}}\|_{Y^*}$ is equivalent to this supremum.

Thus, to prove the desired imbedding is the same as to prove that for every φ in the scalar non-weighted L^p and for every $\psi\in L^q(W^{-q/p})$

$$\sum_{I\in\mathcal{D}}|c_I(\varphi)|\langle\rho\rangle^*_{I,p}(c_I(\psi))|I|\le C\|\varphi\|_{L^p}\|\psi\|_{L^q(W^{-q/p})}$$

(and this clearly *is* a "typical dyadic imbedding theorem"). So, it sounds quite funny, but in order to prove the imbedding, it is better *to not know* what Y^* is in explicit terms !

After the scalar non-weighted case the reader will probably not be surprised by the idea to switch to the Young inequality

$$\sum_{I\in\mathcal{D}}|c_I(\varphi)|\langle\rho\rangle^*_{I,p}(c_I(\psi))|I|\le C\Big(\int_{\mathbb{R}}|\varphi(t)|^p dt+\int_{\mathbb{R}}\|W(t)^{-1/p}\psi(t)\|^q_E\,dt\Big)$$

(the original inequality, as before, can be got from here by taking the infimum of the right hand part over all transformations $\varphi\mapsto\lambda\varphi$, $\psi\mapsto\lambda^{-1}\psi$ $(\lambda>0)$ which leave the left hand part invariant).

10.4. Scalar weighted L^p

To distinguish with the vector case, we will denote the weight by $w(t)$. We have $\langle\rho\rangle^*_{I,p}(x)=\langle w\rangle_I^{-1/p}|x|$ $(x\in\mathbb{C})$ and therefore the inequality to prove can be rewritten as

$$\sum_{I\in\mathcal{D}}|c_I(\varphi)|\frac{|c_I(\psi)|}{\langle w\rangle_I^{1/p}}|I|\le C\Big(\int_{\mathbb{R}}|\varphi(t)|^p dt+\int_{\mathbb{R}}|\psi(t)|^q w(t)^{-q/p}dt\Big).$$

We will show that again this inequality is implied by a condition weaker than $(\mathcal{A}_{p,q})=(A_p)$: the same (dyadic) (A_∞) condition

$$\langle w\rangle_I\le C\exp\big\{\langle\log w\rangle_I\big\}\qquad\text{for every}\quad I\in\mathcal{D}\,,$$

of course. To check that $w\in(A_p)\iff w,w^{-q/p}\in(A_\infty)$ is left to the reader as an easy exercise.

Without loss of generality we may assume that φ and ψ are real-valued positive functions.

10.4.1. The Bellman function: definition and properties

The averages to deal with are now

$$F_I := \langle \varphi^p \rangle_I\,; \qquad G_I := \langle \psi^q w^{-q/p} \rangle_I\,; \qquad u_I := \langle w \rangle_I;$$
$$f_I := \langle \varphi \rangle_I\,; \qquad g_I := \langle \psi \rangle_I\,; \qquad \omega_I := \langle \log w \rangle_I .$$

Again we will try to find an universal Bellman type function, i.e. a function $\mathcal{B}(F,f,G,g,u,\omega)$ satisfying the following properties:

1) Domain: $0 < f \le F^{1/p}$, $0 < g \le G^{1/q} u^{1/p}$, $e^\omega \le u$;

2) Range: $0 \le \mathcal{B}(F,f,G,g,u,\omega) \le 2(F+G)$ (of course, 2 stands here only for technical reasons;

3) Key property:

$$\mathcal{B}(F,f,G,g,u,\omega) \ge$$
$$\ge \Xi(x) u^{-1/p} \left|\frac{f_+ - f_-}{2}\right| \left|\frac{g_+ - g_-}{2}\right| + \frac{1}{2}\Big\{\mathcal{B}(F_+, f_+, \dots) + \mathcal{B}(F_-, f_-, \dots)\Big\}$$

whenever all three corteges (F,f,G,g,u,ω), $(F_\pm, f_\pm, G_\pm, g_\pm, u_\pm, \omega_\pm)$ belong to the domain and

$$(F,f,G,g,u,\omega) = \frac{1}{2}\big\{(F_+, f_+, G_+, g_+, u_+, \omega_+) + (F_-, f_-, G_-, g_-, u_-, \omega_-)\big\}$$

(here $x = ue^{-\omega}$ and $\Xi : [1,+\infty) \to (0,+\infty)$ is a strictly positive decreasing function).

The word-by-word repetition of the reasoning in Sections 6.2.2, 6.3 shows that to find such a function is enough to prove the imbedding and that instead of checking the key property 3) in the finite difference form one may check its infinitesimal version

3$'$) $d^2\mathcal{B} \le -2\Xi(x) u^{-1/p} |df||dg|$.

To give the formal proof we now need only to present a Bellman function. To check its properties (we mainly need to check the strict convexity 3$'$) should not be a problem, especially in the age of computers. So, an impatient reader could go to Section 10.4.5 to see the answer. Pages of computations below are presented to explain the reader *how one can find* a Bellman function. A general idea we use here is very simple: look at already known particular cases (non-weighted L^p and weighted L^2 in our situation) and use your imagination.

10.4.2. The Bellman function: how to guess it?

How could one guess a Bellman type function $\mathcal{B}(F,f,G,g,u,\omega)$ of *six(!)* variables? One of possible ways is the following. We already have the Bellman type function (let's denote it by $\mathcal{B}_0$ now) for the case $w(t)\equiv 1$. Assume that the weight $w(t)$ is constant, but not 1: suppose that $w(t)\equiv u_0$, say. Then to reduce the problem to the non-weighted case, we may use the substitution $\psi\to u_0^{-1/p}\psi$ which is equivalent to the change of the Bellman function $\mathcal{B}_0(F,f,G,g)$ to $\mathcal{B}_0(F,f,G,u_0^{-1/p}g)$. Though the weight $w(t)\in(A_\infty)$ is not necessarily constant, it cannot vary *too* fast from point to point (at least, in average) and thereby it seems reasonable to try essentially the same expression

$$\mathcal{B}(F,f,G,g,u,\omega)=\mathcal{B}_0(F,f,G,u^{-1/p}g)=2(F+G)-f^p-(u^{-1/p}g)^q-\delta\widetilde{\mathcal{B}}(f,u^{-1/p}g)$$

where

$$\widetilde{\mathcal{B}}(f,u^{-1/p}g)=\begin{cases}\dfrac{2}{p}f^p+\big(\dfrac{2}{q}-1\big)(u^{-1/p}g)^q, & \text{if } f^p\ge(u^{-1/p}g)^q;\\ f^2(u^{-1/p}g)^{2-q}, & \text{if } f^p\le(u^{-1/p}g)^q\end{cases}\tag{10.1}$$

when $p>2>q$, or

$$\widetilde{\mathcal{B}}(f,u^{-1/p}g)=\begin{cases}\dfrac{2}{q}(u^{-1/p}g)^q+\big(\dfrac{2}{p}-1\big)f^p, & \text{if } (u^{-1/p}g)^q\ge f^p;\\ (u^{-1/p}g)^2f^{2-p}, & \text{if } (u^{-1/p}g)\le f^p\end{cases}\tag{10.2}$$

when $p<2<q$.

Again we see that the second differential

$$d^2\left[(u^{-1/p}g)^q\right]=d^2[g^qu^{1-q}]=q(q-1)(u^{-1/p}g)^q\left(\frac{dg}{g}-\frac{du}{u}\right)^2$$

degenerates (we used the identity $q/p=q[1-(1/q)]=q-1$). So it seems appropriate to use the same idea as in the L^2 case and to introduce "convexity adding" factors depending on $x=ue^{-\omega}$. We will put these factors onto every term except F and G which results in the expression

$$\mathcal{B}(F,f,G,g,u,\omega)=2(F+G)-f^p\xi(x)-(u^{-1/p}g)^q\zeta(x)-\eta(x)\widetilde{\mathcal{B}}(f,u^{-1/p}g)$$

where $\xi,\zeta,\eta:[1,+\infty)\to(0,1]$.

10.4.3. “Main part” $f^p\xi(x) + (u^{-1/p}g)^q\zeta(x)$

Let us look at how much convexity we can squeeze from the “main part” terms $f^p\xi(x)$ and $(u^{-1/p}g)^q\zeta(x)$. The second differential of $f^p\xi(x)$ is

$$d^2[f^p\xi(x)] = f^p\cdot\begin{pmatrix}\frac{df}{f}\\ \frac{du}{u}\\ d\omega\end{pmatrix}^T\begin{pmatrix}p(p-1)\xi & px\xi' & -px\xi'\\ px\xi' & x^2\xi'' & -x\xi'-x^2\xi''\\ -px\xi' & -x\xi'-x^2\xi'' & x\xi'+x^2\xi''\end{pmatrix}\begin{pmatrix}\frac{df}{f}\\ \frac{du}{u}\\ d\omega\end{pmatrix}$$

$$= f^p\cdot\begin{pmatrix}\frac{df}{f}\\ \frac{du}{u}\\ d\omega-\frac{du}{u}\end{pmatrix}^T\begin{pmatrix}p(p-1)\xi & 0 & -px\xi'\\ 0 & -x\xi' & 0\\ -px\xi' & 0 & x\xi'+x^2\xi''\end{pmatrix}\begin{pmatrix}\frac{df}{f}\\ \frac{du}{u}\\ d\omega-\frac{du}{u}\end{pmatrix}$$

To compute the second differential of $(u^{-1/p}g)^q\zeta(x)$, note again that $q/p = q-1$ and therefore

$$d^2[(u^{-1/p}g)^q\zeta(x)] = d^2[u^{1-q}g^q\zeta(x)] = (u^{-1/p}g)^q\cdot\begin{pmatrix}\frac{dg}{g}, & \frac{du}{u}, & d\omega\end{pmatrix}$$

$$\times\begin{pmatrix}q(q-1)\zeta & -q(q-1)\zeta+qx\zeta' & -qx\zeta'\\ -q(q-1)\zeta+qx\zeta' & q(q-1)\zeta-2(q-1)x\zeta'+x^2\zeta'' & (q-2)x\zeta'-x^2\zeta''\\ -qx\zeta' & (q-2)x\zeta'-x^2\zeta'' & x\zeta'+x^2\zeta''\end{pmatrix}\begin{pmatrix}\frac{dg}{g}\\ \frac{du}{u}\\ d\omega\end{pmatrix}$$

$$= (u^{-1/p}g)^q\cdot\begin{pmatrix}\frac{dg}{g}-\frac{du}{u}\\ \frac{du}{u}\\ d\omega\end{pmatrix}^T\begin{pmatrix}q(q-1)\zeta & qx\zeta' & -qx\zeta'\\ qx\zeta' & 2x\zeta'+x^2\zeta'' & -2x\zeta'-x^2\zeta''\\ -qx\zeta' & -2x\zeta'-x^2\zeta'' & x\zeta'+x^2\zeta''\end{pmatrix}\begin{pmatrix}\frac{dg}{g}-\frac{du}{u}\\ \frac{du}{u}\\ d\omega\end{pmatrix}$$

$$= (u^{-1/p}g)^q\cdot\begin{pmatrix}\frac{dg}{g}-\frac{du}{u}\\ \frac{du}{u}-d\omega\\ d\omega\end{pmatrix}^T\begin{pmatrix}q(q-1)\zeta & qx\zeta' & 0\\ qx\zeta' & 2x\zeta'+x^2\zeta'' & 0\\ 0 & 0 & -x\zeta'\end{pmatrix}\begin{pmatrix}\frac{dg}{g}-\frac{du}{u}\\ \frac{du}{u}-d\omega\\ d\omega\end{pmatrix}$$

These matrices (especially the last one) resemble a lot the matrix

$$\begin{pmatrix} 2\xi & 2x\xi' & 0 \\ 2x\xi' & 2x\xi' + x^2\xi'' & 0 \\ 0 & 0 & -x\xi' \end{pmatrix}$$

we dealt with when considering $L^2(w)$. So it seems quite natural to try the same kind of solution $\xi(x) = \zeta(x) = (1 + \delta x^{-\alpha})/2$ where $0 < \delta < 1$, $\alpha > 0$.

The second differentials $d^2[f^p\xi(x)]$ and $d^2[(u^{-1/p}g)^q\zeta(x)]$ get then

$$\frac{1}{2}f^p \cdot \begin{pmatrix} \dfrac{df}{f} \\ \dfrac{du}{u} \\ d\omega - \dfrac{du}{u} \end{pmatrix}^T \begin{pmatrix} p(p-1) + p(p-1)\delta x^{-\alpha} & 0 & p\alpha\delta x^{-\alpha} \\ 0 & \alpha\delta x^{-\alpha} & 0 \\ p\alpha\delta x^{-\alpha} & 0 & \alpha^2\delta x^{-\alpha} \end{pmatrix} \begin{pmatrix} \dfrac{df}{f} \\ \dfrac{du}{u} \\ d\omega - \dfrac{du}{u} \end{pmatrix}$$

and

$$\frac{1}{2}(u^{-1/p}g)^q \cdot \left(\frac{dg}{g} - \frac{du}{u},\ \frac{du}{u} - d\omega,\ d\omega \right)$$
$$\times \begin{pmatrix} q(q-1) + q(q-1)\delta x^{-\alpha} & -q\alpha\delta x^{-\alpha} & 0 \\ -q\alpha\delta x^{-\alpha} & \alpha(\alpha-1)\delta x^{-\alpha} & 0 \\ 0 & 0 & \alpha\delta x^{-\alpha} \end{pmatrix} \begin{pmatrix} \dfrac{dg}{g} - \dfrac{du}{u} \\ \dfrac{du}{u} - d\omega \\ d\omega \end{pmatrix}$$

respectively.

Recall that $x \geq 1$ in the domain and therefore when $\alpha > 0$ and $[p(p-1)][\alpha^2\delta/2] > [p\alpha\delta]^2 \iff \delta < (p-1)/(2p)$, the matrix $\begin{pmatrix} p(p-1) & 0 & p\alpha\delta x^{-\alpha} \\ 0 & 0 & 0 \\ p\alpha\delta x^{-\alpha} & 0 & (\alpha^2\delta/2)x^{-\alpha} \end{pmatrix}$ is non-negatively defined. This implies immediately that

$$d^2[f^p\xi(x)] \geq \frac{\delta}{2}x^{-\alpha}f^p\Big\{p(p-1)\left(\frac{df}{f}\right)^2 + \alpha\left(\frac{du}{u}\right)^2 + \frac{\alpha^2}{2}\left(d\omega - \frac{du}{u}\right)^2\Big\}.$$

The quadratic form of three variables df, du, $d\omega$ on the right is positively defined: it is clearly non-negative, and can be zero if and only if all three variables are zero.

Using the inequality

$$bY^2 + c(Z-Y)^2 = \frac{b}{2}Y^2 + \left(\frac{b}{2}Y^2 + c(Z-Y)^2\right) \geq$$
$$\geq \frac{b}{2}Y^2 + \left(\frac{2}{b} + \frac{1}{c}\right)^{-1} Z^2 \geq \left(\frac{2}{b} + \frac{1}{c}\right)^{-1}(Y^2 + Z^2),$$

we conclude that

$$d^2[f^p\xi(x)] \geq \gamma_1 x^{-\alpha} f^p \Big\{ \Big(\frac{df}{f}\Big)^2 + \Big(\frac{du}{u}\Big)^2 + (d\omega)^2 \Big\}$$

where

$$\gamma_1 = \frac{\delta}{2}\min\Big\{p(p-1), \Big(\frac{2}{\alpha}+\frac{2}{\alpha^2}\Big)^{-1}\Big\} > 0.$$

Similar argument shows that when $\alpha > 1$ and $[q(q-1)][\alpha(\alpha-1)\delta/2] > [q\alpha\delta]^2 \iff \delta < (q-1)(\alpha-1)/(2\alpha q)$, we have

$$d^2[(u^{-1/p}g)^q\zeta(x)] \geq$$
$$\geq \frac{\delta}{2}x^{-\alpha}(u^{-1/p}g)^q\Big\{q(q-1)\left(\frac{dg}{g}-\frac{du}{u}\right)^2 + \frac{\alpha(\alpha-1)}{2}\left(\frac{du}{u}-d\omega\right)^2 + \alpha\,(d\omega)^2\Big\}.$$

Using the inequality

$$a(X-Y)^2 + b(Y-Z)^2 + cZ^2 \geq \left(\frac{1}{a}+\frac{2}{b}+\frac{3}{c}\right)^{-1}(X^2+Y^2+Z^2)$$

(the proof is left to the reader), we get

$$d^2[(u^{-1/p}g)^q\zeta(x)] \geq \gamma_2 x^{-\alpha}(u^{-1/p}g)^q\Big\{\Big(\frac{dg}{g}\Big)^2 + \Big(\frac{du}{u}\Big)^2 + (d\omega)^2\Big\}$$

with

$$\gamma_2 = \frac{\delta}{2}\Big\{\frac{1}{q(q-1)} + \frac{4}{\alpha(\alpha-1)} + \frac{3}{\alpha}\Big\}^{-1} > 0.$$

Uniting the above estimates, we finally observe that

$$d^2[f^p\xi(x)] + d^2[(u^{-1/p}g)^q\zeta(x)] \geq$$
$$\gamma x^{-\alpha}\Big\{f^p\left(\frac{df}{f}\right)^2 + (u^{-1/p}g)^q\left(\frac{dg}{g}\right)^2 + \Big[f^p + (u^{-1/p}g)^q\Big]\Big[\Big(\frac{du}{u}\Big)^2 + (d\omega)^2\Big]\Big\} \quad (10.3)$$

with $\gamma = \min\{\gamma_1, \gamma_2\} > 0$, provided that $\alpha > 1$ and

$$\delta < \delta_0(\alpha, p, q) = \min\Big\{\frac{p-1}{2p}, \frac{(q-1)(\alpha-1)}{2\alpha q}\Big\}.$$

10.4.4. The "hidden term" $\eta(x)\widetilde{\mathcal{B}}(f, u^{-1/p}g)$

The reader probably remembers that the "hidden term" $\widetilde{\mathcal{B}}$ is of about the same order as the "main part". So it seems quite natural to try to take $\eta(x) = \widetilde{\delta}x^{-\alpha}$ where α is the same as in the main part and $\widetilde{\delta} > 0$ is very-very small.

We have to consider 2 cases:

Case 1: $p > 2 > q$.

"Above the critical curve." Note that "above the critical curve" (i.e. when $f^p \geq (u^{-1/p}g)^q$) addition of $\widetilde{\delta}\widetilde{\mathcal{B}}(f, u^{-1/p}g)$ to the main part is equivalent to replacement of the parameter δ in the definitions of $\xi(x)$ and $\zeta(x)$ by $\delta + (2/p)\widetilde{\delta}$ and $\delta + [(2/q) - 1]\widetilde{\delta}$ respectively. Since both $2/p$, $(2/q) - 1 < 1$, we conclude that if $\widetilde{\delta} > 0$ is so small that $\delta + \widetilde{\delta}$ is still less than $\delta_0(\alpha, p, q)$, the final inequality (10.3) of the previous section will remain true and therefore

$$
\begin{aligned}
-d^2\mathcal{B} \geq \gamma x^{-\alpha}&\Big\{ f^p \left(\frac{df}{f}\right)^2 + (u^{-1/p}g)^q \left(\frac{dg}{g}\right)^2 \Big\} \\
&\geq \gamma x^{-\alpha} 2\sqrt{f^{p-2}(u^{-1/p}g)^{q-2}} \cdot u^{-1/p}|df||dg| \geq 2\gamma x^{-\alpha}u^{-1/p}|df||dg|
\end{aligned}
$$

(recall that when $p > 2 > q$ the condition $f^p \geq (u^{-1/p}g)^q$ defining the region "above the critical curve" is equivalent to $f^{p-2}(u^{-1/p}g)^{q-2} \geq 1$). Thus, the infinitesimal version of the key property is satisfied here with $\Xi(x) = \gamma x^{-\alpha}$.

"Below the critical curve." Recall that below the critical curve the function $\widetilde{\mathcal{B}}(f, u^{-1/p}g)$ is defined by the expression $f^2(u^{-1/p}g)^{2-q}$. The whole "hidden term" $\widetilde{\delta}x^{-\alpha}\widetilde{\mathcal{B}}(f, u^{-1/p}g)$ depends now of four variables f, g, u, ω and, though the dependence is very simple, to write down 4 by 4 matrices seems an unpleasant business. Fortunately, we do not need that. Note that the second differential of the hidden term, whatever it is, has below the critical curve the form

$$\widetilde{\delta}x^{-\alpha}f^2(u^{-1/p}g)^{2-q}\Psi\left(\frac{df}{f}, \frac{dg}{g}, \frac{du}{u}, d\omega\right)$$

where Ψ is a quadratic form with *constant* coefficients depending on p, q and α.

We are going to use the following simple lemma from linear algebra:

Lemma 10.1. *Let $\Phi(x_1, \ldots, x_m) = \sum_{1 \leq i,j \leq m} a_{ij}x_ix_j$ be a quadratic form with real constant coefficients a_{ij} of m real variables $x_1, \ldots, x_m$. If $a_{kk} > 0$, then*

$$\Phi(x_1, \ldots, x_m) \geq \frac{a_{kk}}{2}x_k^2 - A\sum_{j \neq k} x_j^2$$

for sufficiently large $A > 0$ depending on the coefficients a_{ij}.

Proof. Let $a := \max_{1 \leq i,j \leq m} |a_{ij}|$. Denote $\Sigma := \sum_{j \neq k} |x_j|$. We have

$$
\begin{aligned}
&\Phi(x_1, \ldots, x_m) \geq a_{kk}x_k^2 - 2a|x_k|\Sigma - a\Sigma^2 = \\
&\qquad \frac{a_{kk}}{2}x_k^2 + \frac{a_{kk}}{2}\Big(|x_k| - \frac{2a}{a_{kk}}\Sigma\Big)^2 - \Big(a + \frac{2a^2}{a_{kk}}\Big)\Sigma^2 \geq \frac{a_{kk}}{2}x_k^2 - \Big(a + \frac{2a^2}{a_{kk}}\Big)\Sigma^2.
\end{aligned}
$$

Since $\Sigma^2 \leq (m-1)\sum_{j \neq k} x_j^2$, we observe that $A = (m-1)[a + (2a^2)/a_{kk}]$ fits. $\square$

Note now that in the quadratic form $\Psi\Big(\dfrac{df}{f},\dfrac{dg}{g},\dfrac{du}{u},d\omega\Big)$ the coefficient at $\Big(\dfrac{df}{f}\Big)^2$ is 2 and therefore

$$d^2\big[\widetilde{\delta}x^{-\alpha}f^2(u^{-1/p}g)^{2-q}\big]\geq\big[\widetilde{\delta}x^{-\alpha}f^2(u^{-1/p}g)^{2-q}\big]\Big\{\Big(\frac{df}{f}\Big)^2-A\Big[\Big(\frac{dg}{g}\Big)^2+\Big(\frac{du}{u}\Big)^2+(d\omega)^2\Big]\Big\}$$

with some $A=A(\alpha,p,q)>0$.

Combining this estimate with inequality (10.3) of the previous section, we see that "below the critical curve"

$$\begin{aligned}
&-d^2\mathcal{B}\geq\\
&\quad\gamma x^{-\alpha}\Big\{\Big[f^p+\frac{\widetilde{\delta}}{\gamma}f^2(u^{-1/p}g)^{2-q}\Big]\Big(\frac{df}{f}\Big)^2+\Big[(u^{-1/p}g)^q-\frac{\widetilde{\delta}}{\gamma}Af^2(u^{-1/p}g)^{2-q}\Big]\Big(\frac{dg}{g}\Big)^2\\
&\qquad\qquad+\Big[f^p+(u^{-1/p}g)^q-\frac{\widetilde{\delta}}{\gamma}Af^2(u^{-1/p}g)^{2-q}\Big]\Big[\Big(\frac{du}{u}\Big)^2+(d\omega)^2\Big]\Big\}\\
&\quad\geq\gamma x^{-\alpha}\Big\{\frac{\widetilde{\delta}}{\gamma}f^2(u^{-1/p}g)^{2-q}\Big(\frac{df}{f}\Big)^2+\frac{1}{2}(u^{-1/p}g)^q\Big(\frac{dg}{g}\Big)^2\Big\}\geq\big(2\gamma\widetilde{\delta}\big)^{1/2}x^{-\alpha}u^{-1/p}|df||dg|\,,
\end{aligned}$$

provided that $\widetilde{\delta}<\gamma/(2A)$ (recall that $f^2(u^{-1/p}g)^{2-q}\leq(u^{-1/p}g)^q$ below the critical curve).

This means again that the infinitesimal version of the key property is satisfied (now with $\Xi(x)=\big(\gamma\widetilde{\delta}/2\big)^{1/2}x^{-\alpha}$).

Case 2: $p<2<q$

Recall that the function $\widetilde{\mathcal{B}}$ is now

$$\widetilde{\mathcal{B}}(f,u^{-1/p}g)=\begin{cases}\frac{2}{q}(u^{-1/p}g)^q+\big(\frac{2}{p}-1\big)f^p, & \text{if } (u^{-1/p}g)^q\geq f^p;\\ (u^{-1/p}g)^2f^{2-p}, & \text{if } (u^{-1/p}g)^q\leq f^p.\end{cases}$$

The reasoning above the critical curve (i.e. when $(u^{-1/p}g)^q\geq f^p\iff(u^{-1/p}g)^{q-2}f^{p-2}\geq1$) remains essentially unchanged.

Below the critical curve

$$d^2\big[\eta(x)\widetilde{\mathcal{B}}\big]=\widetilde{\delta}x^{-\alpha}(u^{-1/p}g)^2f^{2-p}\Psi\left(\frac{df}{f},\frac{dg}{g},\frac{du}{u},d\omega\right)$$

where Ψ is a quadratic form with constant coefficients depending on p, q and α. Note now that the coefficient at $\Big(\dfrac{dg}{g}\Big)^2$ is 2 and therefore

$$d^2\big[\widetilde{\delta}x^{-\alpha}(u^{-1/p}g)^2f^{2-p}\big]\geq\big[\widetilde{\delta}x^{-\alpha}(u^{-1/p}g)^2f^{2-p}\big]\Big\{\Big(\frac{dg}{g}\Big)^2-A\Big[\Big(\frac{df}{f}\Big)^2+\Big(\frac{du}{u}\Big)^2+(d\omega)^2\Big]\Big\}$$

with some $A=A(\alpha,p,q)>0$.

Again, combining this inequality with (10.3), we see that "below the critical curve"

$$
\begin{aligned}
-\, d^2\mathcal{B} \geq \\
&\gamma x^{-\alpha}\Big\{\Big[f^p - \frac{\widetilde{\delta}}{\gamma}A(u^{-1/p}g)^2 f^{2-p}\Big]\Big(\frac{df}{f}\Big)^2 + \Big[(u^{-1/p}g)^q + \frac{\widetilde{\delta}}{\gamma}(u^{-1/p}g)^2 f^{2-p}\Big]\Big(\frac{dg}{g}\Big)^2 \\
&\qquad + \Big[f^p + (u^{-1/p}g)^q - \frac{\widetilde{\delta}}{\gamma}A(u^{-1/p}g)^2 f^{2-p}\Big]\Big[\Big(\frac{du}{u}\Big)^2 + (d\omega)^2\Big]\Big\} \\
&\geq \gamma x^{-\alpha}\Big\{\frac{1}{2}f^p\Big(\frac{df}{f}\Big)^2 + \frac{\widetilde{\delta}}{\gamma}(u^{-1/p}g)^2 f^{2-p}\Big(\frac{dg}{g}\Big)^2\Big\} \geq \big(2\gamma\widetilde{\delta}\big)^{1/2} x^{-\alpha}u^{-1/p}|df||dg|\,,
\end{aligned}
$$

provided that $\widetilde{\delta} < \gamma/(2A)$ (because now $(u^{-1/p}g)^2 f^{2-p} \leq f^p$ below the critical curve).

10.4.5. Concluding remarks.

Uph-h-h, we are through with scalar $L^p(w)$, and the corresponding Bellman function is

$$\mathcal{B}(F,f,G,g,u,\omega) = 2(F+G) - \frac{1}{2}\big[f^p + (u^{-1/p}g)^q\big](1+\delta x^{-\alpha}) - \widetilde{\delta}x^{-\alpha}\widetilde{\mathcal{B}}(f, u^{-1/p}g);$$

here $\widetilde{\mathcal{B}}(f,u^{-1/p}g)$ is defined by formulae (10.1) and (10.2) from Section 10.4.2; $\alpha > 1$ is arbitrary and $\delta, \widetilde{\delta}$ (to be chosen in this order) are sufficiently small. The corresponding function $\Xi(x)$ in the key property is $\gamma x^{-\alpha}$ ($\gamma = \gamma(\alpha,p,q) > 0$).

Note that again this formula constitutes the complete proof of the imbedding theorem. However hard it was to *invent* this formula, to *check* that the function is strictly concave is the task even a student should be able to carry out (or you may go even further and to give this errand to an artificial "intelligence").

11. Matrix weighted L^p

Recall that our goal is to prove the inequality

$$
\begin{aligned}
\sum_{I\in\mathcal{D}} |c_I(\varphi)|\langle\rho\rangle^*_{I,p}(c_I(\psi))|I| &\leq C\Big(\int_{\mathbb{R}}\varphi(t)^p dt + \int_{\mathbb{R}}\|W(t)^{-1/p}\psi(t)\|^q_E dt\Big) \\
&= C\Big(\int_{\mathbb{R}}\varphi(t)^p dt + \int_{\mathbb{R}}\rho^*_t(\psi(t))^q dt\Big)
\end{aligned}
$$

where $\varphi \in L^p$ is a scalar positive function and $\psi \in L^q(W^{-q/p}) = L^q(\rho^*_t)$. We will try to follow the way we used when considering the case $p = 2$.

Vector L^p space with scalar weight. Let $W(t) = w(t)\,Id$ where w is a scalar positive function satisfying the $(\mathcal{A}_\infty)$-condition. Then the inequality to prove gets

$$\sum_{I\in\mathcal{D}} |c_I(\varphi)|\frac{\|c_I(\psi)\|_E}{\langle w\rangle_I^{1/p}} \leq C\Big(\int_{\mathbb{R}}\varphi(t)^p dt + \int_{\mathbb{R}}\|\psi(t)\|^q_E w(t)^{-q/p} dt\Big)$$

and the Bellman function proving it is just

$$\mathcal{B}(F, f, G, \mathbf{g}, u, \omega) = \mathcal{B}_{\mathrm{scalar}}(F, f, G, \|\mathbf{g}\|_E, u, \omega)$$

where $\mathbf{g}_I = \langle \psi \rangle_I \in E$, $G = \left\langle \|\psi\|_E^q w^{-q/p} \right\rangle_I$ and all other variables are the same as before.

To show this, it is enough to check the infinitesimal version

$$d^2\mathcal{B} \leq -2\Xi(x) u^{-1/p} |df| \|d\mathbf{g}\|_E$$

of the key property. Though we might just repeat all the above computations, we suspect that the reader has got tired of checking the formulae for partial derivatives. So will prefer to give a "soft" proof.

Note first of all that actually we have proved the key property for the scalar Bellman function in a bit stronger form than formally needed. Namely, we proved that

$$-d^2\mathcal{B}_{\mathrm{scalar}} \geq \Xi_1(\dots)(df)^2 + \Xi_2(\dots)(dg)^2$$

where $\Xi_{1,2}(\dots)$ are two positive functions depending on *all* the variables (not on $x = ue^{-\omega}$ only) and such that

$$\Xi_1(\dots)\Xi_2(\dots) \geq \Xi(x)^2 u^{-2/p}$$

everywhere in the domain (exact formulae for $\Xi_{1,2}(\dots)$ depend on whether $p < 2 < q$ or $p > 2 > q$ and on whether we are above or below the critical curve).

The second observation is that the function $\mathcal{B}_{\mathrm{scalar}}$ depends on g in a very simple way; namely, when all the other variables are fixed, $-\mathcal{B}_{\mathrm{scalar}}$ is just a sum of positive powers of g (g^2 or g^q) with positive coefficients and a "g-free" term both below and above the critical curve.

Now everything results from the following lemma.

Lemma 11.1. *Let $\Phi(g, y) = \sum_{k=1}^N A_k(y) g^{\alpha_k} + A_0(y)$ ($\alpha_k > 0$, $A_k(y) \geq 0$ for all $k = 1, \dots, N$; $A_0(y)$ is arbitrary) be a function depending on a real positive variable g and some other set $y = (y_1, \dots, y_m) \in \mathbb{R}^m$ of arguments. Let $\widetilde{\Phi}(\mathbf{g}, y) = \Phi(\|\mathbf{g}\|_E, y)$. If the second differential of Φ satisfies*

$$d^2\Phi \geq \Xi_0(g, y)(dg)^2 + \sum_{k=1}^N \Xi_k(g, y)(dy_k)^2$$

for some nonnegative functions $\Xi_0, \dots, \Xi_N$, then

$$d^2\widetilde{\Phi} \geq \alpha^{-1}\Xi_0(\|\mathbf{g}\|_E, y)\|d\mathbf{g}\|_E^2 + \sum_{k=1}^N \Xi_k(\|\mathbf{g}\|_E, y)(dy_k)^2$$

where $\alpha = \max\{1, \alpha_1 - 1, \dots, \alpha_N - 1\}$.

Applied to $\Phi = -\mathcal{B}_{\mathrm{scalar}}$, this lemma means that the key property for $\mathcal{B}$ is satisfied with, say, $\max\{1, q-1\}^{-1/2}\Xi(x)$ instead of $\Xi(x)$.

Proof. Recall (see Section 6.5 for notation) that $d\|\mathbf{g}\|_E = (d\mathbf{g}, e)_{E,\mathbb{R}}$ where $e = \|\mathbf{g}\|_E^{-1}\mathbf{g}$ and $d^2\|\mathbf{g}\|_E = \|\mathbf{g}\|_E^{-1}\|Qd\mathbf{g}\|_E^2$.

Therefore

$$d^2\widetilde{\Phi}(d\mathbf{g}, dy) = d^2\Phi((d\mathbf{g}, e)_{E,\mathbb{R}}, dy) + \frac{\partial\Phi}{\partial g}\cdot\|\mathbf{g}\|_E^{-1}\|Qd\mathbf{g}\|_E^2\,.$$

Now notice that

$$g^{-1}\frac{\partial\Phi}{\partial g} \ge \alpha^{-1}\frac{\partial^2\Phi}{\partial g^2}$$

(just look at every term separately) and that taking $dy = 0$ in the inequality $d^2\Phi \ge \Xi_0(g,y)(dg)^2 + \sum_{k=1}^N \Xi_k(g,y)(dy_k)^2$, we get

$$\frac{\partial^2\Phi}{\partial g^2} \ge \Xi_0(g,y)\,.$$

Thus,

$$\frac{\partial\Phi}{\partial g}\cdot\|\mathbf{g}\|_E^{-1}\|Qd\mathbf{g}\|_E^2 \ge \alpha^{-1}\Xi_0(\|\mathbf{g}\|_E, y)\|Qd\mathbf{g}\|_E^2$$

which together with the inequality

$$d^2\Phi((d\mathbf{g}, e)_{E,\mathbb{R}}, dy) \ge \Xi_0(\|\mathbf{g}\|_E^2, y)(d\mathbf{g}, e)_E^2 + \sum_{k=1}^N \Xi_k(\|\mathbf{g}\|_E, y)(dy_k)^2$$

and the identity $\|d\mathbf{g}\|_E^2 = (d\mathbf{g}, e)_E^2 + \|Qd\mathbf{g}\|_E^2$ yields the desired result immediately. □

11.1. The Bellman function for $L^p(W)$: preliminary discussion

From here and up to the end of this section we will (almost) completely switch to the "ρ_t-notation".

What arguments should we take for our Bellman function? Clearly, we have 3 positive scalars $F_I = \langle\varphi^p\rangle_I$, $f_I = \langle\varphi\rangle_I$, $G_I = \langle\rho_t^*(\psi(t))^q\rangle_I$, and one vector $g_I = \langle\psi\rangle_I \in E$.

What is U now? The idea to put $U_I = \langle W\rangle_I$ as before is attractive, but unfortunately $\langle\rho\rangle_{I,p}(e)$ is *not* $\|\langle W\rangle_I^{1/p}e\|_E$ when $p \ne 2$. So, we have to define U_I as the average of what is really being averaged, i.e. as $U_I = \langle\rho_t^p\rangle_I$. What is this? Since ρ_t is a norm for every $t \in \mathbb{R}$ and "a norm" is just a function from E to $[0, +\infty)$ with certain properties, U_I is also a function from E to $[0, +\infty)$. A bit closer look shows that $U_I^{1/p}$ is a norm in E.

Let us denote the set of all functions $U : E \to [0, +\infty)$ for which $U^{1/p}$ is a norm by $\mathcal{N}_p(E)$. It is easy to check that $\mathcal{N}_p(E)$ is a convex set.

Remark When $p = 2$ we had $U \in \mathcal{L}_+(E)$, which seems a different thing. Note nevertheless that for every operator $U \in \mathcal{L}_+(E)$ the corresponding quadratic form $(U\cdot,\cdot)_E \in \mathcal{N}_2(E)$. Actually when $\dim E > 1$, the class $\mathcal{N}_2(E)$ contains also a lot of squares of "bad" norms which are not given by any quadratic form. We didn't need them because the quadratic average of "good" norms was a "good" norm again. Now, when we average with power $p \ne 2$, this is no longer true and the whole class $\mathcal{N}_p(E)$ comes into play.

To be able to repeat the construction from Section 7, we need analogues of trace and determinant for $U \in \mathcal{N}_p(E)$.

11.1.1. What is trace now?

Since for $W \in \mathcal{L}_+(E)$ the trace $\operatorname{Tr} W$ is just the sum of values of the quadratic form $(W\cdot,\cdot)_E \in \mathcal{N}_2(E)$ over some orthonormal basis, we may define

$$\operatorname{Tr} U = n^{p/q} \sum_{j=1}^{n} U(e_j)$$

for any $U \in \mathcal{N}_p(E)$ where $e_1, \dots, e_n$ is some *fixed* orthonormal basis in E (fixed, because when $p \neq 2$ the trace defined as above *does* depend on the choice of the basis). The additional factor $n^{p/q}$ is put to provide the inequality

$$U(x) \leq (\operatorname{Tr} U) \|x\|_E^p \qquad \text{for every} \quad x \in E\,.$$

Indeed, if $x = \sum_{j=1}^{n} \alpha_j e_j$ ($\alpha_j \in \mathbb{C}$), we have

$$U^{1/p}(x) \leq \sum_{j=1}^{n} |\alpha_j| U^{1/p}(e_j) \leq \|x\|_E \sum_{j=1}^{n} U^{1/p}(e_j)$$
$$\leq \|x\|_E n^{1/q} \Big(\sum_{j=1}^{n} U(e_j) \Big)^{1/p} = (\operatorname{Tr} U)^{1/p} \|x\|_E\,.$$

11.1.2. What is determinant now?

Consider a "good" norm $\rho(\cdot) = (W\cdot,\cdot)_E^{1/2} = \|W^{1/2}\cdot\|_E$ corresponding to an operator $W \in \mathcal{L}_+(E)$. Let $B_\rho := \{x \in E : \rho(x) \leq 1\}$ be the unit ball in the norm ρ. It is easy to check that $B_\rho = W^{-1/2} B_{\|\cdot\|_E}$ and therefore

$$\operatorname{Vol}(B_\rho) = \left|\det(W^{-1/2})\right|^2 \operatorname{Vol}(B_{\|\cdot\|_E}) = |\det W|^{-1} \operatorname{Vol}(B_{\|\cdot\|_E})\,;$$

here $\operatorname{Vol}(\dots)$ denotes the volume ($2n$-dimensional Lebesgue measure) of the set $\dots$; factor 2 is due to our decision to consider E as a Hilbert space over $\mathbb{C}$, not $\mathbb{R}$. The ratio $v(\rho) = \dfrac{\operatorname{Vol}(B_\rho)}{\operatorname{Vol}(B_{\|\cdot\|_E})}$ makes sense for *any* norm ρ in E and therefore it is reasonable to take the "normalized inverse volume"

$$v^{-1}(U^{1/p}) := \frac{1}{v(U^{1/p})}, \qquad U \in \mathcal{N}_p(E)$$

as an analogue of the determinant.

Lemma 11.2 (trace-determinant inequality). *For every $U \in \mathcal{N}_p(E)$ one has*

$$\left[v^{-1}(U^{1/p})\right]^{p/(2n)} \leq \operatorname{Tr} U.$$

Proof. For any norm ρ we have

$$v(\rho) = \int_S \rho(e)^{-2n} dm(e)$$

where S is the unit sphere in E and dm is the Lebesgue measure on S normalized by the condition $m(S) = 1$. Recalling the inequality $U^{1/p}(e) \le (\operatorname{Tr} U)^{1/p} \|e\|_E$, we get

$$v(U^{1/p}) = \int_S \left[U^{1/p}(e)\right]^{-2n} dm(e) \ge \int_S (\operatorname{Tr} U)^{-2n/p} \|e\|_E^{-2n} dm(e) = (\operatorname{Tr} U)^{-2n/p},$$

which is equivalent to the statement to prove. □

For a norm ρ let $\langle\rho\rangle_{I,0}$ denote its geometric mean, $\langle\rho\rangle_{I,0} := \exp\{\langle\log\rho\rangle_I\}$. Note that for $0 \le p < 1$ the averages $\langle\rho\rangle_{I,p}$ are not necessary norms. But they are homogeneous functions of degree 1, and we still can consider the corresponding volumes.

Lemma 11.3 (Jensen inequality). *Let ρ_t ($t \in \mathbb{R}$) be any family of norms. Then for any $p \ge 0$*

$$v^{-1}(\langle\rho\rangle_{I,p}) \ge \exp\{\langle\log v^{-1}(\rho_t)\rangle_I\}.$$

Proof. We will prove the discrete version of this inequality; namely, we will show that if $\rho_1, \dots, \rho_m$ are any m norms in E and $\rho = (m^{-1}\sum_{j=1}^m \rho_j^p)^{1/p}$, then

$$v^{-1}(\rho) \ge \left\{\prod_{j=1}^m v^{-1}(\rho_j)\right\}^{1/m}.$$

We have

$$v(\rho) = \int_S \rho(e)^{-2n} dm(e) = \int_S \left\{m^{-1}\sum_{j=1}^m \rho_j(e)^p\right\}^{-2n/p} dm(e) \le \int_S \left\{\prod_{j=1}^m \rho_j(e)^{-2n/m}\right\} dm(e)$$
$$\le \prod_{j=1}^m \left\{\int_S \rho_j(e)^{-2n} dm(e)\right\}^{1/m} = \left\{\prod_{j=1}^m v(\rho_j)\right\}^{1/m},$$

which is equivalent to the desired inequality for inverse volumes.

To get the continuous version, one may either apply an appropriate limit procedure, or just use the continuous version of the Holder inequality at the last step. □

11.2. $(\mathcal{A}_{p,0})$-condition.

As before, we will say that a family $\{\rho_t\}$ of norms (or the weight W) satisfies (dyadic) $(\mathcal{A}_{p,0})$ condition if the inverse Jensen inequality

$$v^{-1}(\langle\rho\rangle_{I,p}) \le C\exp\{\langle\log v^{-1}(\rho_t)\rangle_I\}$$

holds for every (dyadic) interval $I \subset \mathbb{R}$.

To show that $\{\rho_t\} \in (\mathcal{A}_{p,q}) \iff \{\rho_t\} \in (\mathcal{A}_{p,0})\,\&\,\{\rho_t^*\} \in (\mathcal{A}_{q,0})$ (or in weight notation $W \in (\mathcal{A}_{p,q}) \iff W \in (\mathcal{A}_{p,0})\,\&\,W^{-q/p} \in (\mathcal{A}_{q,0})$), we need some more work.

11.2.1. Normalizing operators

Let ρ be any norm in E. Consider the family $\mathcal{F}$ of all linear operators $A : E \to E$ satisfying $\rho(A\cdot) \le \|\cdot\|_E$. Clearly, $\mathcal{F}$ is a non-empty compact set of operators. Let V be a solution of the extremal problem

$$A \in \mathcal{F}, \qquad |\det A| \to \max .$$

Since $|\det A|$ is a continuous function of A and since $\varepsilon \cdot Id \in \mathcal{F}$ for $\varepsilon > 0$ small enough, V exists and satisfies $|\det V| > 0$.

Lemma 11.4. *For all* $e \in E$

$$n^{-1}\|e\|_E \le \rho(Ve) \le \|e\|_E .$$

Proof. Since $V \in \mathcal{F}$, we surely have $\rho(Ve) \le \|e\|_E$. So, it is enough to prove the left half of the desired inequality. Assume that it is not true and that for some unit vector $e \in E$ one has $\rho(Ve) < n^{-1}$. Let $r = \rho(Ve)^{-1} > n$. For $\varepsilon \in [0,1]$ consider the operator

$$A_\varepsilon := V\big[(1-\varepsilon)Id + \varepsilon r P_e\big]$$

where $P_e(\cdot) = (\cdot, e)_E e$.

We have

$$\rho(A_\varepsilon\cdot) \le (1-\varepsilon)\rho(V\cdot)+\varepsilon|(\cdot,e)_E| r\rho(Ve) = (1-\varepsilon)\rho(V\cdot)+\varepsilon|(\cdot,e)_E| \le (1-\varepsilon)\|\cdot\|_E+\varepsilon\|\cdot\|_E = \|\cdot\|_E$$

and thereby $A_\varepsilon \in \mathcal{F}$.

On the other hand,

$$|\det A_\varepsilon| = |\det V|\big\{(1-\varepsilon)^{n-1}[1+(r-1)\varepsilon]\big\} .$$

If $r > n$ as we assumed, then the last factor is greater than 1 for small $\varepsilon > 0$ and we get $|\det A_\varepsilon| > |\det V|$, which is impossible. $\square$

We will call the operator V a *normalizing operator* of the norm ρ. The main property of the normalizing operator is the above inequality

$$n^{-1}\|e\|_E \le \rho(Ve) \le \|e\|_E \qquad \forall e \in E.$$

It can be also rewritten in the form

$$n^{-1}\|V^{-1}e\|_E \le \rho(e) \le \|V^{-1}e\|_E \qquad \forall e \in E,$$

which shows that an arbitrary norm ρ in E can be approximated by a "good" norm $\|V^{-1}\cdot\|_E$. Many quantities connected with the norm ρ can be expressed in terms of its normalizing operator V. This allows (if we are not after the best constant, and we have never been after it) to derive results concerning arbitrary norms just from simple well-known facts from linear algebra. To prove that the $(\mathcal{A}_{p,q})$ condition splits into $(\mathcal{A}_{p,0})$ and $(\mathcal{A}_{q,0})$, we need two following lemmas.

Lemma 11.5. *For any norm ρ in E*

$$n^{-2n} \le v^{-1}(\rho)v^{-1}(\rho^*) \le n^{2n}\,.$$

Proof. Since $n^{-1}\|V^{-1}e\|_E \le \rho(e) \le \|V^{-1}e\|_E$, we conclude that $VB_{\|\cdot\|_E} \subset B_\rho \subset nVB_{\|\cdot\|_E}$ and therefore

$$|\det V|^2 \le v(\rho) \le n^{2n}|\det V|^2\,.$$

Switching to the dual norms, we get

$$\|V^*e\|_E \le \rho^*(e) \le n\|V^*e\|_E \qquad \forall e \in E,$$

and therefore

$$n^{-2n}|\det(V^{-1})|^2 \le v(\rho^*) \le |\det(V^{-1})|^2\,.$$

Multiplying these estimates, we get

$$n^{-2n} \le v(\rho)v(\rho^*) \le n^{2n}\,,$$

which is the same as the conclusion of the lemma. □

Lemma 11.6. *Let ρ_1, ρ_2 be two norms in E satisfying $\rho_1 \le \rho_2$ and $v^{-1}(\rho_2) \le Cv^{-1}(\rho_1)$. Then $\rho_2 \le n^{2n}C^{1/2}\rho_1$.*

Proof. Let V_1 and V_2 be the normalizing operators of the norms ρ_1 and ρ_2 respectively. We have

$$n^{-1}\|V_1^{-1}e\|_E \le \rho_1(e) \le \rho_2(e) \le \|V_2^{-1}e\|_E \qquad \forall e \in E,$$

and therefore the operator $B := n^{-1}V_1^{-1}V_2$ is a contraction, i.e. $\|B\| \le 1$.

On the other hand, we have

$$|\det V_2|^2 \ge n^{-2n}v(\rho_2) \ge C^{-1}n^{-2n}v(\rho_1) \ge C^{-1}n^{-2n}|\det V_1|^2$$

and thereby

$$|\det B| = n^{-n}|\det V_2||\det V_1|^{-1} \ge C^{-1/2}n^{-2n}\,.$$

Applying Lemma 7.4, we get $\|B^{-1}\| \le C^{1/2}n^{2n}$, which results in

$$\rho_2(e) \le \|V_2^{-1}e\|_E \le C^{1/2}n^{2n}n^{-1}\|V_1^{-1}e\|_E \le C^{1/2}n^{2n}\rho_1(e) \qquad \forall e \in E\,.$$

□

Assume now that $\{\rho_t\} \in (\mathcal{A}_{p,q})$. By the definition it means that $\langle\rho^*\rangle_{I,q} \le C\langle\rho\rangle_{I,p}^*$. Then $v^{-1}(\langle\rho^*\rangle_{I,q}) \le C^{2n}v^{-1}(\langle\rho\rangle_{I,p}^*)$ and therefore (by Lemma 11.5)

$$v^{-1}(\langle\rho\rangle_{I,p})v^{-1}(\langle\rho^*\rangle_{I,q}) \le C^{2n}n^{2n}\,.$$

On the other hand, $v^{-1}(\rho_t)v^{-1}(\rho_t^*) \ge n^{-2n}$ for every $t \in \mathbb{R}$ (actually, when ρ_t are "good" norms, this product is equal to 1, but it does not matter). Therefore

$$\exp\bigl\{\langle\log v^{-1}(\rho_t)\rangle_I\bigr\}\exp\bigl\{\langle\log v^{-1}(\rho_t^*)\rangle_I\bigr\} \ge n^{-2n}$$

and finally

$$\big[v^{-1}(\langle\rho\rangle_{I,p})\exp\{-\langle\log v^{-1}(\rho_t)\rangle_I\}\big]\big[v^{-1}(\langle\rho^*\rangle_{I,q})\exp\{-\langle\log v^{-1}(\rho_t^*)\rangle_I\}\big]\le n^{4n}C^{2n}\,.$$

Since both the expressions in brackets are at least 1 (due to the Jensen inequality), we get $\{\rho_t\}\in(\mathcal{A}_{p,0})$ and $\{\rho_t^*\}\in(\mathcal{A}_{q,0})$.

Suppose now that we are given that $\{\rho_t\}\in(\mathcal{A}_{p,0})$ and $\{\rho_t^*\}\in(\mathcal{A}_{q,0})$. Then, just writing the above reasoning in the inverse order and using another half of the inequality $n^{-2n}\le v^{-1}(\rho)v^{-1}(\rho^*)\le n^{2n}$ at every step, we arrive to

$$v^{-1}(\langle\rho^*\rangle_{I,q})\le Cv^{-1}(\langle\rho\rangle_{I,p}^*)\,.$$

Since always $\langle\rho^*\rangle_{I,q}\ge\langle\rho\rangle_{I,p}^*$, Lemma 11.6 yields

$$\langle\rho^*\rangle_{I,q}\le n^{2n}C^{1/2}\langle\rho\rangle_{I,p}^*\,,$$

i.e. $\{\rho_t\}\in(\mathcal{A}_{p,q})$.

11.3. The Bellman function: the first attempt.

Now, seems, we are ready to write down the formula for the Bellman function.

The arguments:

$$F_I=\langle\varphi^p\rangle_I\in\mathbb{R};\qquad G_I=\big\langle\rho_t^*(\psi(t))^q\big\rangle_I\in\mathbb{R};\qquad U_I=\langle\rho_t^p\rangle_I\in\mathcal{N}_p(E);$$

$$f_I=\langle\varphi\rangle_I\in\mathbb{R};\qquad g_I=\langle\psi\rangle_I\in E;$$

and, at last,

$$\omega_I=\frac{p}{2n}\big\langle\log v^{-1}(\rho_t)\big\rangle_I=\frac{1}{n}\big\langle\log\det W\big\rangle_I\,.$$

Domain:

$$0<f\le F^{1/p};\qquad (U^{1/p})^*(g)\le G^{1/q};\qquad \big\{v^{-1}(U^{1/p})\big\}^{p/(2n)}\ge e^{\omega}\,.$$

The last two conditions are the Holder inequality (Lemma 7.7) and the Jensen (Lemma 11.3 inequality correspondently. The normalization we chose for ω seems a bit clumsy, but it allows to compare the general case we are investigating now with already considered cases $\dim E=1$ and $p=2$ just literally: notation is in good correspondence.

11.3.1. The key property and construction

We need to construct a function $\mathcal{B}$ of the above six arguments satisfying $0\le\mathcal{B}\le 2(F+G)$ and the following key property:

$$\mathcal{B}(F,f,G,g,U,\omega)\ge\Xi(x)\Big|\frac{f_+-f_-}{2}\Big|(U^{1/p})^*\Big(\frac{g_+-g_-}{2}\Big)+\frac{1}{2}\big\{\mathcal{B}(F_+,f_+,\dots)+\mathcal{B}(F_-,f_-,\dots)\big\}$$

whenever all three corteges of arguments belong to the domain and satisfy

$$(F,f,G,g,U,\omega)=\frac{1}{2}\big\{(F_+,f_+,G_+,g_+,U_+,\omega_+)+(F_-,f_-,G_-,g_-,U_-,\omega_-)\big\}$$

(here $x=\big[v^{-1}(U^{1/p})\big]^{p/(2n)}e^{-\omega}$ is the parameter coming from the $(\mathcal{A}_{p,0})$ condition; again the normalization is chosen to have good correspondence with old notation).

Let us try to follow the lines of Section 7. For every invertible linear operator A in E define

$$\begin{aligned}\mathcal{B}_A(F,f,G,g,U,\omega)&:=\mathcal{B}_{\text{scalar}}(F,f,G,\|A^*g\|_E,\operatorname{Tr}(U\circ A),\omega+\frac{p}{n}\log|\det A|)\\&=2(F+G)-\frac{1}{2}\Big\{f^p+\Big(\frac{\|A^*g\|_E}{[\operatorname{Tr}(U\circ A)]^{1/p}}\Big)^q\Big\}(1+\delta x_A^{-\alpha})-\widetilde{\delta}x_A^{-\alpha}\widetilde{\mathcal{B}}\Big(f,\frac{\|A^*g\|_E}{[\operatorname{Tr}(U\circ A)]^{1/p}}\Big),\end{aligned}$$

where $x_A=\operatorname{Tr}(U\circ A)e^{-\omega}|\det A|^{-p/n}$ and the composition $U\circ A$ is understood in its usual sense, i.e. $(U\circ A)(\cdot)=U(A\cdot)$.

Let us check that the definition is "correct", i.e. that the quantities $g_A=\|A^*g\|_E$, $u_A=\operatorname{Tr}(U\circ A)$ and $\omega_A=\omega+(p/n)\log|\det A|$ satisfy the inequalities $u_A^{-1/p}g_A\le G^{1/q}$ and $u_A\ge e^{\omega_A}$ defining the domain of $\mathcal{B}_{\text{scalar}}$.

The inequality $u_A^{-1/p}g_A\le G^{1/q}$. For any unit vector $e\in E$ we have

$$|(A^*g,e)_E|=|(g,Ae)_E|\le (U^{1/p})^*(g)\cdot U^{1/p}(Ae)\le (U^{1/p})^*(g)\big[\operatorname{Tr}(U\circ A)\big]^{1/p}$$

(recall that $U^{1/p}(Ae)=(U\circ A)^{1/p}(e)\le\big[\operatorname{Tr}(U\circ A)\big]^{1/p}\|e\|_E=\big[\operatorname{Tr}(U\circ A)\big]^{1/p}$).

Therefore

$$g_A=\|A^*g\|_E=\sup_{\|e\|_E=1}|(A^*g,e)_E|\le (U^{1/p})^*(g)\big[\operatorname{Tr}(U\circ A)\big]^{1/p}\le G^{1/q}u_A^{1/p}.$$

The inequality $u_A\ge e^{\omega_A}$: The trace-determinant inequality (Lemma 11.2) gives

$$u_A=\operatorname{Tr}(U\circ A)\ge\big[v^{-1}(U^{1/p}\circ A)\big]^{p/(2n)}=\big[v^{-1}(U^{1/p})\big]^{p/(2n)}|\det A|^{p/n}\ge e^{\omega_A}.$$

Now the key property of the Bellman function for vector L^p with scalar weight yields

$$\begin{aligned}&\mathcal{B}_A(F,f,G,g,U,\omega)\\&\ge\Xi(x_A)\big[\operatorname{Tr}(U\circ A)\big]^{-1/p}\Big|\frac{f_+-f_-}{2}\Big|\cdot\Big\|A^*\frac{g_+-g_-}{2}\Big\|_E+\frac{1}{2}\big\{\mathcal{B}(F_+,f_+,\dots)+\mathcal{B}(F_-,f_-,\dots)\big\}\end{aligned}$$

whenever all three corteges of arguments belong to the domain and satisfy

$$(F,f,G,g,U,\omega)=\frac{1}{2}\big\{(F_+,f_+,G_+,g_+,U_+,\omega_+)+(F_-,f_-,G_-,g_-,U_-,\omega_-)\big\}$$

(here $x_A=u_Ae^{-\omega_A}=\operatorname{Tr}(U\circ A)e^{-\omega}|\det A|^{-p/n}$).

Extremal problem. We are tempted to put $\mathcal{B} = \inf_{\det A \neq 0} \mathcal{B}_{A}$ again. If we were able to show that the infimum is attained at an operator close to the normalizing operator V of the norm $U^{1/p}$, then for our conjectured Bellman function $\mathcal{B}$ we would have the key property with the "strict convexity term" approximately equal to

$$\Xi(x_{V})\big[\mathrm{Tr}(U\circ V)\big]^{-1/p}\Big|\frac{f_+ - f_-}{2}\Big| \cdot \Big\|V^* \frac{g_+ - g_-}{2}\Big\|_E$$

which (the reader is asked to believe us for a while; everything will be checked accurately later) is essentially the same as

$$\Xi(x)\Big|\frac{f_+ - f_-}{2}\Big|(U^{1/p})^*\Big(\frac{g_+ - g_-}{2}\Big)$$

we need.

Let us see whether it is really the case. Forgetting about non-essential (not depending on A) terms and the constant factor $1/2$, we get the extremal problem

$$\det A \neq 0, \qquad \delta x_{A}^{-\alpha} f^p + \Big(\frac{\|A^* g\|_E}{[\mathrm{Tr}(U\circ A)]^{1/p}}\Big)^q (1+\delta x_{A}^{-\alpha}) + 2\widetilde{\delta} x_{A}^{-\alpha}\widetilde{\mathcal{B}}\Big(f, \frac{\|A^* g\|_E}{[\mathrm{Tr}(U\circ A)]^{1/p}}\Big) \longrightarrow \max$$

to investigate.

Again it is natural to factorize $A = VB$ and to denote $\widetilde{U} = U\circ V$ and $\widetilde{g} = V^* g$. Note that

$$\begin{aligned} x_{A} &= \mathrm{Tr}(U\circ A) e^{-\omega} |\det A|^{-p/n} \\ &= \big\{\mathrm{Tr}(\widetilde{U}\circ B)|\det B|^{-p/n}\big\}\big\{|\det V|^{-p/n} e^{-\omega}\big\} =: \big\{\mathrm{Tr}(\widetilde{U}\circ B)|\det B|^{-p/n}\big\}\widetilde{x} \end{aligned}$$

where $\widetilde{x} = |\det V|^{-p/n} e^{-\omega} \asymp \big[v^{-1}(U^{1/p})\big]^{p/(2n)} e^{-\omega} = x$.

Thus our extremal problem can be rewritten as

$$\begin{aligned} &\delta\widetilde{x}^{-\alpha}\frac{|\det B|^{p\alpha/n}}{[\mathrm{Tr}(\widetilde{U}\circ B)]^{\alpha}} f^p + \Big(\frac{\|B^*\widetilde{g}\|_E}{[\mathrm{Tr}(\widetilde{U}\circ B)]^{1/p}}\Big)^q \Big(1 + \delta\widetilde{x}^{-\alpha}\frac{|\det B|^{p\alpha/n}}{[\mathrm{Tr}(\widetilde{U}\circ B)]^{\alpha}}\Big) \\ &\qquad + 2\widetilde{\delta}\,\widetilde{x}^{-\alpha}\frac{|\det B|^{p\alpha/n}}{[\mathrm{Tr}(\widetilde{U}\circ B)]^{\alpha}}\widetilde{\mathcal{B}}\Big(f, \frac{\|B^*\widetilde{g}\|_E}{[\mathrm{Tr}(\widetilde{U}\circ B)]^{1/p}}\Big) \longrightarrow \max . \end{aligned}$$

Again we see that multiplication of B by any scalar factor changes nothing and therefore we may restrict ourselves to operators of norm 1. The first and the third terms are pretty sensitive to degeneration of B and we will hardly have any trouble with them. Let us take a closer look at the middle term

$$\Big(\frac{\|B^*\widetilde{g}\|_E}{[\mathrm{Tr}(\widetilde{U}\circ B)]^{1/p}}\Big)^q \Big(1 + \delta\widetilde{x}^{-\alpha}\frac{|\det B|^{p\alpha/n}}{[\mathrm{Tr}(\widetilde{U}\circ B)]^{\alpha}}\Big) .$$

Let, as before, B_0 be the solution (maybe, degenerated) of the extremal problem

$$\|B\| = 1, \qquad \frac{\|B^*\widetilde{g}\|_E}{[\mathrm{Tr}(\widetilde{U}\circ B)]^{1/p}} \longrightarrow \max .$$

The only way to cut degenerated operators off we know is to show that for some slight perturbation B_ε of the operator B_0 the whole middle term gets greater than $\{[\mathrm{Tr}(\widetilde{U}\circ B_0]^{-1/p}\|B_0^*\widetilde{g}\|_E\}^q$ (then, in order to be better than B_ε, an operator B should make the second factor greater than 1 which would result in a lower bound for $|\det B|$ we long for).

Let us take any operator B_ε satisfying $\|B_\varepsilon - B\| \le \varepsilon$. Since now $\widetilde{U}^{1/p} = U^{1/p}\circ V$ can be almost an arbitrary norm comparable to $\|\cdot\|_E$ (recall the definition of the normalizing operator), the only thing we can say about the expression $\{[\mathrm{Tr}(\widetilde{U}\circ B]^{-1/p}\|B^*\widetilde{g}\|_E\}^q$ is that it is a Lipshitz function of B. Therefore, even in the best case, we can guarantee only that the value of this expression at the operator B_ε is not less than $(1-\varepsilon)$ times its maximal possible value attained at B_0.

On the other hand, if B_0 is an operator of rank 1 (and the L^2-case shows that it is the most probable situation), then, regardless of the choice of B_ε, we have $|\det B_\varepsilon| \le \varepsilon^{n-1}$. Thus, the second factor in the middle term is, roughly speaking, $1+\varepsilon^{p\alpha(n-1)/n}$. Since α should be taken greater than 1, we have no chance to make the whole product $(1-\varepsilon)(1+\varepsilon^{p\alpha(n-1)/n})$ greater than 1 if $p > n/(n-1)$ (constants in the true expressions are not to our favor and we may play with orders only).

Seems we were on the very edge of failure for $p=2$ (recall that we had to take α in a very short interval $\big(1, n/(n-1)\big)$ that time) and now we have failed at last. In fact we were *not* and we have *not*: we just didn't tell the reader the whole truth. And now it is time to open one more card.

11.4. Strong form of the $(\mathcal{A}_{p,0})$ condition.

Note that the definition of the $(\mathcal{A}_{p,q})$ condition

$$\langle\rho^*\rangle_{I,q} \le C\langle\rho\rangle^*_{I,p}$$

makes sense for *every* $q \ge 0$, not for q defined by $1/p + 1/q = 1$ only. In particular, it would be natural to define the $(\mathcal{A}_{p,0})$ condition as

$$\langle\rho^*\rangle_{I,0} := \exp\{\langle\log\rho^*\rangle_I\} \le C\langle\rho\rangle^*_{I,p}\,.$$

We will call this line the $(\mathcal{A}_{p,0})$ condition in the strong form.

What do we have on the left? The function $\langle\rho^*\rangle_{I,0} : E \to [0,+\infty)$ is no longer a norm, but still it is homogeneous of degree 1, i.e.

$$\langle\rho^*\rangle_{I,0}(\beta e) = |\beta|\langle\rho^*\rangle_{I,0}(e) \qquad \text{for every } \beta\in\mathbb{C},\ e\in E.$$

We still can define the unit ball $B_{\langle\rho^*\rangle_{I,0}} = \{x \in E : \langle\rho^*\rangle_{I,0}(x) \le 1\}$ and the normalized inverse volume $v^{-1}(\langle\rho^*\rangle_{I,0})$. The Jensen inequality (Lemma 11.3), yields

$$v^{-1}(\langle\rho^*\rangle_{I,0}) \ge \exp\{\langle\log v^{-1}(\rho_t^*)\rangle_I\}\,.$$

Suppose the family of norms $\{\rho_t\}$ satisfies the $(\mathcal{A}_{p,0})$ condition in the strong form with a constant C. Then

$$v^{-1}(\langle\rho^*\rangle_{I,0}) \le C^{2n} v^{-1}(\langle\rho\rangle^*_{I,p})$$

and Lemma 11.5 implies

$$\left[v^{-1}(\langle\rho\rangle_{I,p}) \exp\left\{-\left\langle \log v^{-1}(\rho_t)\right\rangle_I\right\}\right] \cdot \left[v^{-1}(\langle\rho^*\rangle_{I,0}) \exp\left\{-\left\langle \log v^{-1}(\rho_t^*)\right\rangle_I\right\}\right] \le n^{4n} C^{2n} .$$

Since the second factor is at least 1, the "volume" form of the $(\mathcal{A}_{p,0})$ condition follows immediately. To get the inverse implication, we need the following simple lemma.

Lemma 11.7. *Let ρ be a norm in E. Let*

$$R = \max_{\|e\|_E=1} \rho(e); \qquad r = \min_{\|e\|_E=1} \rho(e) .$$

Then

$$v(\rho) \ge 2^{-2n} r^{-2} R^{-2(n-1)} .$$

Proof. Let e_0 be a unit vector for which $\rho(e_0) = r$. Let $A = (2R)^{-1} Id + (2r)^{-1} P_{e_0}$ where, as before, $P_{e_0} = (\cdot, e_0)_E e_0$. Then for any $x \in E$

$$\rho(Ax) \le (2R)^{-1}\rho(x) + |(x, e_0)_E|(2r)^{-1}\rho(e_0) \le (2R)^{-1} R\|x\|_E + \|x\|_E (2r)^{-1} r = \|x\|_E .$$

Therefore $AB_{\|\cdot\|_E} \subset B_\rho$ and

$$v(\rho) \ge |\det A|^2 \ge 2^{-2n} r^{-2} R^{-2(n-1)} .$$

□

Assume now that for some family of norms $\{\rho_t\}$ and some interval $I \subset \mathbb{R}$ the inequality

$$v^{-1}(\langle\rho\rangle_{I,p}) \le C \exp\left\{\left\langle \log v^{-1}(\rho_t)\right\rangle_I\right\}$$

holds. Our goal is to show that

$$\langle\rho^*\rangle_{I,0} \le C' \langle\rho\rangle^*_{I,p}$$

for some $C' > 0$ depending on the constant C only.

Let V be the normalizing operator of the norm $\langle\rho\rangle_{I,p}$. Note that replacement of the family $\{\rho_t\}$ by the family $\{\rho_t \circ V\}$ changes neither the inequality we are given, nor that to prove. So, without loss of generality, we may assume that $V = Id$, i.e. $n^{-1}\|e\|_E \le \langle\rho\rangle_{I,p}(e) \le \|e\|_E$ for every $e \in E$. Then $\langle\rho\rangle^*_{I,p}(\cdot) \ge \|\cdot\|_E$ and it is enough to prove that $\langle\rho^*\rangle_{I,0}(\cdot) \le C'\|\cdot\|_E$.

For every $t \in I$ let

$$R(t) = \max_{\|e\|_E=1} \rho_t(e), \qquad r(t) = \min_{\|e\|_E=1} \rho(e) .$$

Then, according to the above lemma,

$$v(\rho) \ge 2^{-2n} r(t)^{-2} R(t)^{-2(n-1)}$$

and thereby

$$\log v^{-1}(\rho_t) \le 2n \log 2 + 2 \log r(t) + 2(n-1) \log R(t)\,.$$

Thus

$$n^{-2n} \le v^{-1}(\langle \rho \rangle_{I,p}) \le C \exp\bigl\{\bigl\langle \log v^{-1}(\rho_t)\bigr\rangle_I\bigr\} \\ \le C2^{2n} \exp\bigl\{2(n-1)\bigl\langle \log R(t)\bigr\rangle_I\bigr\} \exp\bigl\{2\bigl\langle \log r(t)\bigr\rangle_I\bigr\}\,.$$

Now observe that for any orthonormal basis $e_1, \ldots, e_n$ of the space E and for any vector $e \in E$ we have

$$\rho_t(e) \le \sum_{j=1}^{n} |(e, e_j)_E| \rho_t(e_j) \le \|e\|_E n^{1/q} \Bigl\{\sum_{j=1}^{n} \rho_t(e_j)^p\Bigr\}^{1/p}\,.$$

From here we can conclude that

$$R(t) \le n^{1/q} \Bigl\{\sum_{j=1}^{n} \rho_t(e_j)^p\Bigr\}^{1/p}$$

and therefore

$$\exp\bigl\{p\bigl\langle \log R(t)\bigr\rangle_I\bigr\} \le \bigl\langle R(t)^p\bigr\rangle_I \le n^{p/q} \sum_{j=1}^{n} \bigl\langle \rho_t(e_j)^p\bigr\rangle_I \le n^{(p/q)+1} = n^p$$

because $\bigl\langle \rho_t(e_j)^p\bigr\rangle_I \le 1$ for every j due to the assumed inequality $\langle \rho \rangle_{I,p}(\cdot) \le \|\cdot\|_E$.

This results in

$$\exp\bigl\{2(n-1)\bigl\langle \log R(t)\bigr\rangle_I\bigr\} \le n^{2(n-1)}$$

and finally in

$$\exp\bigl\{2\bigl\langle \log r(t)\bigr\rangle_I\bigr\} \ge C^{-1}(2n)^{-2n} n^{-2(n-1)}\,,$$

or, which is the same,

$$\exp\Bigl\{\Bigl\langle \log \frac{1}{r(t)}\Bigr\rangle_I\Bigr\} \le C^{1/2}(2n)^n n^{n-1} =: C'\,.$$

It remains to note that the definition of $r(t)$ implies $\rho_t(\cdot) \ge r(t)\|\cdot\|_E$ and thereby $\rho_t^*(\cdot) \le r(t)^{-1}\|\cdot\|_E$ and we get the desired estimate

$$\langle \rho^* \rangle_{I,0}(\cdot) \le \exp\Bigl\{\Bigl\langle \log \frac{1}{r(t)}\Bigr\rangle_I\Bigr\} \|\cdot\|_E \le C'\|\cdot\|_E\,.$$

11.4.1. One more inequality

Though in general $\langle\rho^*\rangle_{I,0}$ is much smaller than $\langle\rho^*\rangle_{I,q}$, we still always have

$$\langle\rho^*\rangle_{I,0} \geq \langle\rho\rangle^*_{I,p}\,.$$

Indeed, for every two vectors $y, z \in E$ and for every $t \in \mathbb{R}$ the inequality $|(y,x)_E| \leq \rho_t^*(y)\rho_t(z)$ holds and therefore

$$\langle\rho^*\rangle_{I,0}(y) = \langle\rho_t(y)\rangle_{I,0} \geq \left\langle \frac{|(y,z)_E|}{\rho_t(z)} \right\rangle_{I,0} = |(y,z)_E|\big[\langle\rho\rangle_{I,0}(z)\big]^{-1}.$$

Thus

$$|(y,z)_E| \leq \langle\rho^*\rangle_{I,0}(y)\langle\rho\rangle_{I,0}(z) \leq \langle\rho^*\rangle_{I,0}(y)\langle\rho\rangle_{I,p}(z)\,,$$

which, by the definition of the dual norm $\langle\rho\rangle^*_{I,p}$, is equivalent to the statement to prove.

11.5. The Bellman function: final version

We will introduce one more variable:

$$\Omega_I = -p\langle\log\rho_t^*\rangle_I\,.$$

Note that Ω is a function from E to $(-\infty,+\infty)$ such that e^Ω is a homogeneous function of degree $-p$, i.e.

$$e^{\Omega(\beta\cdot)} = |\beta|^{-p}e^{\Omega(\cdot)} \qquad \text{for every } \beta \in \mathbb{C}.$$

The above inequality $\langle\rho^*\rangle_{I,0} \geq \langle\rho\rangle^*_{I,p}$ means that

$$\big[e^{-\Omega}\big]^{1/p} \geq (U^{1/p})^*\,,$$

which is the inequality to add to the definition of the domain.

A simple consequence of this inequality is that for every unit vector $z \in E$ and for every invertible operator A we have

$$\begin{aligned}1 = |(z,z)_E|^p &= \big|\big(Az,(A^*)^{-1}z\big)_E\big|^p \\ &\leq \big[U^{1/p}(Az)\big]^p\big[(U^{1/p})^*\big((A^*)^{-1}z\big)\big]^p \leq \operatorname{Tr}(U\circ A)\exp\big\{-\Omega\big((A^*)^{-1}z\big)\big\}\,.\end{aligned}$$

Now we see how to "unload" some power from $|\det A|$ in the expression for the Bellman function written in Section 11.3: we may put a part of it onto e^Ω !

11.5.1. Definition of the Bellman function

Let $\sigma_1,\sigma_2 > 0$, $\sigma_1+\sigma_2 = 1$. Let us take two parameters: an invertible operator A and a unit vector $z \in E$.

Put

$$\begin{aligned}
&\mathcal{B}_{A,z}(F,f,G,g,U,\omega,\Omega)\\
&\qquad := \mathcal{B}_{\text{scalar}}\Big(F,f,G,\|A^*g\|_E,\operatorname{Tr}(U\circ A),\sigma_1\big(\omega+\frac{p}{n}\log|\det A|\big)+\sigma_2\Omega((A^*)^{-1}z)\Big)\\
&\qquad = 2(F+G)-\frac{1}{2}\Big\{f^p+\Big(\frac{\|A^*g\|_E}{[\operatorname{Tr}(U\circ A)]^{1/p}}\Big)^q\Big\}(1+\delta x_{A,z}^{-\alpha})-\widetilde{\delta}x_{A,z}^{-\alpha}\widetilde{\mathcal{B}}\Big(f,\frac{\|A^*g\|_E}{[\operatorname{Tr}(U\circ A)]^{1/p}}\Big),
\end{aligned}$$

where

$$x_{A,z}=\operatorname{Tr}(U\circ A)\big\{e^{\omega}|\det A|^{p/n}\big\}^{-\sigma_1}e^{-\sigma_2\Omega((A^*)^{-1}z)}$$

(the last inequality in the previous section shows that the definition is “correct”).

Define

$$\mathcal{B}(\dots)=\inf\{\mathcal{B}_{A,z}(\dots):\det A\neq 0,\ \|z\|_E=1\}.$$

11.5.2. The extremal problem revisited

Let again V be the normalizing operator of the norm $U^{1/p}$. Using the same factorization $A=VB$ and the notation $\widetilde{U}=U\circ V$, $\widetilde{g}=V^*g$, $\widetilde{x}=|\det V|^{-p/n}e^{-\omega}$, $\widetilde{\Omega}=\Omega\circ(V^*)^{-1}$ and

$$X(B,z)=\left[\frac{|\det B|^{p/n}}{\operatorname{Tr}(\widetilde{U}\circ B)}\right]^{\sigma_1}\left[\frac{\exp\big\{\widetilde{\Omega}\big((B^*)^{-1}z\big)\big\}}{\operatorname{Tr}(\widetilde{U}\circ B)}\right]^{\sigma_2}$$

we come to the extremal problem

$$\begin{aligned}
&\|B\|=1,\ \|z\|_E=1;\\
&\qquad\delta\widetilde{x}^{-\alpha\sigma_1}X(B,z)^\alpha f^p+\left(\frac{\|B^*\widetilde{g}\|_E}{[\operatorname{Tr}(\widetilde{U}\circ B)]^{1/p}}\right)^q\left(1+\delta\widetilde{x}^{-\alpha\sigma_1}X(B,z)^\alpha\right)+\\
&\qquad\qquad+2\widetilde{\delta}\,\widetilde{x}^{-\alpha\sigma_1}X(B,z)^\alpha\widetilde{\mathcal{B}}\Big(f,\frac{\|A^*g\|_E}{[\operatorname{Tr}(U\circ A)]^{1/p}}\Big)\longrightarrow\max
\end{aligned}$$

to solve.

The quantity $\big[\operatorname{Tr}(\widetilde{U}\circ B)\big]^{-1}\exp\big\{\widetilde{\Omega}\big((B^*)^{-1}z\big)\big\}$ Note that as we have already seen, this quantity never exceeds 1. Let us show that, regardless of the choice of the operator B of norm 1, it can be made not less than $n^{-2p}x_1^{-1}$ where

$$x_1=\sup_{y\in E\setminus\{0\}}\frac{e^{-\Omega(y)}}{\big[(U^{1/p})^*y\big]^p}=\left\{\sup_{y\in E\setminus\{0\}}\frac{\langle\rho^*\rangle_{I,0}(y)}{\langle\rho\rangle_{I,p}^*(y)}\right\}^p$$

is the parameter coming from the $(\mathcal{A}_{p,0})$ condition in the strong form.

The main property of a normalizing operator shows that $\widetilde{U}^{1/p}$ is some norm between $n^{-1}\|\cdot\|_E$ and $\|\cdot\|_E$ or, which is the same, $n^{-p}\|\cdot\|_E^p\le\widetilde{U}(\cdot)\le\|\cdot\|_E^p$.

The functions U and Ω are connected by two-sided inequality

$$\big[(U^{1/p})^*(\cdot)\big]^p \le e^{-\Omega(\cdot)} \le x_1\big[(U^{1/p})^*(\cdot)\big]^p,$$

and therefore the same inequality holds for $\widetilde{U}$ and $\widetilde{\Omega}$ (linear transformation $y \mapsto (V^*)^{-1}y$ does not change anything) Thus, we conclude that

$$\|\cdot\|_E^p \le \big[(\widetilde{U}^{1/p})^*(\cdot)\big]^p \le e^{-\widetilde{\Omega}(\cdot)} \le x_1\big[(\widetilde{U}^{1/p})^*(\cdot)\big]^p \le n^p x_1 \|\cdot\|_E^p$$

and therefore

$$e^{\widetilde{\Omega}(\cdot)} \ge n^{-p}x_1^{-1}\|\cdot\|_E^{-p}.$$

For the trace $\mathrm{Tr}(\widetilde{U}\circ B)$ we have

$$\mathrm{Tr}(\widetilde{U}\circ B) = n^{p/q}\sum_{j=1}^n \widetilde{U}(Be_j) \le n^{p/q}\sum_{j=1}^n \|Be_j\|_E^p \le n^{(p/q)+1} = n^p$$

(because $\|B\| = 1$ and $\|e_j\|_E = 1$ for every j), and now the point is that, however degenerated the operator B of norm 1 may be, we always can find *one* unit vector $z \in E$ such that $\|(B^*)^{-1}z\|_E = 1$ and thereby

$$e^{\widetilde{\Omega}((B^*)^{-1}z)} \ge n^{-p}x_1^{-1}\|(B^*)^{-1}z\|_E^{-p} = n^{-p}x_1^{-1}.$$

Combining these two estimates, we see that for any operator B of norm 1

$$\max_{\|z\|_E=1} \frac{\exp\big\{\widetilde{\Omega}\big((B^*)^{-1}z\big)\big\}}{\mathrm{Tr}(\widetilde{U}\circ B)} \in [n^{-2p}x_1^{-1}, 1], \tag{11.1}$$

i.e. is separated from both 0 and ∞.

Therefore, if we take σ_1 so small that $\sigma := \sigma_1\alpha p(n-1)/n < 1$, we will have the second factor in the middle term of order $1+\varepsilon^\sigma$ for small perturbations B_ε of the operator B_0 maximizing the first factor, and the product $(1-\varepsilon)(1+\varepsilon^\sigma)$ will rise above 1 for small $\varepsilon > 0$ even if we have a large constant factor before ε and a very small that before ε^σ. In fact, this means that we are done. The next section will just contain routine computations "justifying" all that we said (we recommend the reader to do these computations by himself and to look at the text only if he gets stuck at some place).

11.5.3. End of the story

Recall that B_0 is the solution of the extremal problem

$$\|B\| = 1, \qquad \frac{\|B^*\widetilde{g}\|_E}{\big[\mathrm{Tr}(\widetilde{U}\circ B)\big]^{1/p}} \longrightarrow \max.$$

Note that since $\widetilde{\mathcal{B}}(f,\cdot)$ is an increasing function when the first argument f is fixed, the operator B_0 maximizes $\widetilde{\mathcal{B}}\big(f, \big[\mathrm{Tr}(\widetilde{U}\circ B)\big]^{-1/p}\|B^*\widetilde{g}\|_E\big)$ as well.

Family B_ε

Lemma 11.8. *For every operator B_0 (in a Hilbert space E, $\dim E = n$) of norm 1 and every $\varepsilon \in [0,1]$ there exists an operator B_ε of norm 1 such that*

$$\|B_\varepsilon - B_0\| \le \varepsilon, \qquad |\det B_\varepsilon| \ge \varepsilon^{n-1}.$$

Proof. We will use well-known fact from linear algebra that every operator B_0 can be represented as $H_1 D_0 H_2$ where $H_{1,2}$ are unitary operators and $D_0 = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ is a diagonal operator with real nonnegative diagonal entries λ_j. Then it is enough to put

$$B_\varepsilon = H_1 D_\varepsilon H_2$$

where $D_\varepsilon = \operatorname{diag}(\max\{\lambda_1, \varepsilon\}, \dots, \max\{\lambda_n, \varepsilon\})$. □

Estimate of $\big[\operatorname{Tr}(\widetilde{U} \circ B_\varepsilon)\big]^{-1/p} \|B_\varepsilon^* \widetilde{g}\|_E$ from below.

The denominator. Recall that $n^{-1}\|\cdot\|_E \le \widetilde{U}^{1/p}(\cdot) \le \|\cdot\|_E$. We saw above that it resulted in $\big[\operatorname{Tr}(\widetilde{U} \circ B)\big]^{1/p} \le n$ for any operator B of norm 1. On the other hand, for any $y \in E$ and any operator B of norm 1 we also have

$$\|By\|_E \le \big[\operatorname{Tr}(\|\cdot\|_E^p \circ B)\big]^{1/p} \|y\|_E \le n\big[\operatorname{Tr}(\widetilde{U} \circ B)\big]^{1/p} \|y\|_E$$

and therefore

$$1 = \|B\| \le n\big[\operatorname{Tr}(\widetilde{U} \circ B)\big]^{1/p}.$$

Uniting these two estimates, we get

$$n^{-1} \le \big[\operatorname{Tr}(\widetilde{U} \circ B)\big]^{1/p} \le n$$

for any operator B of norm 1 and, in particular, for B_0. By the way, it is this estimate from below, which shows that the extremal problem for B_0 is well-posed: the denominator is continuous and separated from 0.

Using the Minkowski inequality

$$\Big(\sum_{j=1}^n a_j^p\Big)^{1/p} \le \Big(\sum_{j=1}^n b_j^p\Big)^{1/p} + \Big(\sum_{j=1}^n |a_j - b_j|^p\Big)^{1/p} \le \Big(\sum_{j=1}^n b_j^p\Big)^{1/p} + n^{1/p} \max_j |a_j - b_j|,$$

we observe that

$$\begin{aligned}
\big[\operatorname{Tr}(\widetilde{U} \circ B_\varepsilon)\big]^{1/p} &= n^{1/q}\Big[\sum_{j=1}^n \widetilde{U}^{1/p}(B_\varepsilon e_j)^p\Big]^{1/p} \\
&\le n^{1/q}\Big[\sum_{j=1}^n \widetilde{U}^{1/p}(B_0 e_j)^p\Big]^{1/p} + n^{1/q} n^{1/p} \max_j \big|\widetilde{U}^{1/p}(B_\varepsilon e_j) - \widetilde{U}^{1/p}(B_0 e_j)\big| \\
&\le \big[\operatorname{Tr}(\widetilde{U} \circ B_0)\big]^{1/p} + n\varepsilon \le (1 + n^2\varepsilon)\big[\operatorname{Tr}(\widetilde{U} \circ B_0)\big]^{1/p}
\end{aligned}$$

because $\widetilde{U}^{1/p}$ is a norm not exceeding $\|\cdot\|_E$, and thereby

$$\big|\widetilde{U}^{1/p}(B_\varepsilon e_j) - \widetilde{U}^{1/p}(B_0 e_j)\big| \le \widetilde{U}^{1/p}\big((B_\varepsilon - B_0)e_j\big) \le \big\|(B_\varepsilon - B_0)e_j\big\|_E \le \|B_\varepsilon - B_0\| \le \varepsilon$$

for every j.

The numerator Note that for $B = Id$ we have

$$\frac{\|B^*\widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U}\circ B)\right]^{1/p}} = \frac{\|\widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U}\circ Id)\right]^{1/p}} \geq n^{-1}\|\widetilde{g}\|_E.$$

Therefore we should also have

$$\frac{\|B_0^*\widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U}\circ B_0)\right]^{1/p}} \geq n^{-1}\|\widetilde{g}\|_E.$$

Since the denominator is not less than n^{-1}, this results in $\|B_0^*\widetilde{g}\|_E \geq n^{-2}\|\widetilde{g}\|_E$. So,

$$\|B_\varepsilon^*\widetilde{g}\|_E \geq \|B_0^*\widetilde{g}\|_E - \varepsilon\|\widetilde{g}\|_E \geq (1-n^2\varepsilon)\|B_0^*\widetilde{g}\|_E$$

and finally

$$\frac{\|B_\varepsilon^*\widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U}\circ B_\varepsilon)\right]^{1/p}} \geq (1-n^2\varepsilon)(1+n^2\varepsilon)^{-1}\frac{\|B_0^*\widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U}\circ B_0)\right]^{1/p}} \geq (1-2n^2\varepsilon)\frac{\|B_0^*\widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U}\circ B_0)\right]^{1/p}}.$$

Note also that the function $\widetilde{\mathcal{B}}$ satisfies

$$\widetilde{\mathcal{B}}(f,\beta\cdot) \geq \beta^{\widetilde{q}}\,\widetilde{\mathcal{B}}(f,\cdot)$$

for any $\beta\in[0,1]$ ($\widetilde{q} = \max\{q,2\}$) and therefore

$$\widetilde{\mathcal{B}}\Big(f,\frac{\|B_\varepsilon^*\widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U}\circ B_\varepsilon)\right]^{1/p}}\Big) \geq (1-2n^2\varepsilon)^{\widetilde{q}}\,\widetilde{\mathcal{B}}\Big(f,\frac{\|B_0^*\widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U}\circ B_0)\right]^{1/p}}\Big).$$

The extremal problem. Recall that it was

$$\|B\| = 1,\ \|z\|_E = 1; \qquad \delta\widetilde{x}^{-\alpha\sigma_1}X(B,z)^\alpha f^p + \left\{\frac{\|B^*\widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U}\circ B)\right]^{1/p}}\right\}^q\left[1+\delta\widetilde{x}^{-\alpha\sigma_1}X(B,z)^\alpha\right]$$
$$+\,2\widetilde{\delta}\,\widetilde{x}^{-\alpha\sigma_1}X(B,z)^\alpha\widetilde{\mathcal{B}}\Big(f,\frac{\|B^*\widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U}\circ B)\right]^{1/p}}\Big) \longrightarrow \max,$$

where

$$X(B,z) = \left[\frac{|\det B|^{p/n}}{\mathrm{Tr}(\widetilde{U}\circ B)}\right]^{\sigma_1}\left[\frac{\exp\{\widetilde{\Omega}((B^*)^{-1}z)\}}{\mathrm{Tr}(\widetilde{U}\circ B)}\right]^{\sigma_2}$$

and $\widetilde{x} = |\det V|^{-p/n}e^{-\omega}$.

As we saw in Section 11.5.2 (see (11.1) there), for any operator B we can choose a unit vector z such that

$$X(B,z) \geq \left[\frac{|\det B|^{p/n}}{\mathrm{Tr}(\widetilde{U}\circ B)}\right]^{\sigma_1}\left(n^{-2p}x_1^{-1}\right)^{\sigma_2} \geq n^{-2p}x_1^{-\sigma_2}|\det B|^{p\sigma_1/n}$$

(because always $\mathrm{Tr}(\widetilde{U} \circ B) \le n^p \le n^{2p}$).

For the operator B_ε it means that it is always possible to find a unit vector $z_\varepsilon \in E$ such that

$$X(B_\varepsilon, z_\varepsilon) \ge n^{-2p} x_1^{-\sigma_2} \varepsilon^{\sigma_1 p(n-1)/n}$$

and therefore

$$\left\{ \frac{\|B_\varepsilon^* \widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U} \circ B_\varepsilon)\right]^{1/p}} \right\}^q (1 + \delta \widetilde{x}^{-\alpha\sigma_1} X(B_\varepsilon, z_\varepsilon)^\alpha)$$
$$\ge \left\{ \frac{\|B_0^* \widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U} \circ B_0)\right]^{1/p}} \right\}^q (1 - 2n^2\varepsilon)^q \left[1 + \delta(n^{2p}\widetilde{x}^{\sigma_1} x_1^{\sigma_2})^{-\alpha} \varepsilon^\sigma\right] =: \mathcal{F}_1(\varepsilon) \left\{ \frac{\|B_0^* \widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U} \circ B_0)\right]^{1/p}} \right\}^q$$

where $\sigma = \sigma_1 \alpha p(n-1)/n$.

Also we see that

$$X(B_\varepsilon, z_\varepsilon)^\alpha \widetilde{\mathcal{B}}\Big(f, \frac{\|B_\varepsilon^* \widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U} \circ B_\varepsilon)\right]^{1/p}}\Big)$$
$$\ge (n^{2p} x_1)^{-\alpha} (1 - 2n^2\varepsilon)^{\widetilde{q}} \varepsilon^\sigma \widetilde{\mathcal{B}}\Big(f, \frac{\|B_0^* \widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U} \circ B_0)\right]^{1/p}}\Big) =: \mathcal{F}_2(\varepsilon) \widetilde{\mathcal{B}}\Big(f, \frac{\|B_0^* \widetilde{g}\|_E}{\left[\mathrm{Tr}(\widetilde{U} \circ B_0)\right]^{1/p}}\Big)$$

and, at last,

$$X(B_\varepsilon, z_\varepsilon)^\alpha f^p \ge (n^{2p} x_1)^{-\alpha} \varepsilon^\sigma f^p =: \mathcal{F}_3(\varepsilon) f^p.$$

If we take $\sigma_1 > 0$ so small that $\sigma < 1$, we can find $\varepsilon = \varepsilon(\widetilde{x}, x_1)$ such that $\mathcal{F}_1(\varepsilon) > 1$, $\mathcal{F}_{2,3}(\varepsilon) > 0$. Clearly, we can even make $\varepsilon(\widetilde{x}, x_1)$ a continuous function of $\widetilde{x}, x_1$ (the reader who likes explicit formulae can easily check that $\varepsilon = (R\widetilde{x}x_1)^{-R}$, where $R > 0$ is a sufficiently large number depending on n, α, δ and σ, fits).

Define $\nu_{1,2,3}(\widetilde{x}, x_1)$ by

$$\mathcal{F}_1\big(\varepsilon(\widetilde{x}, x_1)\big) = 1 + \delta \widetilde{x}^{-\alpha\sigma_1} \nu_1(\widetilde{x}, x_1)^\alpha; \qquad \mathcal{F}_{2,3}\big(\varepsilon(\widetilde{x}, x_1)\big) = \nu_{2,3}(\widetilde{x}, x_1)^\alpha .$$

Clearly, all $\nu_{1,2,3}$ are positive continuous functions of $\widetilde{x}, x_1$.

Let $\nu(\widetilde{x}, x_1) := \min\{\nu_1(\widetilde{x}, x_1), \nu_2(\widetilde{x}, x_1), \nu_3(\widetilde{x}, x_1)\}$.

Then, in order to be better than $(B_\varepsilon, z_\varepsilon)$, a pair (B, z) should satisfy $X(B, z) \ge \nu(\widetilde{x}, x_1)$, i.e.

$$\left[\frac{|\det B|^{p/n}}{\mathrm{Tr}(\widetilde{U} \circ B)}\right]^{\sigma_1} \left[\frac{\exp\{\widetilde{\Omega}((B^*)^{-1} z)\}}{\mathrm{Tr}(\widetilde{U} \circ B)}\right]^{\sigma_2} \ge \nu(\widetilde{x}, x_1)$$

(otherwise *every* term in the expression to maximize is less than the corresponding term for the pair $(B_\varepsilon, z_\varepsilon)$).

Since the second factor never exceeds 1 and $\mathrm{Tr}(U \circ B) \ge n^{-p}$, it means in particular that $|\det B|$ should be at least $n^{-n} \nu(\widetilde{x}, x_1)^{n/(p\sigma_1)}$ and therefore the solution $B_{\mathrm{opt}}, z_{\mathrm{opt}}$ of the extremal problem in the class of non-degenerated operators exists and satisfies

$$X(B_{\mathrm{opt}}, z_{\mathrm{opt}}) \ge \nu(\widetilde{x}, x_1); \qquad \|B_{\mathrm{opt}}^{-1}\| \le |\det B_{\mathrm{opt}}|^{-1} \le n^n \nu(\widetilde{x}, x_1)^{-n/(p\sigma_1)} .$$

Key property of the Bellman function Recall that for the Bellman function $\mathcal{B}$ defined above as

$$\mathcal{B}(\dots) = \inf\bigl\{\mathcal{B}_{A,z}(\dots) : \det A \neq 0, \|z\|_E = 1\bigr\}$$

we always have the key property with the strict convexity term

$$\Xi(x_{A_{\rm opt},z_{\rm opt}})\bigl[\mathrm{Tr}(U\circ A_{\rm opt})\bigr]^{-1/p}\Bigl|\frac{f_+ - f_-}{2}\Bigr|\cdot\Bigl\|A^*_{\rm opt}\frac{g_+ - g_-}{2}\Bigr\|_E$$

where $A_{\rm opt} = VB_{\rm opt}$ and Ξ is the function corresponding to the case of vector L^p with scalar weight.

To run the machine backward and to finish the story, we need the key property with the strict convexity term

$$\Xi'(x,x_1)\Bigl|\frac{f_+ - f_-}{2}\Bigr|(U^{1/p})^*\Bigl(\frac{g_+ - g_-}{2}\Bigr).$$

Let us show that these two expressions are essentially the same. We have

$$x_{A_{\rm opt},z_{\rm opt}} = \widetilde{x}^{\sigma_1}X(B_{\rm opt}, z_{\rm opt})^{-1} \le \widetilde{x}^{\sigma_1}\nu(\widetilde{x},x_1)^{-1};$$

$$\mathrm{Tr}(U\circ A_{\rm opt}) = \mathrm{Tr}(\widetilde{U}\circ B_{\rm opt}) \le n^p$$

and, at last, (recall that $\|V^*\cdot\|_E \ge n^{-1}(U^{1/p})^*(\cdot)$)

$$\begin{aligned}\Bigl\|A^*_{\rm opt}\frac{g_+ - g_-}{2}\Bigr\|_E &= \Bigl\|B^*_{\rm opt}V^*\frac{g_+ - g_-}{2}\Bigr\|_E \\ &\ge \frac{1}{\|B^{-1}_{\rm opt}\|}n^{-1}(U^{1/p})^*\Bigl(\frac{g_+ - g_-}{2}\Bigr) \ge n^{-n}\nu(\widetilde{x},x_1)^{n/(p\sigma_1)}n^{-1}(U^{1/p})^*\Bigl(\frac{g_+ - g_-}{2}\Bigr).\end{aligned}$$

So, finally we observe that the strict convexity term we really have is at least

$$n^{-(n+2)}\nu(\widetilde{x},x_1)^{n/(p\sigma_1)}\Xi\bigl(\widetilde{x}^{\sigma_1}\nu(\widetilde{x},x_1)^{-1}\bigr)\Bigl|\frac{f_+ - f_-}{2}\Bigr|(U^{1/p})^*\Bigl(\frac{g_+ - g_-}{2}\Bigr).$$

It remains to note that (due to the inequality $n^{-1}\|V^{-1}\cdot\|_E \le U^{1/p}(\cdot) \le \|V^{-1}\cdot\|_E$) we have

$$1 \le x = \bigl[v^{-1}(U^{1/p})\bigr]^{p/(2n)}e^{-\omega} \le \widetilde{x} = |\det V|^{-p/n}e^{-\omega} \le n^p x\,,$$

so we may put

$$\Xi'(x,x_1) := \min\bigl\{n^{-(n+2)}\nu(s,t)^{n/(p\sigma_1)}\Xi(s^{\sigma_1}\nu(s,t)^{-1}) : 1 \le s \le n^p x,\ 1 \le t \le x_1\bigr\}$$

(since ν, Ξ are strictly positive continuous functions, Ξ' is strictly positive and continuous as well; the minimum over $t \in [1,x_1]$ is taken to make Ξ' decreasing in both variables x, x_1).

11.5.4. Final remarks

1) It is easy to check that the resulting function $\Xi'(x,x_1)$ behaves like $\gamma x^{-R}x_1^{-R}$ where $\gamma, R > 0$ are some constants depending on $n = \dim E$ only (if we consider α and δ fixed and choose σ_1 and $\varepsilon(\widetilde{x},x_1)$ in some reasonable way for every n).

2) Though we treated $\omega_I = [p/(2n)]\langle\log v^{-1}(U^{1/p})\rangle_I$ and $\Omega_I = -p\langle\log\rho_t^*\rangle_I$ as independent variables, the equivalence of the "volume" and the "strong" forms of the $(\mathcal{A}_{p,0})$ condition shows that they are connected and that, if one wants to describe the true domain (i.e. all corteges $(F, f, G, g, U, \omega, \Omega)$ which can be realized by some functions φ and ψ and some family of norms $\{\rho_t\}$), he should write down some sophisticated relation connecting all three variables U, ω and Ω instead of two simple inequalities

$$\left[v^{-1}(U^{1/p})\right]^{p/(2n)} \geq e^{\omega} \qquad \text{and} \qquad \left[e^{-\Omega}\right]^{1/p} \geq (U^{1/p})^*$$

we used for our "domain".

This observation allows also to exclude any of two variables ω and Ω. All one needs to do for that is to take *maximum* of the constructed above Bellman function of 7 variables over the variable to exclude (of course, the maximum should be taken in the *true* domain). We needed both of them not for *existence* of the Bellman function (it exists because the theorem is true and no other reason is needed), but for possibility to write down a *relatively simple expression* for it (actually to exclude ω is not very hard, but if the reader decides to try to exclude Ω in an explicit way, all we can say to him is "Good luck !").

3) It is worth mentioning that the strong form of the $(\mathcal{A}_{p,0})$ condition, which was invented by one of the authors just to save the construction, is actually the only possible analogue of (A_∞) when $\dim E = \infty$ (we do not know whether it is sufficient for anything there). It also suggests to consider the whole family of $(\mathcal{A}_{p,q})$ conditions corresponding to all possible pairs of nonnegative numbers p and q. This family possesses a lot of nice properties. Actually the whole theory of scalar (A_p)-weights can be transferred to the matrix case except two results which essentially use degeneration $w \in (\mathcal{A}_{p,q}) \iff w \in (\mathcal{A}_{1,q/p})$ of scalar case: the Peter Jones factorization and the Rubio-de-Francia extrapolation theory. By today (May 1, 1996) we do not know what the analogues of these two things are in high dimensions.

4) It is interesting to compare our proof for matrix weighted L^2 with that in [2]. Essentially they coincide, but where they had to *invent a chain of reasoning* and to *prove statements*, we had just to *guess what simple factor one should introduce* to add some convexity and to *calculate second derivatives*. This correspondence is almost one-to-one: every lemma in [2] is just a part of the Bellman function

$$F - (1/2)u^{-1}f^2(1 + \delta(ue^{-\omega})^{-\alpha})$$

and the whole proof is just it.

We leave to the reader to decide up to what extent this correspondence between sophisticated reasonings proving "imbedding theorems" and "structures of convexity" of simple functions in funny regions is universal. And now we are going to leave the "Bellman function approach" and to return to more traditional techniques.

12. Classical Calderon-Zygmund operators

Let T be a bounded operator in scalar, non-weighted L^2. We call T a Calderon-Zygmund operator if there exists a kernel $K : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ such that

1) $(T\varphi)(s) = \int_{\mathbb{R}} K(s,t)\varphi(t)dt$
for every bounded function φ with bounded support and every s for which $\operatorname{dist}(s, \operatorname{supp}\varphi) > 0$;

2) $|K(s,t)| \le \dfrac{Const}{|s-t|}$;

3) $|K(s,t) - K(s_0,t)|,\ |K(t,s) - K(t,s_0)| \le Const \dfrac{|s-s_0|^\varepsilon}{|t-s_0|^{1+\varepsilon}}$
for some $\varepsilon > 0$ whenever $|t-s_0| \ge 2|s-s_0|$.

We will call a Calderon-Zygmund operator classical, if

$$T1 = T^*1 = 0.$$

The formal meaning of this strange line is

$$\lim_{R\to\infty} (T\chi_{[-R,R]}, h_I)_{L^2} = \lim_{R\to\infty} (h_I, T^*\chi_{[-R,R]})_{L^2} = 0$$

for every Haar function h_I.

It is well-known and easy to check that the Hilbert transform is a classical Calderon-Zygmund operator (with $\varepsilon = 1$).

12.1. Matrix of a classical Calderon-Zygmund operator in the Haar basis

Since $(Th_I, h_J) = (h_I, T^*h_J)$ and T^* is an operator of exactly the same kind as T, it is enough to consider the case $|J| \le |I|$.

Lemma 12.1. *Let $s_0, s \in \mathbb{R}$. Let $S \subset \mathbb{R}$ be any bounded measurable set such that $d := dist(s_0, S) \ge 2|s-s_0|$. Then*

$$\int_S |K(s,t) - K(s_0,t)|dt \le Const|s-s_0|^\varepsilon d^{-\varepsilon} \frac{D}{D+d}$$

where $D := \operatorname{diam} S$.

Proof. Note that for every $t \in S$ one has $d \le |t-s_0| \le d+D$. Since $|t-s_0| \ge 2|s-s_0|$ for every $t \in S$, the integral can be estimated by

$$Const \int_S \frac{|s-s_0|^\varepsilon}{|t-s_0|^{1+\varepsilon}} dt \le Const|s-s_0|^\varepsilon \cdot 2 \int_d^{d+D} \frac{dt}{t^{1+\varepsilon}} \le 2Const|s-s_0|^\varepsilon \min\Big\{\frac{D}{d^{1+\varepsilon}}, \frac{1}{\varepsilon d^\varepsilon}\Big\}.$$

Considering separately cases $d \le D$ and $d \ge D$, we see that the last inequality is equivalent to the statement of the lemma. $\square$

Lemma 12.2. *Let φ be any bounded function with bounded support S. Let $J \subset \mathbb{R}$ be any interval satisfying $d := \operatorname{dist}(J,S) \ge |J|$.*

Then

$$|(T\varphi, h_J)| \le Const \|\varphi\|_{L^\infty} |J|^{1+\varepsilon} d^{-\varepsilon} \frac{D}{D+d}$$

where as before $D = \operatorname{diam} S$.

Proof. Let s_0 be the center of J. We have

$$|(T\varphi, h_J)| = \Big| \iint_{J\times S} K(s,t)\varphi(t) h_J(s)\,ds\,dt \Big|$$
$$= \Big| \iint_{J\times S} \big[K(s,t) - K(s_0,t)\big]\varphi(t) h_J(s)\,dsdt \Big| \le \|\varphi\|_{L^\infty} |J| \sup_{s\in J} \int_S |K(s,t) - K(s_0,t)|\,dt\,.$$

Since $d = dist(J,S) \ge |J| \ge 2|s-s_0|$ for every $s \in J$, we conclude that the supremum does not exceed

$$Const|s-s_0|^\varepsilon d^{-\varepsilon} \frac{D}{D+d} \le Const|J|^\varepsilon d^{-\varepsilon} \frac{D}{D+d}$$

and the desired result follows immediately. $\square$

Lemma 12.3. *Let I, J be any two intervals satisfying $|I| \ge |J|$ and $\rho(J,I) \ge |J|$ where $\rho(J,I)$ is the distance from J to the nearest of three points: the middlepoint of the interval I and two its ends (if J contains one of these points, we put $\rho(J,I) = 0$).*

Then

$$|(Th_I, h_J)| \le |J|^{1+\varepsilon} \rho(J,I)^{-\varepsilon} \frac{|I|}{|I| + \rho(J,I)}\,.$$

Proof. If $\rho(J,I) \ge |J|$, two cases are possible:

1) $I \cap J = \varnothing$ and $\rho(J,I) = \operatorname{dist}(I,J)$. Then everything follows immediately from Lemma 12.2 applied to $\varphi = h_I$.

2) J lies "deep inside" I_+ or I_-. Assume for definiteness that $J \subset I_+$. Take $R > 0$ and put

$$\varphi_R = h_I - \chi_{[-R,R]}.$$

For large R the function $\varphi_R \equiv 0$ on I_+ and $\operatorname{dist}(J, \operatorname{supp}\varphi_R) = \rho(J,I)$. Therefore (by Lemma 12.2 again)

$$|(T\varphi_R, h_J)| \le Const|J|^{1+\varepsilon}\rho(J,I)^{-\varepsilon} \frac{2R}{2R + \rho(J,I)}\,.$$

Taking the limit as $R \to \infty$ and recalling that $(T\chi_{[-R,R]}, h_I) \to 0$ as $R \to \infty$, we get

$$|(Th_I, h_J)| \le Const|J|^{1+\varepsilon}\rho(J,I)^{-\varepsilon}.$$

Since surely $\rho(J,I) \le |I|$ in the case under consideration, the factor $|I|\big[|I| + \rho(J,I)\big]^{-1}$ changes nothing.

For the remaining case when $\rho(J,I) < |J|$, we will be completely satisfied with trivial L^2-estimate

$$|(Th_I, h_J)| \le \|T\|_{L^2\to L^2}\|h_I\|_{L^2}\|h_J\|_{L^2} \le Const|I|^{1/2}|J|^{1/2}.$$

□

12.2. Regular and singular pairs

Let $|J| \le |I|$. We will call the pair (I,J) regular if

$$\frac{|J|}{\rho(J,I)} \le \sqrt{\frac{|J|}{|I|}}$$

and singular otherwise.

Lemma 12.4. *For any regular pair (I,J) we have*

$$|(Th_I, h_J)| \le C|I|^{1+\varepsilon/2}|J|^{1+\varepsilon/2}d(I,J)^{-1-\varepsilon}$$

where $d(I,J) = |I| + |J| + \operatorname{dist}(I,J)$ is the "long distance" between the intervals I and J.

Proof. Note that since $|J| \le |I|$, we have $\rho(J,I) \ge |J|$ and therefore the estimate of Lemma 12.3 is applicable.

If $J \cap (3I) = \varnothing$, then $|I| \le \rho(J,I) \le d(I,J) \le 3\rho(J,I)$, and we immediately get

$$|(Th_I, h_J)| \le C|J|^{1+\varepsilon}\rho(J,I)^{-\varepsilon}\frac{|I|}{|I|+\rho(J,I)} \le 3^{1+\varepsilon}C\frac{|J|^{1+\varepsilon}|I|}{d(I,J)^{1+\varepsilon}} \le 3^{1+\varepsilon}C\frac{|I|^{1+\varepsilon/2}|J|^{1+\varepsilon/2}}{d(I,J)^{1+\varepsilon}}.$$

Assume now that $J \cap (3I) \ne \varnothing$. Then $\rho(J,I) \le |I| \le d(I,J) \le 3|I|$ and thereby

$$\begin{aligned}|(Th_I, h_J)| &\le C|J|^{1+\varepsilon}\rho(J,I)^{-\varepsilon}\frac{|I|}{|I|+\rho(J,I)} = C|J|\left(\frac{|J|}{\rho(J,I)}\right)^{\varepsilon}\frac{|I|}{|I|+\rho(J,I)} \\ &\le C|J|\left(\frac{|J|}{|I|}\right)^{\varepsilon/2} \le 3^{1+\varepsilon}C|J|\left(\frac{|J|}{d(I,J)}\right)^{\varepsilon/2}\left(\frac{|I|}{d(I,J)}\right)^{1+\varepsilon/2} = 3^{1+\varepsilon}C\frac{|I|^{1+\varepsilon/2}|J|^{1+\varepsilon/2}}{d(I,J)^{1+\varepsilon}}.\end{aligned}$$

□

12.2.1. Singular pairs

Note that the definition of a singular pair can be rewritten as

$$\rho(J,I) < \sqrt{|I||J|}.$$

Since $\sqrt{|I||J|} \le |I|$, for any singular pair one has $J \cap (3I) \ne \varnothing$ and therefore $|I| \le d(I,J) \le 3|I|$.

Let us fix a positive integer r_0 to be chosen later. If $|J| \ge 2^{-r_0}|I|$, then just from the trivial L^2-estimate we have

$$\begin{aligned}|(Th_I, h_J)| &\le C|I|^{1/2}|J|^{1/2} \\ &= C\frac{|I|^{1+\varepsilon/2}|J|^{1+\varepsilon/2}}{d(I,J)^{1+\varepsilon}}\frac{d(I,J)^{1+\varepsilon}}{|I|^{(1+\varepsilon)/2}|J|^{(1+\varepsilon)/2}} \le 3^{1+\varepsilon}2^{(1+\varepsilon)r_0/2}C\frac{|I|^{1+\varepsilon/2}|J|^{1+\varepsilon/2}}{d(I,J)^{1+\varepsilon}},\end{aligned}$$

i.e. essentially the same estimate as for regular pairs (just with larger constant). Observing that our estimate is symmetric with respect to I and J, we get the following

Theorem 12.5. *For every $r_0 \in \mathbb{N}$ there exists a constant $C = C(r_0, T)$ such that*

$$|(Th_I, h_J)| \le C\frac{|I|^{1+\varepsilon/2}|J|^{1+\varepsilon/2}}{d(I,J)^{1+\varepsilon}}$$

unless the pair (I, J) is "essentially singular" which means that

$$\text{either} \quad |J| \le 2^{-r_0}|I| \quad \text{and} \quad \rho(J,I) \le \sqrt{|I||J|},$$

$$\text{or} \quad |I| \le 2^{-r_0}|J| \quad \text{and} \quad \rho(I,J) \le \sqrt{|I||J|}.$$

13. Classical Calderon-Zygmund operators in matrix weighted L^p

Though we assumed T to act in *scalar* non-weighted L^2, it can be extended onto E-valued non-weighted L^2 by the formula

$$T(\varphi(\cdot)e) = (T\varphi)(\cdot)e, \qquad \varphi \in (\text{scalar})L^2,\ e \in E$$

(clearly, "simple functions" $\varphi(\cdot)e$ span the whole E-valued L^2, the definition is correct, and, as it can be easily seen from a decomposition over some orthonormal basis, the norm $\|T\|_{L^2\to L^2}$ remains the same as in the scalar case).

To prove that T acts in $L^p(W)$ is (due to our definition) the same as to prove the estimate

$$|(T\varphi, \psi)| \le C\|\varphi\|_{L^p(W)}\|\psi\|_{L^q(W^{-q/p})}$$

for any $\varphi \in L^2 \cap L^p(W)$, $\psi \in L^2 \cap L^q(W^{-q/p})$ (clearly, for such pairs φ, ψ both sides of the inequality make sense). We have

$$(T\varphi, \psi) = \sum_{I,J\in\mathcal{D}} (Th_I, h_J)\big(c_I(\varphi), c_J(\psi)\big)_E\,.$$

(for any $\varphi, \psi \in L^2$ the series on the right converges "over rectangles", say) We will show in the next section that this series converges absolutely and can be estimated from above by the product of the norms $\|\varphi\|_{L^p(W)}\|\psi\|_{L^q(W^{-q/p})}$ when we have no essentially singular pair on the right, i.e. when for any essentially singular pair (I, J) either $c_I(\varphi)$ or $c_J(\psi) = 0$. And now our task is to reduce the general case to this particular one.

13.1. Singularity removal

The point is that singular pairs are extremely rare: if you pick up an interval $I \subset \mathbb{R}$ and choose at random an interval $J \supset I$ of length $|J| = 2^r|I|$ where r is some large positive integer, the probability that the pair (I, J) is singular is about $2^{-r/2}$. Unfortunately, intervals of *one fixed* dyadic lattice do not know any probability: they just lie as they lie. To solve the problem, we will use *two random* lattices: one for φ and another one for ψ. Note that, when estimating (Th_I, h_J), we have never assumed anything special about intervals I and J and that the formula

$$(T\varphi, \psi) = \sum_{I \in \mathcal{D}, J \in \mathcal{D}'} (Th_I, h_J)\big(c_I(\varphi), c_J(\psi)\big)_E$$

is still correct.

13.1.1. Average lattice: construction

Our random lattice will contain the intervals of standard length 2^k ($k \in \mathbb{Z}$), but will be "randomly shifted" with respect to the standard lattice $\mathcal{D}_0$ consisting of the intervals $[\ell 2^k, (\ell+1)2^k]$ ($\ell, k \in \mathbb{Z}$). The simplest idea would be to pick up a random variable ξ equidistributed over $\mathbb{R}$ and to define the random lattice as $\xi + \mathcal{D}_0$. Unfortunately, there exists no such ξ and we have to act in a little bit more sophisticated way.

Let Ω be some probability space and let $x(\omega)$ be a random variable equidistributed over $(0, 1)$. Let $\xi_j(\omega)$ be random variables satisfying $\mathbb{P}\{\xi_j = +1\} = \mathbb{P}\{\xi_j = -1\} = 1/2$. Assume also that $x(\omega), \xi_j(\omega)$ are independent. Define the random lattice $\mathcal{D}(\omega)$ as follows:

1) Let $I_0(\omega) = [x(\omega) - 1, x(\omega)] \in \mathcal{D}(\omega)$. This uniquely determines all intervals in $\mathcal{D}(\omega)$ of length 2^k where $k \leq 0$.

2) The intervals $I_k(\omega) \in \mathcal{D}(\omega)$ of length 2^k with $k > 0$ are determined inductively: if $I_{k-1}(\omega) \in \mathcal{D}$ is already chosen, $I_k(\omega)$ is determined by the following rule: $(I_k(\omega))_+ = I_{k-1}(\omega)$ if $\xi_k(\omega) = +1$ and $(I_k(\omega))_- = I_{k-1}(\omega)$ if $\xi_k(\omega) = -1$. By other words, at every step we extend the interval $I_{k-1}(\omega)$ to the left if $\xi_k(\omega) = +1$ and to the right otherwise. Clearly, to know one interval of length 2^k in the lattice is enough to determine all of them.

It is easy to check that the random lattice $\mathcal{D}(\omega)$ constructed in this way satisfies the following

Equidistribution property. Let $I \subset \mathbb{R}$ be any interval, $k \in \mathbb{Z}$. The probability that I contains an endpoint or the middle of some interval $J \in \mathcal{D}(\omega)$ of length 2^k is $\min\{1, 2^{-(k-1)}|I|\}$.

13.1.2. Bad intervals

Let $\mathcal{D}(\omega)$ and $\mathcal{D}'(\omega')$ ($(\omega, \omega') \in \Omega \times \Omega$) be two independent random dyadic lattices. We will call an interval $I \in \mathcal{D}(\omega)$ bad if there exists an interval $J \in \mathcal{D}'(\omega')$ of length $|J| \geq |I|$ such that the pair (I, J) is essentially singular. The definition of bad intervals in $\mathcal{D}'(\omega')$ is the same (now we look for a bigger interval in $\mathcal{D}(\omega)$).

Lemma 13.1. *For any fixed ω and an interval $I \in \mathcal{D}(\omega)$ we have*

$$\mathbb{P}_{\omega'}\{I \textit{ is bad }\} \le 16 \cdot 2^{-r_0/2}.$$

Proof. Recall that the pair (I, J) satisfying $|I| \le |J|$ is essentially singular if $|J| > 2^{r_0}|I|$ and $\rho(I, J) \le \sqrt{|I||J|}$. Let $r > r_0$, $J \in \mathcal{D}'(\omega')$, $|J| = 2^r|I|$. The pair (I, J) is essentially singular if either one of the ends or the middlepoint of J lies in the interval $(2 \cdot 2^{r/2} + 1)J$. According to the equidistribution property above, the probability that $\mathcal{D}'(\omega')$ contains such an interval J is $2^{-(r-1)}(2 \cdot 2^{r/2} + 1) \le 8 \cdot 2^{-r/2}$, say. Considering all possible $r > r_0$, we obtain

$$\mathbb{P}_{\omega'}\{I \text{ is bad }\} \le 8 \sum_{r > r_0} 2^{-r/2} \le 16 \cdot 2^{-r_0/2}.$$

□

13.1.3. Splitting of a function into "bad" and "good" parts

Let $\varphi \in L^2 \cap L^p(W)$ where $W \in (\mathcal{A}_{p,q})$. Then, since the Haar system is a strong unconditional basis in $L^p(W)$, for every fixed $\omega \in \Omega$ we have

$$\varphi = \sum_{I \in \mathcal{D}(\omega)} c_I(\varphi) h_I$$

and

$$\frac{1}{A} \int_{\mathbb{R}} \Big(\sum_{I \in \mathcal{D}(\omega)} \langle \rho \rangle_{I,p} (c_I(\varphi))^2 \chi_I(t) \Big)^{p/2} dt \le \|\varphi\|^p_{L^p(W)} \le A \int_{\mathbb{R}} \Big(\sum_{I \in \mathcal{D}(\omega)} \langle \rho \rangle_{I,p} (c_I(\varphi))^2 \chi_I(t) \Big)^{p/2} dt$$

where A is some large constant depending on that in the $(\mathcal{A}_{p,q})$-condition only. For every $\omega' \in \Omega$ split

$$\varphi = \sum_{I \in \mathcal{D}(\omega)} c_I(\varphi) h_I = \sum_{I \in \mathcal{D}(\omega), I \text{ is good}} c_I(\varphi) h_I + \sum_{I \in \mathcal{D}(\omega), I \text{ is bad}} c_I(\varphi) h_I = \varphi_{\text{good}} + \varphi_{\text{bad}}.$$

We will use here very rough averaging trick. We will just show that in average $\|\varphi_{\text{bad}}\|_{L^p(W)}$ is only a small part of the whole norm $\|\varphi\|_{L^p(W)}$. For that we will need the following simple lemma.

Lemma 13.2. *Let $a_j > 0$, $\sum_j a_j < +\infty$. Let ξ_j be any (not necessarily independent) random variables satisfying $\mathbb{P}\{\xi_j = 1\} = \delta$, $\mathbb{P}\{\xi_j = 0\} = 1 - \delta$. Then for any $\alpha > 0$*

$$\mathbb{E}\Big(\sum_j a_j \xi_j \Big)^\alpha \le \delta^{\min\{\alpha, 1\}} \Big(\sum_j a_j \Big)^\alpha.$$

Proof. For $\alpha = 1$ we obviously have $\mathbb{E}(\sum_j a_j \xi_j) = \delta \sum_j a_j$. Now the estimate for $\alpha < 1$ follows from the Holder inequality. For $\alpha > 1$ we obtain

$$\mathbb{E}\Big(\sum_j a_j \xi_j \Big)^\alpha \le \mathbb{E}\Big(\sum_j a_j \xi_j \Big) \Big(\sum_j a_j \Big)^{\alpha - 1} = \delta \Big(\sum_j a_j \Big)^\alpha.$$

We have

$$\|\varphi_{\mathrm{bad}}\|^p_{L^p(W)} \asymp \int_{\mathbb{R}} \Big(\sum_{I\in\mathcal{D}(\omega),\ \mathrm{I\ is\ bad}} \langle\rho\rangle_{I,p} (c_I(\varphi))^2 \chi_I(t) \Big)^{p/2} dt.$$

Note that for every $t \in \mathbb{R}$ in this integral we have exactly the situation described by the lemma: each interval $I \in \mathcal{D}(\omega)$ containing the point t is bad with probability not greater than $16 \cdot 2^{-r_0/2}$. Therefore for any $t \in \mathbb{R}$

$$\mathbb{E}_{\omega'} \Big(\sum_{I\in\mathcal{D}(\omega),\ \mathrm{I\ is\ bad}} \langle\rho\rangle_{I,p} (c_I(\varphi))^2 \chi_I(t) \Big)^{p/2} \le \big[16 \cdot 2^{-r_0/2}\big]^{\min\{p/2,1\}} \Big(\sum_{I\in\mathcal{D}(\omega)} \langle\rho\rangle_{I,p} (c_I(\varphi))^2 \chi_I(t) \Big)^{p/2}$$

and the same estimate is true for the whole integrals. Thus,

$$\mathbb{E}_{\omega'} \|\varphi_{\mathrm{bad}}\|^p_{L^p(W)} \le \big[16 \cdot 2^{-r_0/2}\big]^{\min\{p/2,1\}} A^2 \|\varphi\|^p_{L^p(W)}$$

and, since it is true for every $\omega \in \Omega$, the same estimate holds for $\mathbb{E}\|\varphi_{\mathrm{bad}}\|^p_{L^p(W)}$. □

Applying the same reasoning to ψ (now we fix ω') and also to the norms in non-weighted L^2 we finally get the following theorem

Theorem 13.3. *For every $\delta > 0$ one can choose $r_0 \in \mathbb{N}$ such that for any $\varphi \in L^2 \cap L^p(W)$, $\psi \in L^2 \cap L^q(W^{-q/p})$ one can find two dyadic lattices $\mathcal{D}$ and $\mathcal{D}'$ such that*

$$\|\varphi_{\mathrm{bad}}\|_{L^p(W)} \le \delta \|\varphi\|_{L^p(W)}, \qquad \|\varphi_{\mathrm{bad}}\|_{L^2} \le \delta \|\varphi\|_{L^2};$$

$$\|\psi_{\mathrm{bad}}\|_{L^q(W^{-q/p})} \le \delta \|\psi\|_{L^p(W^{-q/p})}, \qquad \|\psi_{\mathrm{bad}}\|_{L^2} \le \delta \|\psi\|_{L^2}.$$

(just because it is true for most pairs of lattices $\mathcal{D}(\omega)$ and $\mathcal{D}(\omega')$).

When $\delta < 1$, we also automatically have $\|\varphi_{\mathrm{good}}\|_{L^p(W)} \le 2\|\varphi\|_{L^p(W)}$ etc.

13.1.4. Pulling ourselves by hair

Note that in the right hand part of the identity

$$(T\varphi_{\mathrm{good}}, \psi_{\mathrm{good}}) = \sum_{I\in\mathcal{D}, J\in\mathcal{D}'} (Th_I, h_J) \big(c_I(\varphi_{\mathrm{good}}), c_J(\psi_{\mathrm{good}})\big)_E$$

there is no essentially singular pair with non-zero contribution. As we will see in the next section, this yields the estimate

$$|(T\varphi_{\mathrm{good}}, \psi_{\mathrm{good}})| \le C \|\varphi_{\mathrm{good}}\|_{L^p(W)} \|\psi_{\mathrm{good}}\|_{L^p(W^{-q/p})}$$

with some constant $C > 0$ depending on the constant in the $(\mathcal{A}_{p,q})$ condition, the constants in the definition of Calderon-Zygmund operator and, at last, r_0.

Fix some large $R > 0$ and define

$$M(R) = \sup\big\{ |(T\varphi, \psi)| : \|\varphi\|_{L^p(W)},\ \|\psi\|_{L^q(W^{-q/p})} \le 1;\ \|\varphi\|_{L^2},\ \|\psi\|_{L^2} \le R \big\}.$$

Clearly $M(R) \le \|T\|_{L^2\to L^2} R^2 < \infty$. Let φ, ψ be two functions satisfying the conditions in the definition of $M(R)$ for which $|(T\varphi,\psi)| \ge M(R)/2$, say. Using the lattices delivered by Theorem 13.3, we get

$$|(T\varphi,\psi)| \le |(T\varphi_{\text{good}},\psi)| + |(T\varphi_{\text{bad}},\psi)| \le |(T\varphi_{\text{good}},\psi_{\text{good}})| + |(T\varphi_{\text{good}},\psi_{\text{bad}})| + |(T\varphi_{\text{bad}},\psi)|\,.$$

We have

$$|(T\varphi_{\text{good}},\psi_{\text{good}})| \le C\|\varphi_{\text{good}}\|_{L^p(W)}\|\psi_{\text{good}}\|_{L^p(W^{-q/p})} \le 4C\|\varphi\|_{L^p(W)}\|\psi\|_{L^p(W^{-q/p})} \le 4C;$$

$$|(T\varphi_{\text{good}},\psi_{\text{bad}})| \le 2M(R)\delta$$

(because $2^{-1}\varphi_{\text{good}}$ and $\delta^{-1}\psi_{\text{bad}}$ satisfy the conditions in the definition of $M(R)$) and, at last,

$$|(T\varphi_{\text{bad}},\psi)| \le M(R)\delta$$

(because $\delta^{-1}\varphi_{\text{bad}}$ and ψ satisfy the conditions in the definition of $M(R)$)

So,

$$M(R)/2 \le 4C + 3\delta M(R).$$

Taking $\delta = 1/12$, we get from here $M(R) \le 16C$. Since the estimate does not depend on R, we are done (modulo the "regular" case, of course).

The next section deals with the "regular part" of the matrix.

14. The regular part of the matrix

In this section we will prove that the regular part of the matrix of a classical Calderon–Zygmund operator defines a bounded operator in $L^p(W)$, provided that the weight W satisfies the Muckenhoupt $(\mathcal{A}_{p,q})$ condition.

The estimate of the matrix is relatively simple.

Recall that the "long distance" d between intervals I and J is defined by $d(I,J) := \operatorname{dist}(I,J) + |I| + |J|$. Fix two dyadic lattices $\mathcal{D}$ and $\mathcal{D}'$. For technical convenience we will again assume that the lengths of intervals in both lattices are $2^k,\ k\in\mathbb{Z}$.

Theorem 14.1. *Let a matrix $\{T_{I,J}\}_{I\in\mathcal{D},\mathcal{J}\in\mathcal{D}'}$ satisfy*

$$|T_{I,J}| \le C|I|^{1+\varepsilon/2}|J|^{1+\varepsilon/2} d(I,J)^{-(1+\varepsilon)}\,.$$

Then, if a weight W satisfies the Muckenhoupt $(A_{p,q})$ condition,

$$\sum_{I\in\mathcal{D},J\in\mathcal{D}'} |T_{I,J}|\cdot\left|\big(c_I(\varphi),c_J(\psi)\big)_E\right| \le Const\|\varphi\|_{L^p(W)}\|\psi\|_{L^q(W^{-q/p})}$$

Proof. Since everything is symmetric with respect to the exchange of W, p and $W^{-q/p}$, q, it is enough to consider only half of the matrix, i.e. to take the sum only over $|I| \le |J|$.

Let us "slice" the matrix $\{T_{I,J}\}_{I\in\mathcal{D},J\in\mathcal{D}'}$. For every $r=0,1,2,\ldots$ define the matrix $\{T^{(r)}_{I,J}\}_{I\in\mathcal{D},J\in\mathcal{D}'}$ by putting

$$T^{(r)}_{I,J}=\begin{cases} T_{I,J}, & |I|=2^{-r}|J|\ ;\\ 0, & \text{otherwise}\ .\end{cases}$$

If we are able to show that the norms of $T^{(r)}$ decrease as a geometric progression, i.e.

$$\sum_{I\in\mathcal{D},J\in\mathcal{D}'}|T^{(r)}_{I,J}|\cdot|\big(c_I(\varphi),c_J(\psi)\big)_E|\le 2^{-r\alpha}C\cdot\|\varphi\|_{L^p(W)}\|\psi\|_{L^q(W^{-q/p})}$$

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

To demonstrate the reader the main idea of the proof not obstacled by technical details, we will start with the simplest case of the scalar weighted L^2 and then will gradually advance to the general case.

14.1. Scalar L^2 case.

In this case we have $|\big(c_I(\varphi),c_J(\psi)\big)_E|=|c_I(\varphi)|\cdot|c_J(\psi)|$

Recall that the Haar system is a strong unconditional basis in $L^2(w)$, and therefore the norms $\|\varphi\|_{L^2(w)}$ and $\|\psi\|_{L^2(w^{-1})}$ are equivalent to

$$\Big(\sum_{I\in\mathcal{D}}|c_I(\varphi)|^2\langle w\rangle_I\cdot|I|\Big)^{1/2}\quad\text{and}\quad\Big(\sum_{J\in\mathcal{D}'}|c_J(\psi)|^2\langle w^{-1}\rangle_J\cdot|J|\Big)^{1/2}\quad\text{respectively.}$$

Thus our task is to show that for any two sequences $\{c_I\}_{I\in\mathcal{D}}$ and $\{c'_J\}_{J\in\mathcal{D}'}$

$$\sum_{I\in\mathcal{D},J\in\mathcal{D}'}|T^{(r)}_{I,J}|\cdot|c_I|\cdot|c'_J|\le 2^{-r\alpha}C\cdot\Big(\sum_{I\in\mathcal{D}}|c_I|^2\langle w\rangle_I\cdot|I|\Big)^{1/2}\Big(\sum_{J\in\mathcal{D}'}|c'_J|^2\langle w^{-1}\rangle_J\cdot|J|\Big)^{1/2}.$$

It is convenient to give another representation for the norm $\big(\sum_{I\in\mathcal{D}}|c_I|^2\langle w\rangle_I\cdot|I|\big)^{1/2}$. Consider functions f_k,

$$f_k=\sum_{I\in\mathcal{D}\,:\,|I|=2^k}c_I\chi_I\,.$$

Clearly

$$\sum_{I\in\mathcal{D}}|c_I|^2\langle w\rangle_I\cdot|I|=\sum_{k=-\infty}^{\infty}\|f_k\|^2_{L^2(w)},$$

so we get a natural isometric representation of the coefficient space as a subspace X^0 of the space $X=\oplus_{k=-\infty}^{\infty}X_k=\oplus_{k=-\infty}^{\infty}L^2(w)$. The subspace X^0 can be identified with the set of all sequences $\{f_k\}_{k=-\infty}^{\infty}\in X$ such that each f_k is constant on every dyadic interval of length 2^k in $\mathcal{D}$.

Let $\{g_k\}_{k=-\infty}^{\infty} \in X^* = \oplus_k X_k^* = \oplus_k L^2(w^{-1})$ where

$$g_k = \sum_{J\in\mathcal{D}' : |J|=2^k} c'_J \chi_J$$

be the analogous representation of a coefficient sequence $\{c'_J\}_{J\in\mathcal{D}'}$. Then

$$\sum_{I\in\mathcal{D},J\in\mathcal{D}'} |T^{(r)}_{I,J}| \cdot |c_I| \cdot |c_J| = \sum_{k=-\infty}^{\infty} \iint_{\mathbb{R}^2} K_k^{(r)}(s,t) \cdot |f_{k-r}(s)| \cdot |g_k(t)| ds dt \,,$$

where the kernels $K_k^{(r)}(s,t)$ are defined by

$$K_k^{(r)}(s,t) = \sum_{\substack{I\in\mathcal{D}:|I|=2^{k-r} \\ J\in\mathcal{D}':|J|=2^k}} |T^{(r)}_{I,J}| \cdot |I|^{-1} \cdot |J|^{-1} \chi_I(s)\chi_J(t) \,.$$

Note that "layers" with different k do not interfere, so it is enough to estimate each integral in the sum separately, i.e. to show that

$$\iint_{\mathbb{R}^2} K_k^{(r)}(s,t) \cdot |f_{k-r}(s)| \cdot |g_k(t)| ds dt \le 2^{-r\alpha} C \|f_{k-r}\|_{L^2(w)} \|g_k\|_{L^2(w^{-1})} \,.$$

Clearly the kernels $K_k^{(r)}(s,t)$ can be estimated as

$$K_k^{(r)}(s,t) \le A \cdot 2^{-r\varepsilon/2} \cdot \frac{2^{k\varepsilon}}{\big(2^k + |t-s|\big)^{1+\varepsilon}} \,,$$

where A is some absolute constant.

So, if we prove that the convolution operators $f \mapsto \mathcal{K}_k * f$, where

$$\mathcal{K}_k(s) = \frac{2^{k\varepsilon}}{\big(2^k + |s|\big)^{1+\varepsilon}} \,,$$

are uniformly bounded in $L^2(w)$, we are done.

The uniform boundedness of the above operators is almost obvious. Take any $j \in \mathbb{N}$ and consider a splitting $\mathcal{C}$ of the real line $\mathbb{R}$ into intervals of length 2^j. For any interval I in the splitting the norm of the operator $f \to \langle f\rangle_I \chi_I$ in $L^2(w)$ does not exceed $\|w\|_{A_2}$ according to Lemma 2.1. Since the intervals in the splitting are mutually disjoint, we conclude that the norms of averaging operators $A_{\mathcal{C}} f := \sum_{I\in\mathcal{C}} \langle f\rangle_I \cdot \chi_I$ in $L^2(w)$ are bounded by the same constant $\|w\|_{A_2}$. Now take the average of $A_{\mathcal{C}}$ over all (real) shifts of the splitting $\mathcal{C}$. We get the convolution with the function τ_j (see Figure 8).

So, the operators $f \mapsto f * \tau_j$ are uniformly bounded in $L^2(w)$. Since $\tau_{j+1}(t) \ge 2^{-j-2}\chi_{[-2^j,2^j]}$, one can conclude that the operators $f \mapsto 2^{-j} f * \chi_{[-2^j,2^j]}$ are also uniformly bounded in $L^2(w)$.

It is easy to see that

$$\mathcal{K}_k(t) \le C \sum_{j=0}^{\infty} 2^{-j\varepsilon} \cdot 2^{-j-k} \cdot \chi_{[-2^{j+k},2^{j+k}]}(t) \,.$$

Since $\sum_{j=0}^{\infty} 2^{-j\varepsilon} = (1-2^{-\varepsilon})^{-1} < \infty$ it follows immediately that the convolutions with $\mathcal{K}_k$ are uniformly bounded in $L^2(w)$. $\square$

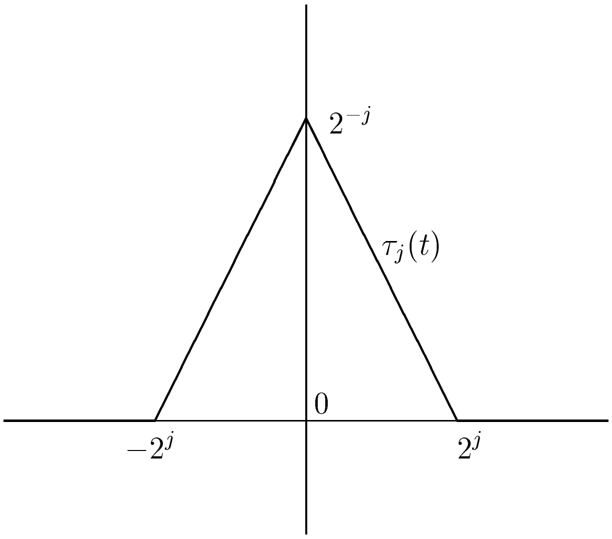

Figure 8: Function τ_j

14.2. Vector L^2 case.

Consider now the case of E-valued weighted L^2. The inequality to prove is now

$$\sum_{I\in\mathcal{D},J\in\mathcal{D}'} |T^{(r)}_{I,J}|\left|\big(c_I(\varphi), c_J(\psi)\big)_E\right| \le 2^{-r\alpha}C\cdot \|\varphi\|_{L^2(W)}\|\psi\|_{L^2(W^{-1})}$$

Since the Haar system is a strong unconditional basis in $L^2(W)$, the norms $\|\varphi\|_{L^2(W)}$ and $\|\psi\|_{L^2(W^{-1})}$ are equivalent to

$$\Big(\sum_{I\in\mathcal{D}}\big\|\langle W\rangle_I^{1/2}c_I(\varphi)\big\|_E^2\cdot|I|\Big)^{1/2} \quad\text{and}\quad \Big(\sum_{J\in\mathcal{D}'}\big\|\langle W^{-1}\rangle_J^{1/2}c_J(\psi)\big\|_E^2\cdot|J|\Big)^{1/2} \quad\text{respectively.}$$

Thus we need to show that for any two sequences of vectors $\{c_I\}_{I\in\mathcal{D}}$ and $\{c'_J\}_{J\in\mathcal{D}'}$ in E

$$\sum_{I\in\mathcal{D},J\in\mathcal{D}'} |T^{(r)}_{I,J}|\cdot\left|\big(c_I, c'_J\big)_E\right| \le 2^{-r\alpha}C\cdot\Big(\sum_{I\in\mathcal{D}}\big\|\langle W\rangle_I^{1/2}c_I\big\|^2\cdot|I|\Big)^{1/2}\Big(\sum_{J\in\mathcal{D}'}\big\|\langle W^{-1}\rangle_J^{1/2}c'_J\big\|^2\cdot|J|\Big)^{1/2}.$$

We will again associate with the sequence $\{c_I\}_{I\in\mathcal{D}}$ the sequence of functions

$$f_k = \sum_{I\in\mathcal{D},\,|I|=2^k} c_I\chi_I\,.$$

Clearly,

$$\sum_{I\in\mathcal{D}} \left\| \langle W\rangle_I^{1/2} c_I \right\|_E^2 \cdot |I| = \sum_{k=-\infty}^{\infty} \|f_k\|^2_{L^2(W)} .$$

So we again get a natural isometric representation of the coefficient space as a subspace $X^0 \subset X = \oplus_{k=-\infty}^{\infty} X_k = \oplus_{k=-\infty}^{\infty} L^2(W)$ consisting of all sequences $\{f_k\}_{k=-\infty}^{\infty}$ such that each function f_k is constant on dyadic intervals in $\mathcal{D}$ of length 2^k.

Similarly, putting

$$g_k = \sum_{J\in\mathcal{D}',\, |J|=2^k} c'_J \chi_J ,$$

we can identify the coefficient space of sequences $\{c'_J\}_{J\in\mathcal{D}'}$ with the subspace of $X^* = \oplus_{k=-\infty}^{\infty} X_k^* = \oplus_{k=-\infty}^{\infty} L^2(W^{-1})$ consisting of those sequences $\{g_k\}_{k=-\infty}^{\infty}$ for which each function g_k is constant on dyadic intervals in $\mathcal{D}'$ of length 2^k.

In this representation we can rewrite the estimate we need as

$$\sum_{I\in\mathcal{D}, J\in\mathcal{D}'} |T^{(r)}_{I,J}| \cdot \left| \left(c_I, c'_J \right)_E \right| =$$

$$= \sum_{k=-\infty}^{\infty} \iint_{\mathbb{R}^2} K_k^{(r)}(s,t) \cdot \left| \left(f_{k-r}(s), g_k(t) \right)_E \right| ds dt \le 2^{-r\alpha} C \|\{f_k\}\|_X \|\{g_k\}\|_{X^*} ,$$

with the same kernels $K_k^{(r)}(s,t)$ as before. Again "layers" with different k do not interfere, so it is enough to estimate each integral in the sum separately.

Recall that the kernels $K_k^{(r)}(s,t)$ can be estimated as

$$K_k^{(r)}(s,t) \le A \cdot 2^{-r\varepsilon/2} \cdot \frac{2^{k\varepsilon}}{(2^k + |t-s|)^{1+\varepsilon}} ,$$

where A is some absolute constant. So, if we prove that for the kernels $\mathcal{K}_k$,

$$\mathcal{K}_k(s,t) = \frac{2^{k\varepsilon}}{(2^k + |t-s|)^{1+\varepsilon}} ,$$

we have the uniform estimate

$$\iint_{\mathbb{R}^2} \mathcal{K}_k(s,t) \left| \left(f(s), g(t) \right)_E \right| ds dt \le C \|f\|_{L^2(W)} \|g\|_{L^2(W^{-1})}$$

for all $f \in L^2(W)$, $g \in L^2(W^{-1})$, then we are done.

So far we were just repeating word by word the reasoning for the scalar case. But now we should deviate from it. In that proof we used the majoration property for positive kernels, namely the fact that if an integral operator with a positive kernel is bounded in (a scalar) $L^2(w)$, then an integral operator with any smaller kernel is bounded as well and its norm can be estimated from above by the norm of the first operator. Unfortunately, the similar statement for general matrix-valued weights is not true, and we need to introduce a new step. Namely, replacing f by $W^{-1/2}f$ and g by $W^{1/2}g$ we can see that the estimate we need immediately follows from

$$\iint_{\mathbb{R}^2} \mathcal{K}_k(s,t)\|W^{-1/2}(s)W^{1/2}(t)\| \cdot \|f(s)\|_E \|g(t)\|_E dsdt \le C\|f\|_{L^2}\|g\|_{L^2} \quad \forall f, g \in L^2 .$$

And this inequality is clearly scalar: it includes only scalar functions $\|f(\cdot)\|_E$ and $\|g(\cdot)\|_E$. So it is enough to show that all integral operators with kernels $\mathcal{K}_k(s,t)\|W^{-1/2}(s)W^{1/2}(t)\|$ are uniformly bounded in the non-weighted scalar L^2.

Fix $j \in \mathbb{Z}$. For a splitting $\mathcal{C}$ of $\mathbb{R}$ into intervals of the length 2^j consider the "averaging" operator

$$(A_{\mathcal{C}}f)(t) = \sum_{I \in \mathcal{C}} \frac{1}{|I|} \int_I \|W^{-1/2}(s)W^{1/2}(t)\|f(s)ds , \qquad f \in L^2 .$$

Let us show that the operators $A_{\mathcal{C}}$ are uniformly bounded. Indeed the operators are block diagonal, so it is enough to estimate each block (integration over each interval $I \in \mathcal{C}$) separately. The norm of each block can be estimated by its Hilbert–Schmidt norm

$$\Big(\frac{1}{|I|^2} \iint_{I \times I} \|W^{-1/2}(s)W^{1/2}(t)\|^2 dsdt\Big)^{1/2} .$$

Note that for every two operators $A, B \in \mathcal{L}_+(E)$

$$\|A^{1/2}B^{1/2}\|^2 = \|A^{1/2}BA^{1/2}\| \le \operatorname{Tr}(A^{1/2}BA^{1/2}) = \operatorname{Tr}(AB) \le n\|A^{1/2}B^{1/2}\|^2 .$$

So, since trace is a linear functional,

$$\begin{aligned}\frac{1}{|I|^2} \iint_{I \times I} \|W^{-1/2}(s)W^{1/2}(t)\|^2 dsdt &\le \frac{1}{|I|^2} \iint_{I \times I} \operatorname{Tr}\big\{W^{-1}(s)W(t)\big\} dsdt \\ &= \operatorname{Tr}\big\{\langle W^{-1}\rangle_I \langle W\rangle_I\big\} \le n\|\langle W^{-1}\rangle_I^{1/2}\langle W\rangle_I^{1/2}\|^2 \le n\|W\|_{\mathcal{A}_{2,2}} < +\infty ,\end{aligned}$$

proving the statement.

Now take the average of the operators $A_{\mathcal{C}}$ over all shifts of the grid $\mathcal{C}$. We get the integral operator with kernel

$$\tau_j(s-t)\|W^{-1/2}(s)W^{1/2}(t)\| ,$$

where the function τ_j is as defined on Figure 8. Since $\tau_{j+1} \ge 2^{-j-2}\chi_{[-2^j, 2^j]}$, the integral operators with kernels $2^{-j}\chi_{[-2^j,2^j]}(s-t)\|W^{-1/2}(s)W^{1/2}(t)\|$ are uniformly bounded as well.

Finally, the inequalities

$$\frac{2^{k\varepsilon}}{(2^k + |s|)^{1+\varepsilon}} \le C \sum_{j=0}^{\infty} 2^{-j\varepsilon} \cdot 2^{-j-k} \cdot \chi_{[-2^{j+k}, 2^{j+k}]}(s)$$

imply that the integral operators with kernels

$$\frac{2^{k\varepsilon}}{(2^k + |s-t|)^{1+\varepsilon}} \cdot \|W^{-1/2}(s)W^{1/2}(t)\|$$

are uniformly bounded in L^2.

14.3. Vector L^p case.

In this section we will prove Theorem 14.1 in full generality. Recall that the inequality to prove is

$$\sum_{I\in\mathcal{D},J\in\mathcal{D}'}|T_{I,J}^{(r)}|\left|\big(c_I(\varphi),c_J(\psi)\big)_E\right|\le 2^{-r\alpha}C\cdot\|\varphi\|_{L^p(W)}\|\psi\|_{L^q(W^{-q/p})}$$

Though we know that

$$\|\varphi\|_{L^p(W)}\asymp\Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\langle\rho\rangle_{I,p}(c_I(\varphi))^2\chi_I(t)\Big)^{p/2}dt\Big\}^{1/p},$$

this representation of the norm is not convenient for us. Instead we will observe that since the Haar system is a strong unconditional basis in $L^p(W)$, for every function $\varphi=\sum_{I\in\mathcal{D}}c_I(\varphi)h_I\in L^p(W)$ and for any sequence of complex numbers $\{\varepsilon_I\}_{I\in\mathcal{D}}$ such that $|\varepsilon_I|=1$ for every $I\in\mathcal{D}$ the norm

$$\Big\|\sum_I\varepsilon_Ic_I(\varphi)h_I(\cdot)\Big\|_{L^p(W)}=\Big\{\int_{\mathbb{R}}\Big\|\sum_{I\in\mathcal{D}}\varepsilon_IW(t)^{1/p}c_I(\varphi)h_I(t)\Big\|_E^p dt\Big\}^{1/p}$$

is comparable to $\|\varphi\|_{L^p(W)}$. Taking the average over all possible sequences $\{\varepsilon_I\}$ and applying the Khinchin lemma, we conclude that the norm $\|\varphi\|_{L^p(W)}$ is equivalent to

$$\Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\|W(t)^{1/p}c_I(\varphi)\chi_I(t)\|_E^2\Big)^{p/2}dt\Big\}^{1/p},$$

Note that when $p\ne 2$ this expression for the norm is *different* from the above expression involving norms $\langle\rho\rangle_{I,p}(c_I(\varphi))$ and their equivalence is a nontrivial result which *does not* hold for an arbitrary weight as it was for $p=2$.

Similarly for $\psi=\sum_{J\in\mathcal{D}'}c_J(\psi)h_J\in L^q(W^{-q/p})$ the norm $\|\psi\|_{L^q(W^{-q/p})}$ is equivalent to

$$\Big\{\int_{\mathbb{R}}\Big(\sum_{J\in\mathcal{D}'}\|W(t)^{-1/p}c_J(\psi)\chi_J(t)\|_E^2\Big)^{q/2}dt\Big\}^{1/q}.$$

So it is enough to show that for any coefficients $c_I,c'_J\in E$

$$\sum_{I\in\mathcal{D},J\in\mathcal{D}'}|T_{I,J}^{(r)}|\cdot\left|\big(c_I,c'_J\big)\right|\le$$
$$\le 2^{-r\alpha}C\cdot\Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\|W(t)^{1/p}c_I\chi_I(t)\|_E^2\Big)^{p/2}dt\Big\}^{1/p}\Big\{\int_{\mathbb{R}}\Big(\sum_{J\in\mathcal{D}'}\|W(t)^{-1/p}c'_J\chi_J(t)\|_E^2\Big)^{q/2}dt\Big\}^{1/q}.$$

Again it is convenient to replace the coefficient sequence $\{c_I\}$ by the corresponding function

$$f:\mathbb{R}\times\mathbb{Z}\to E,\qquad f(t,k):=\sum_{I\in\mathcal{D},\,|J|=2^k}c_I\chi_I\,,\qquad(t\in\mathbb{R},k\in\mathbb{Z}).$$

The norm

$$\Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\|W(t)^{1/p}c_I\chi_I(t)\|_E^2\Big)^{p/2}dt\Big\}^{1/p} \quad \text{is equivalent to the norm} \quad \|W^{1/p}f\|_{L^p_{\ell^2(E)}}$$

where $(W^{1/p}f)(t,k) := W^{1/p}(t)f(t,k)$ and the space $L^p_{\ell^2(E)}$ (which we will denote just by X for notational simplicity) consists of E-valued functions $F(t,k)$ on $\mathbb{R}\times\mathbb{Z}$ with the norm given by

$$\|F\|_X = \Big\{\int_{\mathbb{R}}\Big(\sum_{k\in\mathbb{Z}}\|F(t,k)\|_E^2\Big)^{p/2}dt\Big\}^{1/p}.$$

Clearly the dual space $X^* = L^q_{\ell^2(E)}$ where $1/p+1/q=1$, and for the sequence $\{c'_J\}$ the norm

$$\Big\{\int_{\mathbb{R}}\Big(\sum_{J\in\mathcal{D}'}\|W(t)^{-1/p}c'_J\chi_J(t)\|_E^2\Big)^{q/2}dt\Big\}^{1/q} \quad \text{is equivalent to} \quad \|W^{-1/p}g\|_{X^*}$$

where

$$g(t,k) = \sum_{J\in\mathcal{D}',\,|J|=2^k} c'_J\chi_J(t)\,, \qquad t\in\mathbb{R}, k\in\mathbb{Z}.$$

Thus we can rewrite the estimate we need as

$$\sum_{I\in\mathcal{D},J\in\mathcal{D}'}|T^{(r)}_{I,J}|\cdot\big|\big(c_I,c'_J\big)_E\big| =$$

$$= \sum_{k=-\infty}^{\infty}\iint_{\mathbb{R}^2}K^{(r)}_k(s,t)\cdot\big|\big(f(s,k-r),g(t,k)\big)_E\big|dsdt \le 2^{-r\alpha}C\|W^{1/p}f\|_X\|W^{-1/p}g\|_{X^*}\,,$$

where the kernels $K^{(r)}_k(s,t)$ are defined as

$$K^{(r)}_k(s,t) = \sum_{\substack{I\in\mathcal{D},\,|I|=2^{k-r}\\ J\in\mathcal{D}',\,|J|=2^k}}|T^{(r)}_{I,J}|\cdot|I|^{-1}\cdot|J|^{-1}\chi_I(s)\chi_J(t)\,.$$

We will show that the above estimate holds for any f and g for which $W^{1/p}f\in X$, $W^{-1/p}g\in X^*$, not only for f and g representing some coefficient sequences.

Consider a shift operator S defined on functions f on $\mathbb{R}\times\mathbb{Z}$ by $Sf(t,k)=f(t,k-1)$. Since clearly $\|W^{1/p}S^rf\|_X = \|W^{1/p}f\|_X$ for all $r\in\mathbb{Z}$, we can replace f by $S^{-r}f$ and check the inequality

$$\sum_{k=-\infty}^{\infty}\iint_{\mathbb{R}^2}K^{(r)}_k(s,t)\cdot\big|\big(f(s,k),g(t,k)\big)_E\big|dsdt \le 2^{-r\alpha}C\|W^{1/p}f\|_X\|W^{-1/p}g\|_{X^*}\,.$$

for the "diagonal integral operator".

Unfortunately we can not estimate each element of the diagonal (each integral operator with kernel $K^{(r)}_k(s,t)$) separately because, unlike the L^2 case, the norm in $X=L^p_{\ell^2(E)}$ cannot be expressed in terms of the norms $\|f(\cdot,k)\|$.

Fortunately, even without the separation of the layers we can apply the same idea as in the vector L^2 case. The kernels $K_k^{(r)}(s,t)$ can be estimated as

$$K_k^{(r)}(s,t) \le A \cdot 2^{-r\varepsilon/2} \cdot \frac{2^{k\varepsilon}}{(2^k + |t-s|)^{1+\varepsilon}}\,,$$

where A is some absolute constant. So, if we prove that for the kernels $\mathcal{K}_k$,

$$\mathcal{K}_k(s,t) = \frac{2^{k\varepsilon}}{(2^k + |t-s|)^{1+\varepsilon}}\,,$$

we have the estimate

$$\sum_{k=-\infty}^{\infty} \iint_{\mathbb{R}^2} \mathcal{K}_k(s,t) \cdot \left| \big(f(s,k), g(t,k)\big)_E \right| dsdt \le C \|W^{1/p} f\|_X \|W^{-1/p} g\|_{X^*}\,,$$

then we are done.

Changing notation $W^{1/p}f$ to f and $W^{1/p}g$ to g, we can rewrite the desired inequality as

$$\sum_{k=-\infty}^{\infty} \iint_{\mathbb{R}^2} \mathcal{K}_k(s,t) \cdot \left| \big(W^{-1/p}(s) f(s,k)\,,\, W^{1/p}(t) g(t,k)\big)_E \right| dsdt \le C \|f\|_X \|g\|_{X^*}\,.$$

Note that

$$\begin{aligned}\left| \big(W^{-1/p}(s) f(s,k)\,,\, W^{1/p}(t) g(t,k)\big)_E \right| &= \left| \big(f(s,k)\,,\, W^{-1/p}(s) W^{1/p}(t) g(t,k)\big)_E \right| \\ &\le \left\| W^{-1/p}(s) W^{1/p}(t) \right\| \cdot \left\| f(s,k) \right\|_E \left\| g(t,k) \right\|_E\end{aligned}$$

and therefore it is enough to prove that

$$\sum_{k=-\infty}^{\infty} \iint_{\mathbb{R}^2} \mathcal{K}_k(s,t) \cdot \left\| W^{-1/p}(s) W^{1/p}(t) \right\| \cdot \left\| f(s,k) \right\|_E \left\| g(t,k) \right\|_E dsdt \le C \|f\|_X \|g\|_{X^*}\,.$$

This inequality is clearly scalar: only scalar functions $\|f(s,k)\|_E$, $\|g(t,k)\|_E$ (we will denote them by the same letters f and g) are involved, so it is enough to prove that the "integral operator"

$$(\mathcal{K} f)(t,k) = \int_{\mathbb{R}} \mathcal{K}_k(s,t) \cdot \left\| W^{-1/p}(s) W^{1/p}(t) \right\| \cdot f(s,k) ds \qquad (f \in L^p_{\ell^2})$$

is bounded in the ℓ^2-valued space L^p.

Let $\mathcal{D}$ be any dyadic grid (it has nothing common with the grid used for φ: at this stage we have completely switched to the functions f and g and all letters are free again). Consider the averaging operator $A_{\mathcal{D}}$ defined by

$$(A_{\mathcal{D}} f)(t,k) = \sum_{I \in \mathcal{D}\,,\, |I| = 2^k} \frac{1}{|I|} \int_I \left\| W^{-1/p}(s) W^{1/p}(t) \right\| \cdot f(s,k)\, ds\,, \qquad s \in \mathbb{R}, k \in \mathbb{Z}.$$

Lemma 14.2. *If the weight W satisfies the Muckenhoupt condition $(\mathcal{A}_{p,q})$, where $1/p + 1/q = 1$, then for all dyadic grids $\mathcal{D}$ the corresponding averaging operators $A_{\mathcal{D}}$ are uniformly bounded in ℓ^2-valued L^p. Here "uniformly" means that the estimate for the norm does not depend on the dyadic grid $\mathcal{D}$.*

The theorem follows immediately from the lemma. Namely, averaging the operators $A_{\mathcal{D}}$ over all shifts of the dyadic grid $\mathcal{D}$, we observe that the "integral operator"

$$(\mathcal{T}f)(t,k) = \int_{\mathbb{R}} \tau_k(s-t) \cdot \left\| W^{-1/p}(s) W^{1/p}(t) \right\| \cdot f(s,k)\, ds$$

is bounded in ℓ^2-valued L^p. Since the shift operator S, $(Sf)(t,k) = f(t,k-1)$ and all its integer powers are isometries in ℓ^2-valued L^p, the integral operators $\mathcal{T}_j = S^{-j}\mathcal{T}S^j$, $j \in \mathbb{Z}$, or in the explicit form

$$(\mathcal{T}_j f)(t,k) = \int_{\mathbb{R}} \tau_{k+j}(s-t) \cdot \left\| W^{-1/p}(s) W^{1/p}(t) \right\| \cdot f(s,k)\, ds$$

have the same norm, and so all are uniformly bounded.

Since $\tau_{j+1} \geq 2^{-j-2}\chi_{[-2^j,2^j]}$ and

$$\frac{2^{k\varepsilon}}{(2^k + |s|)^{1+\varepsilon}} \leq C \sum_{j=0}^{\infty} 2^{-j\varepsilon} \cdot 2^{-j-k} \cdot \chi_{[-2^{j+k},2^{j+k}]}(s)\,,$$

we can conclude that

$$\frac{2^{k\varepsilon}}{(2^k + |s|)^{1+\varepsilon}} \leq C' \sum_{j=1}^{\infty} 2^{-j\varepsilon} \cdot 2^{-j-k} \cdot \tau_{j+k}(s)\,.$$

This means that the "integral operator" $\mathcal{K}$ (which we have to show be bounded) satisfies

$$\mathcal{K} \prec C' \sum_{j=1}^{\infty} 2^{-j\varepsilon} \mathcal{T}_j\,.$$

Here the symbol $\prec$ means that the kernel of the integral operator on the left is estimated entriwise by the kernel of the operator on the right (recall that all kernels are positive and "diagonal"). Therefore

$$\|\mathcal{K}\|_{L^p_{\ell^2} \to L^p_{\ell^2}} \leq C' \sum_{j=1}^{\infty} 2^{-j\varepsilon} \|\mathcal{T}_j\|_{L^p_{\ell^2} \to L^p_{\ell^2}} \leq C \leq \infty\,,$$

and we are done (modulo Lemma 14.2).

Proof of Lemma 14.2. Let V_I be the normalizing operator of the norm $\langle\rho\rangle_{I,p}$. Recall that for every vector $e \in E$ we have

$$\langle\rho\rangle_{I,p}(e) \asymp \|V_I^{-1} e\|_E\,.$$

Then the Muckenhoupt $(A_{p,q})$ condition allows to conclude from here that

$$\langle \rho^* \rangle_{I,q}(e) \asymp \langle \rho \rangle^*_{I,p}(e) \asymp \|V_I^* e\|_E .$$

Define an operator-valued function V on $\mathbb{R} \times \mathbb{Z}$ by

$$V(t,k) = \sum_{I\in\mathcal{D}\,,\,|I|=2^k} V_I \chi_I(t) .$$

Consider the realization of the Triebel-Lizorkin space $\mathcal{TL}_p$ as a subspace in $L^p_{\ell^2}$ discussed above in Section 9.1.4. We will prove that the operators $\mathcal{J}_1 : \mathcal{TL}_p \to L^p_{\ell^2}$,

$$(\mathcal{J}_1 \widetilde{f})(t,k) = \|W^{1/p}(t)V(t,k)\| \cdot \widetilde{f}(t,k), \qquad \widetilde{f} \in \mathcal{TL}_p,\ t\in\mathbb{R}, k\in\mathbb{Z}$$

and $\mathcal{J}_2 : \mathcal{TL}_q \to L^q_{\ell^2}$,

$$(\mathcal{J}_2 \widetilde{g})(t,k) = \|W^{-1/p}(t)\big(V(t,k)^*\big)^{-1}\| \cdot \widetilde{g}(t,k), \qquad \widetilde{g} \in \mathcal{TL}_q,\ t\in\mathbb{R}, k\in\mathbb{Z}$$

are uniformly bounded. If we show this, we are done. Indeed, the adjoint operator $\mathcal{J}_2^* : L^p_{\ell^2} \to \mathcal{TL}_p$ is given by

$$(\mathcal{J}_2^* f)(t,k) = \sum_{I\in\mathcal{D},|I|=2^k} \frac{1}{|I|}\int_I \|W^{-1/p}(s)\big(V(s,k)^*\big)^{-1}\| \cdot f(s,k)\, ds .$$

Note now that for every interval $I\in\mathcal{D}$ of length 2^k and every two points $s,t\in I$ we have

$$\begin{aligned}
&\|W^{-1/p}(s)\big(V(s,k)^*\big)^{-1}\| \cdot \|W^{1/p}(t)V(t,k)\| = \|W^{-1/p}(s)\big(V_I^*\big)^{-1}\| \cdot \|W^{1/p}(t)V_I\| \\
&= \|W^{-1/p}(s)\big(V_I^*\big)^{-1}\|\cdot\|V_I^* W^{1/p}(t)\| \ge \|W^{-1/p}(s)\big(V_I^*\big)^{-1} V_I^* W^{1/p}(t)\| = \|W^{-1/p}(s)W^{1/p}(t)\|
\end{aligned}$$

and therefore $A_{\mathcal{D}} \prec \mathcal{J}_1\mathcal{J}_2^*$, proving the lemma.

The boundedness of the operator $\mathcal{J}_1$ (the reasoning for $\mathcal{J}_2$ is exactly the same) is an immediate consequence of the norm equivalence property

$$\begin{aligned}
&\Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\|W(t)^{1/p}c_I\chi_I(t)\|_E^2\Big)^{p/2}dt\Big\}^{1/p} \\
&\qquad\asymp \Big\|\sum_{I\in\mathcal{D}} c_I h_I(\cdot)\Big\|_{L^p(W)} \asymp \Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\langle\rho\rangle_{I,p}(c_I)^2\chi_I(t)\Big)^{p/2}dt\Big\}^{1/p}
\end{aligned}$$

for the coefficient sequences $\{c_I\}\subset E$.

Recalling that $\langle\rho\rangle_{I,p}(c_I) \asymp \|V_I^{-1}c_I\|_E$, we can rewrite it as

$$\Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\|W(t)^{1/p}c_I\chi_I(t)\|_E^2\Big)^{p/2}dt\Big\}^{1/p} \asymp \Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\|V_I^{-1}c_I\|_E^2\chi_I(t)\Big)^{p/2}dt\Big\}^{1/p} .$$

Changing notation $V_I^{-1}c_I$ to c_I, we get

$$\Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\big\|W(t)^{1/p}V_I c_I\chi_I(t)\big\|_E^2\Big)^{p/2}dt\Big\}^{1/p}\asymp\Big\{\int_{\mathbb{R}}\Big(\sum_{I\in\mathcal{D}}\|c_I\|_E^2\chi_I(t)\Big)^{p/2}dt\Big\}^{1/p}=\big\|\{c_I\}_{I\in\mathcal{D}}\big\|_{\mathcal{TL}_p(E)}$$

or, in the function representation $F(s,k)=\sum_{I\in\mathcal{D},|I|=2^k}c_I\chi_I(s)$,

$$\big\|W^{1/p}(s)V(s,k)F(s,k)\big\|_{L^p_{\ell^2(E)}}\asymp\big\|F\big\|_{\mathcal{TL}_p(E)}$$

where $\mathcal{TL}_p(E)$ is the E-valued Triebel-Lizorkin space. Let $e\in E$ be any unit vector, $\widetilde{f}\in\mathcal{TL}_p$. Putting $F(s,k)=\widetilde{f}(s,k)e$, we observe that

$$\big\|\,\|W^{1/p}(s)V(s,k)e\|_E\cdot\widetilde{f}(s,k)\,\big\|_{L^p_{\ell^2}}\asymp\big\|\widetilde{f}\big\|_{\mathcal{TL}_p}\,.$$

Since for every orthonormal basis $e_1,\ldots,e_n$ we have

$$\|W^{1/p}(s)V(s,k)\|\asymp\sum_{j=1}^n\|W^{1/p}(s)V(s,k)e_j\|_E\,,$$

we conclude that

$$\big\|\|W^{1/p}(s)V(s,k)\|\widetilde{f}(s,k)\big\|_{L^p_{\ell^2}}\asymp\big\|\widetilde{f}\big\|_{\mathcal{TL}_p}\,,$$

i.e. $\|\mathcal{J}_1\widetilde{f}\|_{L^p_{\ell^2}}\asymp\big\|\widetilde{f}\big\|_{\mathcal{TL}_p}$, which is even more than we need.

15. Appendix 1. Strong unconditional bases of wavelets

In this section we will outline some results about wavelet bases in weighted L^p spaces. In particular we will show that the coefficient space of wavelet decompositions is canonically isomorfic to the corresponding *non-weighted* Triebel-Lizorkin space.

15.1. Biorthogonal wavelets

We will consider the so-called biorthogonal wavelets. Fix some dyadic lattice $\mathcal{D}$ and associate with any interval $I\subset\mathcal{D}$ two functions φ_I and ψ_I. Assume that these functions satisfy the following minimal requirements ($\alpha\in(0,1)$ is some small positive number).

1. Fast decay:
$$|\varphi_I(t)|,\ |\psi_I(t)|\le C\left(\frac{|I|}{d(t,I)}\right)^{1+\alpha},$$
where $d(t,I)=\operatorname{dist}(t,I)+|I|$;

2. Biorthogonality:
$$\int_{\mathbb{R}}\varphi_I(t)\overline{\psi}_J(t)dt=\begin{cases}|I|, & I=J;\\ 0, & I\ne J\end{cases}$$

3. Smoothness:

$$|\varphi_I(t_1) - \varphi_I(t_2)|,\ |\psi_I(t_1) - \psi_I(t_2)| \le C \left(\frac{|t_1 - t_2|}{|I|} \right)^{\alpha}$$

(due to the fast decay property it is indeed smoothness, not growth restriction: when $|t_1 - t_2| > |I|$ the condition holds automatically);

4. Completeness: the linear span $\mathcal{L}\{\varphi_I : I \in \mathcal{D}\}$ is dense in the scalar L^2 (note that we only require the completeness of $\{\varphi_I\}_{I\in\mathcal{D}}$; we will get the completeness of another system automatically);

5. Mean value zero: $\int_{\mathbb{R}} \varphi_I(t)dt = \int_{\mathbb{R}} \psi_I(t)dt = 0$.

Usually in wavelet theory the functions φ_I and ψ_I are obtained from a pair of functions φ and ψ by dilations and shifts. Namely, for a dyadic interval I let us denote by t_I its left endpoint, and define

$$\varphi_I(t) := \varphi\big(|I|^{-1}(t - t_I)\big)\,, \qquad \psi_I(t) := \psi\big(|I|^{-1}(t - t_I)\big)\,.$$

If we assume that φ and ψ satisfy 1 and 3 (say with $I = [0,1]$), then all functions φ_I, ψ_I automatically satisfy these properties. Conditions 1&2 for such functions imply 5, so we do not have to assume the latter.

Suppose we have a weighted space $L^p(W)$ of E-valued functions with matrix $(\mathcal{A}_{p,q})$ weight. Given the systems $\{\varphi_I\}_{I\in\mathcal{D}}$, $\{\psi_I\}_{I\in\mathcal{D}}$ satisfying 1–5, we can define the *wavelet subspaces* $\mathcal{E}_I$, $I \in \mathcal{D}$ by $\mathcal{E}_I = \varphi_I(\cdot)E = \{\varphi_I(\cdot)\mathbf{e} : \mathbf{e} \in E\}$. The corresponding coordinate projections $\mathcal{P}_I$ are defined for $f \in L^2 \cap L^p(W)$ by the formula

$$\mathcal{P}_I f(t) := \varphi_I(t) \int_{\mathbb{R}} f(s)\overline{\psi}_I(s)ds\,. \tag{15.1}$$

The main result of this section is the following theorem.

Theorem 15.1. *Let functions φ_I and ψ_I satisfy the conditions 1–5 above, and let $\{\mathcal{E}_I\}_{I\in\mathcal{D}}$ be the corresponding family of the wavelet subspaces. Suppose a matrix weight W satisfy the Muckenhoupt $(\mathcal{A}_{p,q})$ condition $(1/p + 1/q = 1)$. Then the system $\{\mathcal{E}_I\}_{I\in\mathcal{D}}$ is a strong unconditional basis in $L^p(\rho_t) = L^p(W)$, and the coefficient space is the same as for the Haar system, i.e.*

$$\Big\|\sum_{I\in\mathcal{D}} c_I\varphi_I(\cdot)\Big\|_{L^p(W)} \asymp \Big\|\big\{|I|^{-1/p}\|c_I\varphi_I(\cdot)\|_{L^p(W)}\big\}_{I\in\mathcal{D}}\Big\|_{\mathcal{TL}_p}$$
$$\asymp \Big\|\sum_{I\in\mathcal{D}} c_I h_I(\cdot)\Big\|_{L^p(W)} \asymp \Big\|\big\{\langle\rho\rangle_{I,p}(c_I)\big\}_{I\in\mathcal{D}}\Big\|_{\mathcal{TL}_p}$$

(the last equivalence is simply the norm equivalence property for the Haar system, which has already been proved while every other one is a non-trivial statement).

Proof. Consider an operator T defined (at least on the Haar functions) by $Th_I = \varphi_I$, $I \in \mathcal{D}$. Suppose we have proved that T is a bounded operator in $L^p(W)$, i.e. that

$$\Big\|\sum_{I\in\mathcal{D}} c_I\varphi_I(\cdot)\Big\|_{L^p(W)} \le C\Big\|\sum_{I\in\mathcal{D}} c_I h_I(\cdot)\Big\|_{L^p(W)}$$

for any sequence of vectors $c_I \in E$, $I \in \mathcal{D}$ with finitely many non-zero terms. Then the theorem is proved.

Indeed, since everything (except the property 4 which will not be used in the proof of the boundedness of T) is symmetric with respect to the exchange of W, p and $W^{-q/p}$, q, the operator T', $T'h_I = \psi_I$ is a bounded operator in $L^q(W^{-q/p})$.

The boundedness of T implies that for every $\mathbf{e} \in E$

$$\big\|\varphi_I\mathbf{e}\big\|_{L^p(W)} \le C\big\|h_I\mathbf{e}\big\|_{L^p(W)} = C|I|^{1/p}\langle\rho\rangle_{I,p}(\mathbf{e})\,, \tag{15.2}$$

Similarly

$$\big\|\psi_I\mathbf{e}'\big\|_{L^q(W^{-q/p})} \le C\big\|h_I\mathbf{e}'\big\|_{L^q(W^{-q/p})} = C|I|^{1/q}\langle\rho^*\rangle_{I,q}(\mathbf{e}')\,. \tag{15.3}$$

Since

$$\begin{aligned}\big|(\mathbf{e},\mathbf{e}')_E\big| = |I|^{-1}\big|(\varphi_I\mathbf{e},\psi_I\mathbf{e}')\big| \le |I|^{-1}\big\|\varphi_I\mathbf{e}\big\|_{L^p(W)}\big\|\psi_I\mathbf{e}'\big\|_{L^q(W^{-q/p})} \le \\ \le |I|^{-1/p}C\big\|\varphi_I\mathbf{e}\big\|_{L^p(W)}\langle\rho^*\rangle_{I,q}(\mathbf{e}')\,,\end{aligned}$$

the definition of the dual norm implies

$$\big\|\varphi_I\mathbf{e}\big\|_{L^p(W)} \ge C^{-1}|I|^{1/p}\langle\rho^*\rangle^*_{I,q}(\mathbf{e}) \asymp |I|^{1/p}\langle\rho\rangle_{I,p}(\mathbf{e}) = \big\|h_I\mathbf{e}\big\|_{L^p(W)}\,.$$

Together with the inequality (15.2) it results in

$$\big\|\varphi_I\mathbf{e}\big\|_{L^p(W)} \asymp \big\|h_I\mathbf{e}\big\|_{L^p(W)} = |I|^{1/p}\langle\rho\rangle_{I,p}(\mathbf{e})\,.$$

Similarly one can get that

$$\big\|\psi_I\mathbf{e}\big\|_{L^q(W^{-q/p})} \asymp \big\|h_I\mathbf{e}\big\|_{L^q(W^{-q/p})} = |I|^{1/q}\langle\rho^*\rangle_{I,q}(\mathbf{e})\,.$$

Let $\widetilde{P}_I$ denote the orthogonal projection in E-valued L^2 onto the Haar subspace $h_I(\cdot)E$,

$$\widetilde{P}_I f(s) = h_I(s)\int_I f(t)h_I(t)dt\,.$$

Since W satisfies the Muckenhoupt condition $(\mathcal{A}_{p,q})$, the operator $\widetilde{P}_I$ can be extended from the dense set $L^2\cap L^p(W)$ to a bounded operator in $L^p(W)$. The operator $T\widetilde{P}_I(T')^*$ coincides on the dense set $L^2 \cap L^p(W)$ with the projection $\mathcal{P}_I$ given by (15.1), so all the projections $\mathcal{P}_I$ are uniformly bounded.

Therefore to prove the theorem we can use Criterion 3 (Lemma 4.12). We are given the system of subspaces $\mathcal{E}_I$, $I \in \mathcal{D}$, and the corresponding projections are given by (15.1).

By Remark 4.13 the corresponding imbedding theorems can be rewritten as

$$\|\sum_{I\in\mathcal{D}} c_I\varphi_I(\cdot)\|_{L^p(W)} \le C\Big\|\{\langle\rho\rangle_{I,p}(c_I)\}_{I\in\mathcal{D}}\Big\|_{\mathcal{TL}_p} \asymp \|\sum_{I\in\mathcal{D}} c_I h_I(\cdot)\|_{L^p(W)},$$

and

$$\|\sum_{I\in\mathcal{D}} c_I\psi_I(\cdot)\|_{L^q(W^{-q/p})} \le C\Big\|\{\langle\rho^*\rangle_{I,q}(c_I)\}_{I\in\mathcal{D}}\Big\|_{\mathcal{TL}_q} \asymp \|\sum_{I\in\mathcal{D}} c_I h_I(\cdot)\|_{L^q(W^{-q/p})}$$

for all (finite) systems of vectors $cciI \in E$. But that is exactly the boundedness of the operators T and T'.

It remains to check the "Parceval identity"

$$(f\,,\,g) = \sum_{I\in\mathcal{D}}(\mathcal{P}_I f\,,\,\mathcal{P}_I^* g)\,.$$

It clearly holds for all $g \in$ (E-valued) L^2 and for all $f \in \mathcal{L}\{\varphi_I(\cdot)E\,:\,I\in\mathcal{D}\}$ (just write down what $\mathcal{P}_I^* g$ is). Note that $\mathcal{L}\{\varphi_I(\cdot)E\,:\,I\in\mathcal{D}\}$ is dense in the E-valued L^2. We know that the imbedding theorems and therefore the "Bessel inequality" hold *for every* matrix-weighted space $L^p(W)$ with weight $W\in(\mathcal{A}_{p,q})$. In particular, the Bessel inequality holds in the non-weighted ($W = Id$) space L^2. Thus the subspaces $\mathcal{E}_I$ form a strong unconditional basis in L^2 and the immediate conclusion is that the "Parceval identity" holds for every $f, g\in L^2$. This implies that it holds for all $f\in L^2\cap L^p(W)$ and $g\in L^2\cap L^q(W^{-q/p})$ (both sets are dense in $X = L^p(W)$ and $X^* = L^q(W^{-q/p})$ respectively) and the theorem is proved modulo the boundedness of the operator T. □

Proof of the boundedness of the operator T. Recall that we need to show that for any sequence of vectors $c_I\in E$ ($I\in\mathcal{D}$) containing finitely many non-zero elements one has

$$\Big\|\sum_{I\in\mathcal{D}} c_I\varphi_I(\cdot)\Big\|_{L^p(W)} \le C\Big\|\sum_{I\in\mathcal{D}} c_I h_I(\cdot)\Big\|_{L^p(W)}.$$

It is enough to check that for every function $g\in L^2\cap L^q(W^{-q/p})$ the inequality

$$\Big|\Big(\sum_{I\in\mathcal{D}} c_I\varphi_I, g\Big)\Big| = \Big|\sum_{I\in\mathcal{D}}(c_I\varphi_I, g)\Big| \le C\Big\|\sum_{I\in\mathcal{D}} c_I h_I\Big\|_{L^p(W)}\|g\|_{L^q(W^{-q/p})}$$

holds.

To do this, let us decompose g over the Haar system corresponding to a random dyadic lattice $\mathcal{D}(\omega)$. We have

$$\Big(\sum_{I\in\mathcal{D}} c_I\varphi_I, g\Big) = \sum_{I\in\mathcal{D},J\in\mathcal{D}(\omega)}(\varphi_I, h_J)\big(c_I, c_J(g)\big)_E$$

where, as before, $c_J(g) = |J|^{-1}\int_J g(t)h_J(t)dt$. Since $g,\varphi_I\in L^2$, the series converges absolutely.

Again we will call an interval $I\in\mathcal{D}$ bad if there exists an interval $J\in\mathcal{D}(\omega)$ such that $|J| > 2^{r_0}|I|$ and $\rho(I,J)\le\sqrt{|I||J|}$.

Lemma. *If I is good, then*

$$|(\varphi_I, h_J)| \le C|I|^{1+\varepsilon/2}|J|^{1+\varepsilon/2}d(I,J)^{-1-\varepsilon}$$

with $\varepsilon = 2\alpha^2(1+3\alpha)^{-1} > 0$.

Suppose that this estimate is proved. Then from Theorem 14.1 we immediately get

$$\Big|\sum_{I\in\mathcal{D}, I \text{ is good}} \big(c_I\varphi_I, g\big)\Big| \le C'\Big\|\sum_{I\in\mathcal{D}} c_I h_I\Big\|_{L^p(W)} \|g\|_{L^q(W^{-q/p})}$$

Now notice that when r_0 is large enough, the probability $\mathbb{P}\{I \text{ is good}\}$ is positive and does not depend on I. Let us denote this probability by β. We have

$$\beta\sum_{I\in\mathcal{D}}\big(c_I\varphi_I, g\big) = \mathbb{E}\sum_{I\in\mathcal{D}, I \text{ is good}}\big(c_I\varphi_I, g\big)\,,$$

which immediately implies the desired inequality with the constant $\beta^{-1}C'$. □

Proof of the lemma. It will be very similar to the reasoning we used when estimating (Th_I, h_J). We have 3 following bounds:

1) Fast decay yields

$$|(\varphi_I, h_J)| \le C\int_J \frac{|I|^{1+\alpha}}{d(t,I)^{1+\alpha}}dt \le C\int_J \frac{|I|^{1+\alpha}}{[|I| + dist(I,J)]^{1+\alpha}}dt = C\frac{|I|^{1+\alpha}|J|}{[|I| + dist(I,J)]^{1+\alpha}}\,.$$

2) Smoothness yields

$$|(\varphi_I, h_J)| \le \int_{J_-}\big|\varphi_I\big(t+\frac{|J|}{2}\big) - \varphi_I(t)\big|dt \le C\int_{J_-}\left(\frac{|J|}{|I|}\right)^{\alpha} dt \le C\frac{|J|^{1+\alpha}}{|I|^{\alpha}}\,.$$

3) Fast decay together with the condition $\int_{\mathbb{R}}\varphi_I(t)dt = 0$ imply that if I is good and $|J| > 2^{r_0}|I|$, then

$$|(\varphi_I, h_J)| \le 4\int_{\rho(I,J)}^{\infty}|\varphi_I(t)|dt \le C|I|^{1+\alpha}\rho(I,J)^{-\alpha} \le C|I|^{1+\alpha/2}|J|^{-\alpha/2}$$

(to get this bound when $I \subset J$, use the same trick with subtracting 1 or -1 from h_J which we applied when estimating (Th_I, h_J))

Now consider two cases:

A) $|J| \le \max\{2^{r_0}|I|, dist(I,J)\}$

Then, clearly, $d(I,J) \le 2^{r_0+1}[|I| + dist(I,J)]$ and the estimate (1) can be rewritten as

$$|(\varphi_I, h_J)| \le C2^{(r_0+1)(1+\alpha)} \frac{|I|^{1+\alpha}|J|}{d(I,J)^{1+\alpha}}.$$

For any $\gamma \in (0,1)$ we can interpolate between this inequality and one delivered by (2):

$$|(\varphi_I, h_J)| \le C' \left\{ \frac{|I|^{1+\alpha}|J|}{d(I,J)^{1+\alpha}} \right\}^{1-\gamma} \left\{ \frac{|J|^{1+\alpha}}{|I|^\alpha} \right\}^\gamma = C' \frac{|I|^{1+\alpha-\gamma(1+2\alpha)}|J|^{1+\alpha\gamma}}{d(I,J)^{(1+\alpha)(1-\gamma)}}.$$

Solving the equation $1+\alpha-\gamma(1+2\alpha) = 1+\alpha\gamma$ for γ we get $\gamma = \alpha(1+3\alpha)^{-1}$ and the desired estimate follows immediately.

B) $|J| > \max\{2^{r_0}|I|, dist(I,J)\}$.

Then we clearly have $d(I,J) \le 3|J|$ and the bound n (3) yields

$$|(\varphi_I, h_J)| \le C\frac{|I|^{1+\alpha/2}}{|J|^{\alpha/2}} = C\frac{|I|^{1+\alpha/2}|J|^{1+\alpha/2}}{d(I,J)^{1+\alpha}} \cdot \left(\frac{d(I,J)}{|J|}\right)^{1+\alpha} \le C3^{1+\alpha}\frac{|I|^{1+\alpha/2}|J|^{1+\alpha/2}}{d(I,J)^{1+\alpha}},$$

which is even better than claimed because $\varepsilon = 2\alpha^2(1+3\alpha)^{-1} < \alpha$ for every $\alpha > 0$. $\square$

16. Appendix 2. Maximal function

In this section we will show that the maximal function operator $\mathcal{M}$ acts in $L^p(W)$ if (and, obviously, only if) $W \in (\mathcal{A}_{p,q})$. We will start with the scalar case.

Theorem 16.1. *If $w \in (A_p)$, then for any $\varphi \in L^p(w)$*

$$\|\mathcal{M}\varphi\|_{L^p(w)} \le C\|\varphi\|_{L^p(w)}$$

where $(\mathcal{M}\varphi)(t) := \sup_{I\ni t}|I|^{-1}\int_I|\varphi(s)|ds$.

Proof. First of all note that it is enough to prove the statement for the *dyadic* maximal function $(\mathcal{M}_\mathcal{D}\varphi)(t) := \sup_{I\ni t, I\in\mathcal{D}}|I|^{-1}\int_I|\varphi(s)|ds$ (the general result can be obtained by averaging over all dyadic grids $\mathcal{D}$). We may also restrict ourselves to positive functions $\varphi \in L^\infty$.

Reduction to a dyadic imbedding theorem. Though we might follow the lines of Section 1.2, for diversity (or for uniformness, depends on what to compare to) we will use another way. Namely, let us denote by $I(t)$ ($t \in \mathbb{R}$) an interval in $\mathcal{D}$ for which

$$|I(t)|^{-1}\int_{I(t)}|\varphi(s)|ds \ge \frac{1}{2}\mathcal{M}_\mathcal{D}\varphi(t).$$

Such an interval exists due to our (irrelevant in any other respect) assumption $\varphi \in L^\infty$ and, moreover, we can make a measurable choice (i.e. such choice that the set

$$E_I := \{t \in \mathbb{R} : I(t) = I\}$$

is measurable for any $I \in \mathcal{D}$).

Clearly $E_I \subset I$ for every $I \in \mathcal{D}$ and the sets E_I are mutually disjoint. Now we have

$$\int_{\mathbb{R}} \mathcal{M}_{\mathcal{D}}\varphi(t)^p dt = \sum_{I\in\mathcal{D}} \int_{E_I} \mathcal{M}_{\mathcal{D}}\varphi(t)^p dt \le 2^p \sum_{I\in\mathcal{D}} \Big(\int_{E_I} w(t)dt\Big) \langle\varphi\rangle_I^p \,.$$

Thus Theorem 16.1 is equivalent to the following dyadic imbedding:

$$\sum_{I\in\mathcal{D}} \mu_I \langle\varphi\rangle_I^p \le C\|\varphi\|_{L^p(w)}^p$$

where $\mu_I = \int_{E_I} w(t)dt$.

What can the sequence $\{\mu_I\}$ be? Clearly

$$\sum_{J\subset I} \mu_J = \sum_{J\subset I} \int_{E_J} w(t)dt \le \int_I w(t)dt$$

because $E_J \subset J \subset I$ for any $J \subset I$. A bit closer look shows that this Carleson type condition is the *only* essential restriction, but formally we will not use this observation.

The Bellman function. We will use the Bellman function approach to prove this dyadic imbedding. The averages involved are

$$F_I = \langle\varphi^p w\rangle_I\,; \qquad u_I = \langle w\rangle_I\,; \qquad M_I = |I|^{-1} \textstyle\sum_{J\subset I} \mu_J\,;$$
$$f_I = \langle\varphi\rangle_I\,; \qquad v_I = \langle w^{-q/p}\rangle_I\,.$$

The Bellman function $\mathcal{B}(F, f, u, v, M)$ should satisfy the following properties:

1) Domain $D(C)$: $0 \le f \le F^{1/p}v^{1/q}$, $\quad 0 \le M \le u$, $\quad 1 \le uv^{p/q} \le C$ (Holder inequality, Carleson type condition and the bound for the (A_p) norm).

2) Range: $0 \le \mathcal{B} \le C'F$ (our belief that the result is true)

3) Key property: for any $0 \le \vartriangle\! M \le M$ and for every three corteges (F, f, u, v, M), $(F_\pm, f_\pm, u_\pm, v_\pm, M_\pm) \in D(C)$ satisfying

$$(F, f, u, v, M - \vartriangle\! M) = \frac{1}{2}\big\{(F_+, f_+, u_+, v_+, M_+) + (F_-, f_-, u_-, v_-, M_-)\big\}$$

one has

$$\mathcal{B}(F, f, u, v, M) \ge \vartriangle\! M f^p + \frac{1}{2}\big\{\mathcal{B}(F_+, f_+, u_+, v_+, M_+) + \mathcal{B}(F_-, f_-, u_-, v_-, M_-)\big\}\,.$$

Again we will construct an universal Bellman function, i.e. a function defined in the whole region

$$D = \cup_C D(C) = \{(F, f, u, v, M) : 0 \le f \le F^{1/p} v^{1/q}, \quad 0 \le M \le u, \quad 1 \le u v^{p/q}\}$$

and satisfying $0 \le \mathcal{B} \le F$ and the following key property:

$$\mathcal{B}(F, f, u, v, M) \ge \Xi(x) \Delta M f^p + \frac{1}{2}\{\mathcal{B}(F_+, f_+, u_+, v_+, M_+) + \mathcal{B}(F_-, f_-, u_-, v_-, M_-)\}$$

where $x = uv^{p/q}$ and $\Xi : [1, +\infty) \to (0, +\infty)$ is some strictly positive decreasing function.

The word by word repetition of the reasoning we used when proving Carleson Imbedding Theorem shows that instead of checking the key property in the finite difference form, we may check its infinitesimal version

3′) $d^2\mathcal{B} \le 0$;

3″) $\dfrac{\partial \mathcal{B}}{\partial M} \ge \Xi(x) f^p$.

Now to finish the proof, we need only to present a Bellman type function. Here it goes:

$$\mathcal{B}(F, f, u, v, M) = F - \frac{f^p}{2v^{p/q}}\{1 + \delta[(u + M)v^{p/q}]^{-\alpha}\}$$

where $\alpha > q/p$ and $\delta > 0$ is small enough.

Checking the properties. Let us check (3′) and (3″).

3″):

$$\frac{\partial \mathcal{B}}{\partial M} = \frac{\delta\alpha}{2\{(u + M)v^{p/q}\}^{\alpha+1}} f^p \ge 2^{-\alpha-2} \delta\alpha\, x^{-(\alpha+1)} f^p$$

(the last estimate is due to the inequality $0 \le M \le u$).

3′): Recall that $p/q = p - 1$ and therefore we need to show that

$$d^2\Big[\frac{f^p}{v^{p/q}}\{1 + \delta X^{-\alpha}\}\Big] \ge 0$$

where $X = (u + M)v^{p-1}$.

This can be done by direct computation of all second order partial derivatives and checking that the resulting 3 by 3 matrix is positively defined. Let us show another way which involves almost no computation at all and which allows to check the concavity of all functions appeared in this paper in almost no time.

Note that

$$d^2\left(\frac{f^p}{v^{p-1}}\right) = p(p-1)\frac{f^p}{v^{p-1}}\left(\frac{df}{f} - \frac{dv}{v}\right)^2 .$$

The second differential $d^2\big(f^p v^{-(p-1)} X^{-\alpha}\big)$, whatever it is, has the form

$$\frac{f^p}{v^{p-1}} X^{-\alpha} \Psi\Big(\frac{df}{f}, \frac{dv}{v}, \frac{du_1}{u_1}\Big)$$

where $u_1 = u + M$ and Ψ is a quadratic form with *constant* coefficients.

Lemma 16.2. *Let $\Phi(y_1, y_2, y_3) = \sum_{i,j=1}^{3} a_{ij} y_i y_j$ be any quadratic form with real coefficients. If the form $\Phi(y, y, y_3)$ is strictly positively defined, (i.e. if $\Phi(y, y, y_3) \geq \gamma(y^2 + y_3^2)$ for some $\gamma > 0$), then*

$$\Phi(y_1, y_2, y_3) \geq \frac{\gamma}{2}(y_2^2 + y_3^2) - A(y_1 - y_2)^2 \geq -A(y_1 - y_2)^2$$

for some $A > 0$ depending on the coefficients a_{ij} only.

The proof of this lemma is analogous to the proof of the similar lemma in Section 10.4.4 and is left to the reader as a simple excercise.

Thereby, if we show that Ψ is strictly positive on the hyperplane $\frac{df}{f} = \frac{dv}{v}$, we will have

$$d^2\Big[\frac{f^p}{v^{p/q}}\{1 + \delta X^{-\alpha}\}\Big] \geq \frac{f^p}{v^{p-1}}\big[p(p-1) - \delta X^{-\alpha} A\big]\Big(\frac{df}{f} - \frac{dv}{v}\Big)^2 .$$

Since always $X \geq 1$, it means that to make the second differential nonnegative, it is enough to take $\delta < p(p-1)A^{-1}$.

Now observe that to show that Ψ is strictly positively defined on the hyperplane $\frac{df}{f} = \frac{dv}{v}$ is the same as to show that the restriction of the function $f^p v^{-(p-1)} X^{-\alpha}$ to the hyperplane $f = v$ is strictly convex, i.e. to show that the function $v^p v^{-(p-1)}(u_1 v^{p-1})^{-\alpha} = u_1^{-\alpha} v^{1-\alpha(p-1)}$ of two variables u_1 and v is strictly convex. It is really true if both powers are negative, i.e. if $\alpha > (p-1)^{-1} = q/p$. □

The vector case: definition of the maximal function. Now let us consider the case of E-valued functions φ. The main problem here is (however strange it may seem) to *define* the maximal function. If the weight is scalar (i.e. $W(t) = w(t)Id$), we may just write

$$(\mathcal{M}\varphi)(t) := \sup_{I \ni t} |I|^{-1} \int_I \|\varphi(s)\|_E ds$$

and the function

$$\mathcal{B}(F, f, u, v, M) = F - \frac{\|f\|_E^p}{2v^{p/q}}\{1 + \delta\big[(u + M)v^{p/q}\big]^{-\alpha}\}$$

with the same α and δ as before constitutes the proof of the inequality

$$\|\mathcal{M}\varphi\|_{L^p(w)} \leq C\|\varphi\|_{L^p(W)} .$$

Note that even in this simplest case the space on the left is not the same as on the right: the maximal function of an E-valued function $\varphi \in L^p(W)$ lies in *scalar* $L^p(w)$. Actually there is no reasonable way to define a maximal function operator which really "acts in $L^p(W)$" when $\dim E > 1$.

Of course, when saying that the Bellman type function above constitutes the proof of this inequality we didn't mean that one *needs* to prove it from the very beginning (it is exactly the same scalar inequality in which for some strange reason one denoted an arbitrary positive scalar function from $L^p(w)$ by $\|\varphi\|_E$). Rather we meant that the key property of the scalar Bellman function holds for this new expression with one vector argument as well (which at this stage should be obvious to the reader).

The definition for the case when W is an arbitrary $\mathcal{A}_{p,q}$ weight comes from the idea that the maximal function should serve as a tool for estimation of convolution operators and other similar things. Let again denote the norm $\|W(t)^{1/p}\cdot\|_E$ by $\rho_t(\cdot)$. Define

$$\widetilde{\mathcal{M}}\varphi(t)=\sup\big\{\rho_t\big(|I|^{-1}\int_I k(s)\varphi(s)ds\big)\ :\ I\ni t;\ k:I\to\mathbb{C},\ \|k\|_{L^\infty}\le 1\big\}\,.$$

In the scalar case we clearly have $\widetilde{\mathcal{M}}\varphi=w^{1/p}\mathcal{M}\varphi$

Theorem 16.3.

$$\|\widetilde{\mathcal{M}}\varphi\|_{L^p}\le C\|\varphi\|_{L^p(W)}$$

(on the left the norm is taken in scalar non-weighted L^p*).*

Proof. As before we may consider the dyadic maximal function

$$\widetilde{\mathcal{M}}_{\mathcal{D}}\varphi(t)=\sup\big\{\rho_t\big(|I|^{-1}\int_I k(s)\varphi(s)ds\big)\ :\ I\ni t,\ I\in\mathcal{D};\ k:I\to\mathbb{C},\ \|k\|_{L^\infty}\le 1\big\}\,.$$

Define

$$\mathcal{F}_I:=\big\{|I|^{-1}\int_I k(s)\varphi(s)ds\ :\ k:I\to\mathbb{C},\ \|k\|_{L^\infty}\le 1\big\}\qquad(I\in\mathcal{D})\,.$$

Clearly $\mathcal{F}_I$ is a compact convex subset of E, symmetric with respect to the origin (recall that $L^p(W)\subset L^1_{\text{loc}}$ for any Muckenhoupt weight W). The definition of the maximal function can be then rewritten as

$$\widetilde{\mathcal{M}}_{\mathcal{D}}\varphi(t)=\sup\big\{\rho_t(f)\ :\ I\ni t,\ I\in\mathcal{D};\ f\in\mathcal{F}_I\big\}\,.$$

Let again $I(t)\ni t$ be an interval in $\mathcal{D}$ for which

$$\sup\big\{\rho_t(f)\ :\ f\in\mathcal{F}_{I(t)}\big\}>\frac{1}{2}\widetilde{\mathcal{M}}_{\mathcal{D}}\varphi(t)\,.$$

Let, at last, f_t be a vector in $\mathcal{F}_{I(t)}$ such that $\rho_t(f_t)>2^{-1}\widetilde{\mathcal{M}}_{\mathcal{D}}\varphi(t)$. Clearly we can make a measurable choice, i.e. such a choice that all the sets $E_I=\{t\in\mathbb{R}: I(t)=I\}$ and the function $t\mapsto f_t$ are Lebesgue measurable.

We have

$$\int_{\mathbb{R}}\widetilde{\mathcal{M}}_{\mathcal{D}}\varphi(t)^p dt=\sum_{I\in\mathcal{D}}\int_{E_I}\widetilde{\mathcal{M}}_{\mathcal{D}}\varphi(t)^p dt\le 2^p\sum_{I\in\mathcal{D}}\int_{E_I}\rho_t(f_t)^p dt\,.$$

The quantity $\int_{E_I}\rho_t(f_t)^p dt$ seems not very easy to control because f_t are different for different t. Note nevertheless that every convex symmetric set in a finite dimensional space E is comparable to a symmetric convex hull of not more than $n = \dim E$ vectors. This shows that actually

$$\sup_{f_t\in\mathcal{F}_I}\int_{E_I}\rho_t(f_t)^p dt \asymp \sup_{f\in\mathcal{F}_I}\int_{E_I}\rho_t(f)^p dt = \sup_{f\in\mathcal{F}_I}\mu_I(f)$$

where $\mu_I := \int_{E_I}\rho_t^p dt \in \mathcal{N}_p(E)$. The quantity on the right, obviously, is a function of "averages" μ_I and $\mathcal{F}_I$, which justifies the idea of Bellman function approach. It is funny that this observation which is very important philosophically (it shows that we lose nothing and the object we are looking for exists if the theorem is true) will play no role in "running the machine backward": the strict convexity term we will get will majorate the integral $\int_{E_I}\rho_t(f_t)^p dt$ for another, completely trivial reason.

The Bellman function: definition and properties. Let us write down the averages which will serve as arguments of our Bellman function. They are

$$F_I = \langle\rho_t(\varphi(t))^p\rangle_I \in \mathbb{R}; \qquad M_I = |I|^{-1}\textstyle\sum_{J\subset I}\mu_J \in \mathcal{N}_p(E);$$

$$U_I = \langle\rho_t^p\rangle_I \in \mathcal{N}_p(E); \qquad V_I = \langle(\rho_t^*)^q\rangle_I \in \mathcal{N}_q(E);$$

and ... $\mathcal{F}_I$! Though the convex set $\mathcal{F}_I$ in the role of a variable seems a little bit unusual, it still satisfies the martingale property

$$\mathcal{F}_I = \frac{1}{2}\big(\mathcal{F}_{I_-} + \mathcal{F}_{I_+}\big) := \Big\{\frac{f_- + f_+}{2} \ :\ f_\pm \in \mathcal{F}_{I_\pm}\Big\}$$

and this is all that really matters. Actually, you need even less: *any* reasonable relation between averages for I and those for I_+ and I_- would suffice. The importance of having martingale type relations comes from the fact that they correspond to the convexity type property for the Bellman function while other relations result in much more complicated (sometimes even not local) conditions to satisfy.

The domain for our Bellman function is defined by the following three inequalities:

$$(V^{1/q})^*(f) \le F^{1/p} \quad \forall f\in\mathcal{F}; \qquad M \le U; \qquad (V^{1/q})^* \le U^{1/p} \le C(V^{1/q})^*$$

(Holder inequality, Carleson type condition and $(\mathcal{A}_{p,q})$ condition respectively). Let A be an invertible linear operator in E; $z\in E$, $\|z\|_E = 1$.

Let $\eta > 0$ be some small number to be chosen later. Define

$$\mathrm{Tr}_z\, U := U(z) + \eta\,\mathrm{Tr}\, U\,.$$

(the reason for such surpressing all directions except one will get clear soon). Define

$$\begin{aligned}\mathcal{B}_{A,z}(F,\mathcal{F},U,V,M) &= \inf_{f\in\mathcal{F}}\mathcal{B}_{\mathrm{scalar}}(F, \|Af\|_E, \mathrm{Tr}_z\, U, \mathrm{Tr}\, V, \mathrm{Tr}_z\, M)\\ &= F - \frac{\sup_{f\in\mathcal{F}}\|Af\|_E^p}{2[\mathrm{Tr}(V\circ A^*)]^{p/q}}\big(1+\delta X_{A,z}^{-\alpha}\big)\end{aligned}$$

where

$$X_{A,z} = \big[\mathrm{Tr}_z(U \circ A^{-1}) + \mathrm{Tr}_z(M \circ A^{-1})\big]\big[\mathrm{Tr}(V \circ A^*)\big]^{p/q}.$$

Let us check that the definition is correct, i.e. that the cortege $(F, \|Af\|_E, \mathrm{Tr}_z U, \mathrm{Tr}\, V, \mathrm{Tr}_z M)$ always belongs to the domain of $\mathcal{B}_{\mathrm{scalar}}$. Indeed, since $M \le U$, we immediately get $\mathrm{Tr}_z M \le \mathrm{Tr}_z U$. Also, since $V^{1/q} \ge (U^{1/p})^*$, we get

$$\begin{aligned} 1 = |(z,z)_E| = |(A^{-1}z, A^*z)_E| &\le U^{1/p}(A^{-1}z)(U^{1/p})^*(A^*z) \\ &\le U^{1/p}(A^{-1}z)V^{1/q}(A^*z) \le [\mathrm{Tr}_z(U \circ A^{-1})]^{1/p}[\mathrm{Tr}(V \circ A^*)]^{1/q}. \end{aligned}$$

At last, for any $f \in \mathcal{F}$ and any unit vector $e \in E$

$$|(Af, e)_E| = |(f, A^*e)_E| \le (V^{1/q})^*(f)V^{1/q}(A^*e) \le F^{1/p}\big[\mathrm{Tr}(V \circ A^*)\big]^{1/q}$$

and therefore $\|Af\|_E \le F^{1/p}\big[\mathrm{Tr}(V \circ A^*)\big]^{1/q}$. Let

$$\mathcal{B}(\dots) = \inf_{A,z} \mathcal{B}_{A,z}(\dots).$$

Running the machine backward. Suppose that for a cortege $(F, \mathcal{F}, U, V, M)$ corresponding to an interval I (since there will be no other interval to consider, we will omit the index almost everywhere) the infimum is attained at some invertible operator A_{opt} and an unit vector z_{opt}. Let also f_{opt} be the vector in $\mathcal{F}$ for which

$$\|A_{\mathrm{opt}} f_{\mathrm{opt}}\|_E = \sup_{f \in \mathcal{F}} \|A_{\mathrm{opt}} f\|_E.$$

Let $(F_\pm, \mathcal{F}_\pm, U_\pm, V_\pm, M_\pm)$ be the corteges corresponding to $I_\pm$. We have

$$(F, \mathcal{F}, U, V, M - |I|^{-1}\mu_I) = \frac{1}{2}\big\{(F_-, \mathcal{F}_-, U_-, V_-, M_-) + (F_+, \mathcal{F}_+, U_+, V_+, M_+)\big\}.$$

Since $f_{\mathrm{opt}} \in \mathcal{F}$, there exist vectors $f_\pm \in \mathcal{F}_\pm$ such that $f_{\mathrm{opt}} = 2^{-1}(f_- + f_+)$. Thus we have

$$\begin{aligned} &|I|\mathcal{B}(F, \mathcal{F}, U, V, M) \\ &\quad = |I|\mathcal{B}_{\mathrm{scalar}}\big(F, \|A_{\mathrm{opt}} f_{\mathrm{opt}}\|_E, \mathrm{Tr}_{z_{\mathrm{opt}}}(U \circ A_{\mathrm{opt}}^{-1}), \mathrm{Tr}(V \circ A_{\mathrm{opt}}^*), \mathrm{Tr}_{z_{\mathrm{opt}}}(M \circ A_{\mathrm{opt}}^{-1})\big) \ge \\ &\Xi(x_{A_{\mathrm{opt}}, z_{\mathrm{opt}}})\,\mathrm{Tr}_{z_{\mathrm{opt}}}(\mu_I \circ A_{\mathrm{opt}}^{-1})\|A_{\mathrm{opt}} f_{\mathrm{opt}}\|_E^p + |I_-|\mathcal{B}_{\mathrm{scalar}}\big(F_-, \|A_{\mathrm{opt}} f_-\|_E, \mathrm{Tr}_{z_{\mathrm{opt}}}(U_- \circ A_{\mathrm{opt}}^{-1}), \dots\big) \\ &\qquad + |I_+|\mathcal{B}_{\mathrm{scalar}}\big(F_+, \|A_{\mathrm{opt}} f_+\|_E, \mathrm{Tr}_{z_{\mathrm{opt}}}(U_+ \circ A_{\mathrm{opt}}^{-1}), \dots\big) \\ &\quad \ge \Xi(x_{A_{\mathrm{opt}}, z_{\mathrm{opt}}})\,\mathrm{Tr}_{z_{\mathrm{opt}}}(\mu_I \circ A_{\mathrm{opt}}^{-1})\|A_{\mathrm{opt}} f_{\mathrm{opt}}\|_E^p + |I_-|\mathcal{B}(F_-, \mathcal{F}_-, \dots) + |I_+|\mathcal{B}(F_+, \mathcal{F}_+, \dots), \end{aligned}$$

where $x_{A,z} = \mathrm{Tr}_z(U \circ A^{-1})[\mathrm{Tr}(V \circ A^*)]^{p/q}$.

Note that

$$\mathrm{Tr}_{z_{\mathrm{opt}}}(\mu_I \circ A_{\mathrm{opt}}^{-1})\|A_{\mathrm{opt}} f_{\mathrm{opt}}\|_E^p \geq \eta\, \mathrm{Tr}(\mu_I \circ A_{\mathrm{opt}}^{-1})\|A_{\mathrm{opt}} f_{\mathrm{opt}}\|_E^p$$
$$= \eta \int_{E_I} \mathrm{Tr}(\rho_t^p \circ A_{\mathrm{opt}}^{-1})\|A_{\mathrm{opt}} f_{\mathrm{opt}}\|_E^p \, dt \geq \eta \int_{E_I} \mathrm{Tr}(\rho_t^p \circ A_{\mathrm{opt}}^{-1})\|A_{\mathrm{opt}} f_t\|_E^p \, dt \geq \eta \int_{E_I} \rho_t(f_t)^p dt$$

for any family of vectors $f_t \in \mathcal{F}$. So, if we show that for sufficiently small $\eta = \eta(C) > 0$ the quantity $x_{A_{\mathrm{opt}}, z_{\mathrm{opt}}}$ is bounded from above whenever $U^{1/p} \leq C(V^{1/q})^*$, we will have done.

The extremal problem. It is enough to show that for any vector $f \in E$ the solution $A_{\mathrm{opt}}, z_{\mathrm{opt}}$ of the extremal problem

$$\det A \neq 0,\ \|z\|_E = 1; \qquad \frac{\|Af\|_E^p}{[\mathrm{Tr}(V \circ A^*)]^{p/q}}\left(1 + \delta X_{A,z}^{-\alpha}\right) \longrightarrow \max$$

exists and satisfies $x_{A_{\mathrm{opt}}, z_{\mathrm{opt}}} \leq C'(n, C)$. Let A_0 be the ajoint operator to the normalizing operator of the norm $V^{1/q}$. Let $\widetilde{V} = V \circ A_0^*$, $\widetilde{U} = U \circ A_0^{-1}$, $\widetilde{M} = M \circ A_0^{-1}$ and, at last, $\widetilde{f} = A_0 f$. Then the factorization $A = BA_0$ allows to rewrite the extremal problem as

$$\det B \neq 0,\ \|z\|_E = 1; \qquad \frac{\|B\widetilde{f}\|_E^p}{[\mathrm{Tr}(\widetilde{V} \circ B^*)]^{p/q}}\left(1 + \delta \widetilde{X}_{B,z}^{-\alpha}\right) \longrightarrow \max$$

where
$$\widetilde{X}_{B,z} = X_{BA_0,z} = \left[\mathrm{Tr}_z(\widetilde{U} \circ B^{-1}) + \mathrm{Tr}_z(\widetilde{M} \circ B^{-1})\right]\left[\mathrm{Tr}(\widetilde{V} \circ B^*)\right]^{p/q}.$$
The rest of the reasoning is very similar to that in Section 11.5.3, only much simpler technically. Again we observe that multiplication of B by a scalar factor changes nothing and therefore we may restrict ourselves to operators B of norm 1.

Let B_0 be the solution of the extremal problem

$$\|B\| = 1; \qquad \frac{\|B\widetilde{f}\|_E}{[\mathrm{Tr}(\widetilde{V} \circ B^*)]^{1/q}} \longrightarrow \max .$$

By the definition of the normalizing operator we have $n^{-q}\|\cdot\|_E^q \leq \widetilde{V}(\cdot) \leq \|\cdot\|_E^q$ and thereby (due to the Muckenhoupt condition and the inequality $M \leq U$, we get

$$\widetilde{M}(\cdot) \leq \widetilde{U}(\cdot) \leq C^p n^p \|\cdot\|_E^p .$$

Let B_ε $(0 \leq \varepsilon \leq 1)$ be the family of operators such that $\|B_\varepsilon\| = 1$, $\|B_\varepsilon - B_0\| \leq \varepsilon$ and $\|B_\varepsilon^{-1}\| \leq \varepsilon^{-1}$ (the construction is the same as in Section 11.5.3). Then for operators B_ε we have again

$$\frac{\|B_\varepsilon \widetilde{f}\|_E^p}{[\mathrm{Tr}(\widetilde{V} \circ B_\varepsilon^*)]^{p/q}} \geq (1 - 2n^2\varepsilon)^p \frac{\|B_0 \widetilde{f}\|_E^p}{[\mathrm{Tr}(\widetilde{V} \circ B_0^*)]^{p/q}} .$$

Also, as before, $[\mathrm{Tr}(\widetilde{V}\circ B_\varepsilon^*)]^{p/q}\le n^p$ and

$$\mathrm{Tr}_z(\widetilde{U}\circ B_\varepsilon^{-1})+\mathrm{Tr}_z(\widetilde{M}\circ B_\varepsilon^{-1})\le 2\,\mathrm{Tr}_z(\widetilde{U}\circ B_\varepsilon^{-1})=2\widetilde{U}(B_\varepsilon^{-1}z)+2\eta\,\mathrm{Tr}(\widetilde{U}\circ B_\varepsilon^{-1})\,.$$

Note now that

$$\mathrm{Tr}(\widetilde{U}\circ B_\varepsilon^{-1})\le C^pn^p\,\mathrm{Tr}(\|\cdot\|_E^p\circ B_\varepsilon^{-1})\le C^pn^{2p}\|B_\varepsilon^{-1}\|^p\le C^pn^{2p}\varepsilon^{-p}$$

and that, since $\|B_\varepsilon\|=1$, we always can find *one* unit vector $z_\varepsilon\in E$ for which $\|B_\varepsilon^{-1}z_\varepsilon\|_E=1$ and thereby $\widetilde{U}(B_\varepsilon^{-1}z_\varepsilon)\le C^pn^p$. This results in

$$\widetilde{X}_{B_\varepsilon,z_\varepsilon}\le 2C^pn^{2p}(1+n^p\eta\varepsilon^{-p})$$

and finally in

$$\frac{\|B_\varepsilon\widetilde{f}\|_E^p}{[\mathrm{Tr}(\widetilde{V}\circ B_\varepsilon^*)]^{p/q}}\big(1+\delta\widetilde{X}_{B_\varepsilon,z_\varepsilon}^{-\alpha}\big)\ge(1-2n^2\varepsilon)^p\big[1+\delta2^{-\alpha}n^{-2p\alpha}C^{-p\alpha}(1+n^p\eta\varepsilon^{-p})^{-\alpha}\big]\frac{\|B_0\widetilde{f}\|_E^p}{[\mathrm{Tr}(\widetilde{V}\circ B_0^*)]^{p/q}}\,.$$

Take some $\varepsilon>0$ for which

$$(1-2n^2\varepsilon)^p\big[1+\delta2^{-\alpha}n^{-2p\alpha}C^{-p\alpha}(1+n^p)^{-\alpha}\big]>1$$

and put $\eta=\varepsilon^p$. Clearly everything depends on n and C only. Denoting the difference between the quantity on the left and 1 by $\delta\nu^\alpha$ $(\nu=\nu(n,C)>0)$, we see that in order to be better than $(B_\varepsilon,z_\varepsilon)$, a pair (B,z) should satisfy $X_{B,z}\le\nu^{-1}$ and therefore

$$x_{A_{\mathrm{opt}},z_{\mathrm{opt}}}\le X_{A_{\mathrm{opt}},z_{\mathrm{opt}}}=\widetilde{X}_{B_{\mathrm{opt}},z_{\mathrm{opt}}}\le\nu^{-1}\,,$$

finishing the proof. □

References

[1] R. Hunt, B. Muckenhoupt and R. Wheeden, Weighted norm inequalities for conjugate function and Hilbert transform, *Trans. Amer. Math. Soc.,* **176** (1973), pp. 227–251.

[2] S. Treil and A. Volberg, *Wavelets and the angle between past and future*, Univ. of Göttingen preprint, Geometrie und Analysis, heft 7 (1995), pp. 1–33; to appear in Journal of Functional Analysis.

[3] F. Nazarov and S. Treil, *The weighted norm inequaities for Hilbert Transform are now trivial*, to appear in C. R. Acad Sci. Paris.

[4] D.I.Burkholder, *Boundary value problems and sharp inequalities for martingale transforms*, The Annals of Probability (1984), Vol. 12, No. 3, pp. 647–702.