THE CARLESON EMBEDDING THEOREM WITH MATRIX WEIGHTS

AMALIA CULIUC AND SERGEI TREIL

Abstract. In this paper we prove the weighted martingale Carleson embedding theorem with matrix weights both in the domain and in the target space.

1. Introduction and main results

The main result of this paper is the matrix weighted martingale Carleson embedding theorem, where matrix weights appear in both the domain and the target space. The need for such result is motivated by the attempt to generalize the two weight estimates for well localized operators from [6], to the case of matrix-valued measures. The main part of the estimate in [6] is the two weight inequality for paraproducts, and this estimate for the matrix-valued measures can be reduced exactly to the embedding theorem treated in this paper.

Earlier versions of the matrix weighted Carleson Embedding Theorem theorem under fairly strong additional assumptions (such as the weight belonging to the $A_2$ class) go back to [7] and, more recently, [3], [1]. Two weight estimates with matrix weights for well-localized operators, also under additional assumptions, were treated in [4] and [1] (see also [2]), but the result is still not known in full generality.

The weighted embedding theorem presented in this paper does not assume any properties for the matrix weight except local boundedness, and produces an embedding constant that depends polynomially on the dimension of the space. As in the scalar case, our embedding theorem states the Carleson measure condition, which is just a simple testing condition, implies the embedding. For matrix weights the Carleson measure condition (condition (ii) in Theorem 1.1 or condition (iii) in Theorem 1.2) is an inequality between positive semidefinite matrices.

In the case of scalar weights in the domain, the right hand side of the inequality is a multiple of the identity matrix $I$: in this situation, sacrificing constants, one can replace matrices by their norms, and the matrix embedding theorem trivially follows from the scalar one. Of course, the constants obtained by such trivial reduction are far from optimal: constants of optimal order were obtained using more complicated reasoning in [5].

In our case, both sides of the Carleson measure condition are general positive semidefinite matrices, so the simple strategy of replacing matrices by norms or traces does not work. A more complicated idea, in the spirit of the argument in [5], is used to get the result.

1.1. Setup.

1.1.1. Atomic filtered spaces. Let $(\mathcal{X}, \mathcal{F}, \sigma)$ be a sigma-finite measure space with an atomic filtration $\mathcal{F}_n$, that is, a sequence of increasing sigma-algebras $\mathcal{F}_n \subset \mathcal{F}$ such that for each $\mathcal{F}_n$ there exists a countable collection $D_n$ of disjoint sets of finite measure with the property that every set of $\mathcal{F}_n$ is a union of sets in $D_n$. 

2010 Mathematics Subject Classification. Primary 42B20, 60G42, 60G46.
Key words and phrases. Carleson embedding theorem, matrix weights, Bellman function.
Supported in part by the National Science Foundation under the grant DMS-1301579.
We will call the sets $I \in D_n$ atoms, and denote by $D$ the collection of all atoms, $D = \cup_{n \in \mathbb{Z}} D_n$. We allow a set $I$ to belong to several generations $D_n$, so formally an atom $I \in D_n$ is a pair $(I, n)$. To avoid overloading the notation, we skip the “time” $n$ and write $I$ instead of $(I, n)$; if we need to “extract” the time $n$, we will use the symbol $\text{rk} I$. Namely, if $I$ denotes the atom $(I, n)$ then $n = \text{rk} I$.

The inclusion $I \subset J$ for atoms should be understood as inclusion for the sets together with the inequality $\text{rk} I \geq \text{rk} J$. However, the union (intersection) of atoms is just the union (intersection) of the corresponding sets and “times” $n$ are not taken into account.

A standard example of such a filtration is the dyadic lattice $D$ on $\mathbb{R}^N$, which explains the choice of notation. However, in what follows, $D$ will always denote a general collection of atoms and $I \in D$ will stand for an atom in $D$, and not necessarily for a dyadic interval.

1.1.2. Matrix-valued measures. Let $\mathcal{F}_0$ be the collection of sets $E \cap F$ where $E \in \mathcal{F}$ and $F$ is a finite union of atoms. A $d \times d$ matrix-valued measure $W$ on $\mathcal{X}$, is a countably additive function on $\mathcal{F}_0$ with values in the set of $d \times d$ positive semidefinite matrix. Equivalently, $W = (w_{j,k})_{j,k=1}^d$ is a $d \times d$ matrix whose entries $w_{j,k}$ are (possibly signed or even complex-valued) measures, finite on atoms, and such that for any $E \in \mathcal{F}_0$ the matrix $(w_{j,k}(E))_{j,k=1}^d$ is positive semidefinite. Note that the measure $W$ is always finite on atoms.

The weighted space $L^2(W)$ is defined as the set of all measurable $\mathbb{F}^d$-valued functions ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) such that
\[
\|f\|_{L^2(W)}^2 := \int_{\mathcal{X}} \langle W(dx)f(x), f(x) \rangle_{\mathbb{F}^d} < \infty;
\]
as usual we take the quotient space over the set of functions of norm 0.

1.2. Main results.

**Theorem 1.1.** Let $W$ be a $d \times d$ matrix-valued measure and let $\tilde{A}_I$ be positive semidefinite $d \times d$ matrices. The following statements are equivalent:

(i) \[
\sum_{I \in D} \left\| \tilde{A}_I^{1/2} \int_I W(dx)f(x) \right\|^2 \leq A\|f\|_{L^2(W)}^2
\]

(ii) \[
\sum_{I \in D} W(I)\tilde{A}_I W(I) \leq BW(I_0) \text{ for all } I_0 \in D.
\]

Moreover, for the best constants $A$ and $B$ we have $B \leq A \leq CB$, where $C = C(d)$ is a constant depending only on the dimension $d$.

Note that the underlying measure $\sigma$ is absent from the statement of the theorem: we do not need $\sigma$ in the setup, we only need the filtration $\mathcal{F}_n$. Alternatively, we can pick $\sigma$ to make the setup more convenient. For example, if we define $\sigma := \text{tr} W := \sum_{k=1}^d w_{k,k}$, then the measures $w_{j,k}$ are absolutely continuous with respect to $\sigma$. Thus, we can always assume that our matrix-valued measure is an absolutely continuous one $Wd\sigma$, where $W$ is a matrix weight, i.e. a locally integrable (meaning integrable on all atoms $I$) matrix-valued function with values in the set of positive semidefinite matrices.

For a measurable function $f$ we denote by $\langle f \rangle_I$ its average,
\[
\langle f \rangle_I := \sigma(I)^{-1} \int_I f d\sigma;
\]
if $\sigma(I) = 0$ we put $\langle f \rangle_I = 0$. The same definition is used for the vector and matrix-valued functions.

In what follows we will often use $|E|$ for $\sigma(E)$ and $dx$ for $d\sigma$. 
The theorem below is the restatement of Theorem 1.1 in this setup, if we put $A_I = |I|^{-1}A_I$. More precisely, Theorem 1.1 is just the equivalence (ii) $\iff$ (iii) in Theorem 1.2. The equivalence (i) $\iff$ (ii) will be explained below.

**Theorem 1.2.** Let $W$ be a $d \times d$ matrix-valued weight and let $A_I$, $I \in D$ be a sequence of positive semidefinite $d \times d$ matrices. Then the following are equivalent:

(i) $\sum_{I \in D} \left\| A_I^{1/2} \langle W I \rangle_I \right\|^2 |I| \leq A \| f \|_{L^2}$.

(ii) $\sum_{I \in D} \left\| A_I^{1/2} \langle W f \rangle_I \right\|^2 |I| \leq A \| f \|_{L^2(W)}^2$.

(iii) $\frac{1}{|I_0|} \sum_{I \in D, I \subset I_0} \langle W \rangle_I A_I \langle W \rangle_I |I| \leq B \langle W \rangle_{I_0}$ for all $I_0 \in D$.

Moreover, $B \leq A \leq CB$, where $C = C(d) = e \cdot d^3(d+1)^2$.

**Acknowledgement.** The authors would like to thank F. Nazarov for suggesting the crucial idea of using a “Bellman function with a parameter” argument, similar to the one used earlier in [5].

## 2. Proof of the main result

### 2.1. Trivial reductions.

The equivalence of (i) and (ii) is trivial. In (i), perform the change of variables $f := W^{1/2}f$ to obtain (ii) and similarly, in (ii) set $f := W^{-1/2}f$ to obtain (i). Note that here we do not need to assume that the weight $W$ is invertible a.e.: we just interpret $W^{-1/2}$ as the Moore–Penrose inverse of $W^{1/2}$.

The implication (i) $\implies$ (iii) and the estimate $A \geq B$ are obvious by setting $f = W^{1/2}1_I e$, $e \in \mathbb{F}^d$ in (i). Equivalently, to show that (ii) $\implies$ (iii) one just needs to apply (ii) to the test functions $f = 1_I e$.

So it only remains to prove that (iii) $\implies$ (i), or equivalently, that (iii) $\implies$ (ii).

### 2.1.1. Invertibility of $W$.

Let us notice that without loss of generality we can assume that the weight $W$ is invertible a.e., and even that the weight $W^{-1}$ is uniformly bounded.

To show that, define for $\alpha > 0$ the weight $W_\alpha$ by $W_\alpha(s) := W(s) + \alpha I$, and let

$$A_I^\alpha := \langle W_\alpha \rangle_I^{-1} \langle W \rangle_I A_I \langle W \rangle_I (W_\alpha)^{-1}.$$  

If (iii) is satisfied, then trivially

$$\frac{1}{|I_0|} \sum_{I \in D, I \subset I_0} \langle W_\alpha \rangle_I A_I^\alpha \langle W_\alpha \rangle_I |I| \leq B \langle W \rangle_{I_0} \leq B \langle W_\alpha \rangle_{I_0}.$$  

If Theorem 1.2 holds for invertible weights $W$, we get that for all $f \in L^2(W) \cap L^2$

$$\sum_{I \in D} \left\| A_I^\alpha / 2 \langle W_\alpha f \rangle_I \right\|^2 |I| \leq A \| f \|_{L^2(W_\alpha)}^2.$$  

Noticing that $\| f \|_{L^2(W_\alpha)} \to \| f \|_{L^2(W)}$, $\langle W_\alpha f \rangle_I \to \langle W f \rangle_I$, $A_I \to A_I$ as $\alpha \to 0^+$ we immediately get (ii) for all $f \in L^2(W) \cap L^2$; taking the limit inside the sum is justified because an infinite sum of non-negative numbers is the supremum of all finite subsums, and finite sums commute with limits.

Since the estimate (ii) holds on a dense set, extending the embedding operator by continuity we trivially get that it holds for all $f \in L^2(W)$. 

---

**THE CARLESON EMBEDDING THEOREM WITH MATRIX WEIGHTS 3**
2.2. The Bellman functions. By homogeneity we can assume without loss of generality that \( B = 1 \). As we discussed above, we only need to prove the implication (iii) \( \implies \) (i).

Following a suggestion by F. Nazarov we will do so by a “Bellman function with a parameter” argument similar to one presented in \([5]\). Denote
\[
F_I = \|f\|_{L^2(I)}^2 := \langle |f|^2 \rangle_I
\]
(2.1)
\[
M_I = \frac{1}{|I|} \sum_{J \subset I} \langle W \rangle_J A_J \langle W \rangle_J
\]
(2.2)
\[
x_I = \langle W^{1/2} f \rangle_I.
\]
(2.3)

For any real number \( s, 0 \leq s < \infty \), define the Bellman function
\[
B_s(I) = B_s(F_I, x_I, M_I) = \left( \langle (W)_{I+s} M_I \rangle^{-1} x_I, x_I \right)_{\mathbb{R}^d}.
\]
(2.4)

Notice that \( F_I \) is not involved in the definition of \( B_s(I) \), but it will be used in the estimates.

The functions \( B_s \) satisfy the following properties:

(i) The range property: \( 0 \leq B_s(I) \leq F_I \);
(ii) The key inequality:
\[
B_s(I) + s R_I(s) \leq \sum_{I' \in \text{ch}(I)} \frac{|I'|}{|I|} B_s(I')
\]
(2.5)

where
\[
R_I(s) = \|A_{1/2} (W)_{I} (\langle W \rangle_{I+s} M_I)^{-1} x_I\|^2.
\]

The inequality \( B_s(I) \geq 0 \) is trivial, and the inequality \( B_s(I) \leq F_I \) follows immediately from Lemma 3.1 below. The key inequality (2.5) is a consequence of Lemma 3.3, which we also prove below.

2.3. From Bellman functions to the estimate. Let us rewrite (2.5) as
\[
|I| B_s(I) + |I| s R_I(s) \leq \sum_{I' \in \text{ch}(I)} |I'| B_s(I').
\]

Then, applying this estimate to each \( B_s(I') \), and then to each descendant of each \( I' \), we get, going \( m \) generations down,
\[
|I| B_s(I) + \sum_{I' \in \mathcal{D}, I' \subset I \atop \text{rk } I' < \text{rk } I + m} s R_{I'}(s) |I'| \leq \sum_{I' \in \mathcal{D}, I' \subset I \atop \text{rk } I' = \text{rk } I + m} |I'| B_s(I') \leq \|f 1_I\|_{L^2}^2;
\]

here in the last inequality we used the fact that \( B_s(I) \leq F_I = \langle \|f(\cdot)\|_I^2 \rangle = |I|^{-1} \|f 1_I\|_{L^2}^2 \).

Letting \( m \to \infty \) and ignoring the non-negative term \( s B_s(I) \) in the left had side, we get that
\[
s \sum_{I' \in \mathcal{D}, I' \subset I} R_{I'}(s) |I'| \leq \|f 1_I\|_{L^2}^2.
\]

Summing the above inequality over all \( I \in \mathcal{D}_n \) we obtain
\[
s \sum_{I' \in \mathcal{D}, \text{rk } I' \geq n} R_{I'}(s) |I'| \leq \|f\|_{L^2}^2.
\]

Then, letting \( n \to -\infty \) and replacing \( I' \) by \( I \), we arrive to the estimate
\[
s \sum_{I \in \mathcal{D}} R_I(s) \leq \|f\|_{L^2}^2.
\]
(2.6)
Note that $R_I(0) = \|A_I^{1/2}x_I\| = \|A_I^{1/2}(W_I^{1/2}f)_I\|$, so to prove (i) we need to estimate $\sum_I R_I(0)$, and we only have the estimate of $s \sum_I R_I(s)!$

In the scalar case we trivially (since $M_I < \langle W \rangle_I$) have $R_I(0) \leq 4R_I(1)$, which gives us (i) with constant $4B$.

But due to non-commutativity, such estimate fails in the matrix case, so an extra trick is needed. The final step in the proof of Theorem 1.2 is the following lemma:

**Lemma 2.1.** For $\varepsilon > 0$

\[ R_I(0) \leq C(\varepsilon, d) \frac{1}{\varepsilon} \int_0^\varepsilon s R_I(s) ds. \]

Moreover, for $\varepsilon = 2/d$ we can have $C(d) = e \cdot d^3(d + 1)^2$.

Applying the lemma to (2.6), we get

\[ \sum_{I \in D} R_I(0) \leq e \cdot d^3(d + 1)^2 \|f\|_{L^2}^2, \]

which proves the theorem (modulo Lemma 2.1).

**2.4. Proof of Lemma 2.1.** Observe that it follows from the cofactor inversion formula that the entries of the matrix $((W)_I + sM_I)^{-1}$ are of the form $p_{j,k}(s)/Q(s)$, where

\[ Q(s) = Q_I(s) = \text{det}((W)_I + sM_I) \]

is a polynomial of degree at most $d$, and $p_{j,k}(s)$ are polynomials of degree at most $d - 1$.

Therefore $R_I$ is a rational function in $s$, $R_I(s) = \tilde{P}_I(s)/|Q_I(s)|^2$, where $\tilde{P}_I(s)$ is a polynomial of degree at most $2(d - 1)$ and $\tilde{P}_I(s) \geq 0$. We can then write $\tilde{P}_I(s) = |P_I(s)|^2$, where $P_I$ has degree at most $d - 1$. Therefore $R_I(s) = |P_I(s)/Q_I(s)|^2$.

By hypothesis, $M_I \leq \langle W \rangle_I$, so the operator $(W)_I + sM_I$ is invertible for all $s$ such that $\text{Re}(s) > -1$. Thus the zeroes of $Q_I(s)$ are all in the half plane $\text{Re}(s) \leq -1$. Let $\lambda_1, \lambda_2, ..., \lambda_d$ be the roots of the polynomial $Q_I(s)$ counting multiplicity. We have

\[ \left| \frac{Q_I(s)}{Q_I(0)} \right| = \prod_{k=1}^d \left| \frac{s - \lambda_k}{\lambda_k} \right|. \]

For a fixed $s$ and $\text{Re} \lambda_k \geq -1$ the term $|s - \lambda_k|/|\lambda_k|$ attains its maximum at $\lambda_k = -1$. Therefore, on the interval $[0, \varepsilon]$,

\[ \left| \frac{Q_I(s)}{Q_I(0)} \right| \leq (1 + \varepsilon)^d. \] (2.7)

From the estimate above,

\[ \int_0^\varepsilon s \left| \frac{P_I(s)}{Q_I(0)} \right|^2 ds \leq (1 + \varepsilon)^{2d} \int_0^\varepsilon s R_I(s) ds \] (2.8)

It will suffice then to find a constant $C_1 = C_1(\varepsilon, d)$ such that for any polynomial $p$ of degree at most $d - 1$

\[ |p(0)|^2 \leq C_1 \int_0^\varepsilon s |p(s)|^2 \frac{ds}{\varepsilon}. \] (2.9)

Note that if we do not care about the constant $C(d)$ we can stop here: we just consider the space of polynomials of degree at most $d$ endowed with the norm $\|p\| := \varepsilon^{-1} \int_0^\varepsilon s |p(s)|^2 ds$ and the linear functional $p \mapsto p(0)$. Since any linear functional on a finite-dimensional normed space is bounded, we immediately get (2.9).
If we want to estimate the constant $C(d)$, some extra work is needed. First, making the change of variables $x = 2s/\varepsilon$ we can see that (2.9) is equivalent (with the same constant $C_1$) to

$$|p(0)|^2 \leq C_1 \frac{\varepsilon}{4} \int_0^2 x |p(x)|^2 \, dx$$

or, equivalently, to the estimate

$$|p(1)|^2 \leq C_1 \frac{\varepsilon}{4} \int_{-1}^1 (1 - x) |p(x)|^2 \, dx \tag{2.10}$$

for all polynomials $p$, deg $p \leq d - 1$.

Consider the Jacobi polynomials $P_n^{(1,0)}$ which are orthogonal polynomials with respect to the weight $w(x) = (1 - x) = (1 - x)^1 (1 + x)^0$. Denote by $J_n^{(1,0)}$ the normalized Jacobi polynomials, $J_n^{(1,0)} := \|P_n^{(1,0)}\|^{-1}_{L^2(w)} P_n^{(1,0)}$.

Since $P_n^{(1,0)}(1) = n + 1$ and $\|P_n^{(1,0)}\|^2_{L^2(w)} = 2/(n + 1)$, we have that

$$J_n^{(1,0)}(1)^2 = (n + 1)^3/2. \tag{2.11}$$

Writing $p = \sum_{n=0}^{d-1} c_n J_n^{(1,0)}$ we get that

$$\int_{-1}^1 (x - 1)^2 (P(x))^2 \, dx = \|P\|^2_{L^2(w)} = \sum_{n=0}^{d-1} |c_n|^2$$

and that by (2.11)

$$P(1) = \sum_{n=0}^{d-1} c_n \frac{(n + 1)^3/2}{\sqrt{2}}.$$

From Cauchy–Schwarz,

$$|p(1)|^2 \leq \left( \sum_{n=0}^{d-1} |c_n|^2 \right)^{1/2} \left( \sum_{n=0}^{d-1} \frac{(n + 1)^3}{2} \right) = \frac{1}{8} d^2 (d + 1)^2 \|P\|^2_{L^2(w)}.$$

Comparing this with (2.10) we can see that (2.10) and consequently (2.9) hold with

$$C_1 = C_1(\varepsilon, d) = \varepsilon^{-1} d^2 (d + 1)^2/2.$$

From (2.9) and (2.8)

$$R_I(0) \leq C(\varepsilon, d) \frac{1}{\varepsilon} \int_0^\varepsilon s R_I(s) \, ds,$$

with $C(\varepsilon, d) = \varepsilon^{-1} d^2 (d + 1)^2 (1 + \varepsilon)^{2d}/2$.

For $\varepsilon = 1/(2d)$, we have indeed that $C(d) = d^3 (d + 1)^2 (1 + \frac{1}{2d})^{2d} \leq d^3 (d + 1)^2$.

3. Verifying properties of $B_s$

It remains to show that $B_s$ satisfies the Bellman function properties. The range property is proved in the following proposition:

Lemma 3.1. For $B_s$ defined above in (2.4), $B_s(I) \leq F_I$.
Proof. Let \( e \in \mathbb{F}^n \). Since \( W \) is self-adjoint, an application of the Cauchy-Schwarz inequality gives

\[
\left| \langle f, e \rangle \right| \leq \left( \int f^* f \right)^{1/2} \left( \int W^{1/2} e^* W^{1/2} e \right)^{1/2}.
\]

Therefore, recalling the notation \((2.1), (2.3)\), we get that for any vector \( e \),

\[
\frac{|\langle x_I, e \rangle|^2}{\langle \langle W \rangle_I e, e \rangle} \leq F_I.
\]

Using Lemma 3.2 below we can write

\[
\langle (\langle W \rangle_I + M_I)^{-1} x, x \rangle = \sup_{e \neq 0} \frac{|\langle x_I, e \rangle|^2}{\langle \langle W \rangle_I e, e \rangle} \leq \sup_{e \neq 0} \frac{|\langle x_I, e \rangle|^2}{\langle \langle W \rangle_I e, e \rangle} \leq F_I.
\]

which means exactly that \( B_s(I) \geq 0 \).

\[\square\]

Lemma 3.2. Let \( A \geq 0 \) be an invertible operator in a Hilbert space \( \mathcal{H} \). Then for any vector \( x \in \mathcal{H} \)

\[
\langle A^{-1} x, x \rangle = \sup_{e \in \mathcal{H} : e \neq 0} \frac{|\langle x, e \rangle|^2}{\langle A e, e \rangle}
\]

Proof. By definition,

\[
\langle A^{-1} x, x \rangle = \| A^{-1/2} x \|^2
\]

\[
= \sup_{a \in \mathcal{H} : \| a \| \neq 0} \frac{\langle A^{-1/2} x, a \rangle^2}{\| a \|^2}
\]

\[
= \sup_{a \in \mathcal{H} : \| a \| \neq 0} \frac{|\langle x, A^{-1/2} a \rangle|^2}{\| a \|^2}.
\]

Making the change of variables \( a = A^{1/2} e \) we conclude

\[
\langle A^{-1} x, x \rangle = \sup_{e \in \mathcal{H} : \| e \| \neq 0} \frac{|\langle x, e \rangle|^2}{\langle A e, e \rangle}.
\]

\[\square\]

The main estimate \((2.5)\) is the consequence of the following lemma:

Lemma 3.3. Let \( \mathcal{H} \) be a Hilbert space. For \( x \in \mathcal{H} \) and for \( U \) being a bounded invertible positive operator in \( \mathcal{H} \) define

\[
\phi(U, x) := \langle U^{-1} x, x \rangle_{\mathcal{H}}.
\]

Then the function \( \phi \) is convex, and, moreover, if

\[
x_0 = \sum_k \theta_k x_k, \quad \Delta U := U_0 - \sum_k \theta_k U_k
\]

where \( 0 \leq \theta_k \leq 1, \sum_k \theta_k = 1 \), then

\[
\sum_k \theta_k \phi(U_k, x_k) - \phi(U_0, x_0) \geq \langle U_0^{-1} \Delta U U_0^{-1} x_0, x_0 \rangle_{\mathcal{H}}
\]

\[\text{(3.2)}\]
To see that this lemma implies (2.5), fix $s > 0$. Denoting

$$U^s_I = \langle W \rangle_I + sM_I, \quad x_I = \langle W^{1/2}f \rangle_I,$$

we see that

$$B_s(I) = \phi(U^s_I, x_I).$$

Let $I_k, k \geq 1$ be the children of $I$, and let $\theta_k = |I_k|/|I|$. Notice that $\langle W \rangle_I = \sum_k \theta_k \langle W \rangle_{I_k}$, $M_I = \sum_k \theta_k M_{I_k} + s\langle W \rangle_I A_I \langle W \rangle_I$, so

$$U^s_I - \sum_k \theta_k U_{I_k} =: \Delta U^s = s\langle W \rangle_I A_I \langle W \rangle_I.$$

Therefore, applying Lemma 3.3 with $U_0 = U^s_I$, $x_0 = x_I$, $U_k = U^s_{I_k}$, $x_k = x_{I_k}$, $\Delta U = \Delta U^s$ we get (3.2), that translates exactly to the estimate (2.5).

**Proof of Lemma 3.3.** The function $\phi$ and the right hand side of (3.2) are invariant under the change of variables

$$x \mapsto U_0^{-1/2}x, \quad U \mapsto U_0^{-1/2}UU_0^{-1/2},$$

so it is sufficient to prove (3.2) only for $U_0 = I$.

In this case define function $\Phi(\tau)$, $0 \leq \tau \leq 1$ as

$$\Phi(\tau) = \sum_k \theta_k \left( (I + \tau \Delta U_k)^{-1} (x_0 + \tau \Delta x_k), (x_0 + \tau \Delta x_k) \right)_H - \langle x_0, x_0 \rangle_H,$$

where $\Delta x_k = x_k - x_0$ and $\Delta U_k = U_k - U_0 = U_k - I$. Using the power series expansion of $(I + \tau \Delta U)^{-1}$ we get

$$\Phi(\tau) = \tau \left( 2 \sum_k \theta_k \langle \Delta x_k, x_0 \rangle_H - \sum_k \theta_k \langle \Delta U_k x_0, x_0 \rangle \right)$$

$$+ \tau^2 \left( \sum_k \theta_k \langle \Delta U_k^2 x_0, x_0 \rangle + \sum_k \theta_k \langle \Delta U_k, \Delta x_k \rangle - 2 \sum_k \theta_k \langle \Delta U_k x_0, \Delta x_k \rangle_H \right) + o(\tau^2).$$

Notice that $\sum_k \theta_k \Delta x_k = \sum_k \theta_k (x_k - x_0) = 0$ and also $\sum_k \theta_k \Delta U_k = -\Delta U$. Hence

$$\Phi(\tau) = \tau \langle \Delta U x_0, x_0 \rangle + \tau^2 \sum_k \theta_k \left( \|\Delta U_k x_0\|^2 + \|\Delta x_k\|^2 - 2\langle \Delta U_k x_0, \Delta x_k \rangle \right) + o(\tau^3).$$

(3.4)

Using the above formula for $x_0 = (x_1 + x_2)/2$, $U_0 = (U_1 + U_2)/2$ (so $\Delta U = 0$) we get that the second differential of $\phi$ at $U = I$ is non-negative (the function $\phi$ is clearly analytic, so all the differentials are well defined).

The change of variables (3.3) implies that the second differential of $\phi$ is nonnegative everywhere. In particular, this implies that $\Phi''(\tau) \geq 0$, so $\Phi$ is convex.

Returning to the general choice of arguments $U, x$, we can see from (3.4) that

$$\Phi'(0) = \langle \Delta U x_0, x_0 \rangle_H.$$

Since $\Phi$ is convex and $\Phi(0) = 0$,

$$\Phi(1) \geq \Phi'(0) = \langle \Delta U x_0, x_0 \rangle_H.$$
REFERENCES

[1] K. Bickel, B. Wick, A study of the matrix Carleson embedding theorem with applications to sparse operators, arXiv:1503.06493

[2] K. Bickel, S. Petermichl, B. Wick Bounds for the Hilbert Transform with Matrix $A_2$ Weights, arXiv:1402.3886 [math.CA].

[3] J. Isralowitz, H. Kwon, S. Pott A matrix weighted $T_1$ theorem for matrix kernelled Calderon-Zygmund operators, arXiv:1401.6570

[4] R. Kerr Martingale transforms, the dyadic shift and the Hilbert transform: a sufficient condition for boundedness between matrix weighted spaces, arXiv:0906.4028

[5] F. Nazarov, G. Pisier, S. Treil, A. Volberg, Sharp estimates in vector Carleson imbedding theorem and for vector paraproducts, J. Reine Angew. Math. 542 (2002), 147–171.

[6] F. Nazarov, S. Treil, and A. Volberg, Two weight inequalities for individual Haar multipliers and other well localized operators, Math. Res. Lett. 15 (2008), no. 3, 583–597.

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

Department of Mathematics, Brown University, Providence, RI 02912, USA
E-mail address: amalia@math.brown.edu

Department of Mathematics, Brown University, Providence, RI 02912, USA
E-mail address: treil@math.brown.edu