ABSTRACT. It is well-known that dyadic martingale transforms are a good model for Calderón-Zygmund singular integral operators. In this paper we extend some results on weighted norm inequalities to vector-valued functions. We prove that, if $W$ is an $A_2$ matrix weight, then the weighted $L^2$-norm of a Calderón-Zygmund operator with cancellation has the same dependence on the $A_2$ characteristic of $W$ as the weighted $L^2$-norm of an appropriate matrix martingale transform. Thus the question of the dependence of the norm of matrix-weighted Calderón-Zygmund operators on the $A_2$ characteristic of the weight is reduced to the case of dyadic martingales and paraproducts. We also show a slightly different proof for the special case of Calderón-Zygmund operators with even kernel, where only scalar martingale transforms are required. We conclude the paper by proving a version of the matrix-weighted Carleson Embedding Theorem.

Our method uses a Bellman function technique introduced by S. Treil to obtain the right estimates for the norm of dyadic Haar shift operators. We then apply the representation theorem of T. Hytönen to extend the result to general Calderón-Zygmund operators.

1. INTRODUCTION

In the 1970’s, R.A. Hunt, B. Muckenhoupt and R.L. Wheeden [3] and R.R. Coifman and C. Fefferman [6] showed that a Calderón-Zygmund singular integral operator is bounded on the weighted space $L^p(w)$ if and only if the scalar weight $w$ belongs to the so-called $A_p$ class. For the last two decades, an important open problem in Harmonic Analysis was to characterize the dependence of the operator norm on the $A_p$ characteristic, $[w]_{A_p}$, of the weight. For $p=2$ this dependence was conjectured to be linear in $[w]_{A_2}$; the problem has become known as the $A_2$ conjecture. The first step was taken by J. Wittwer [20], who proved the $A_2$ conjecture for dyadic martingale transforms. Using Bellman function techniques, S. Petermichl and A. Volberg [19] showed the conjecture for the Beurling-Ahlfors transform. It took a few more years until the $A_2$ conjecture was proved for the Hilbert transform by S. Petermichl (see [17]). The conjecture was finally settled for general Calderón-Zygmund operators in 2010 by T. Hytönen [10]. The main ingredient in his proof is the pointwise representation of a general Calderón-Zygmund operator as a weighted average over an infinite number of randomized dyadic systems of some simpler operators (called dyadic Haar shifts) in such a way that the estimates for the dyadic Haar shifts depend polynomially on the complexity.

A natural problem is to try to extend these results to vector-valued functions. S. Treil and A. Volberg introduced the correct definition of a matrix $A_p$ weight (see [21]). M. Goldberg [6], F. Nazarov and S. Treil [14] and A. Volberg [25] showed that certain Calderón-Zygmund operators are bounded on $L^p(W)$ when $1 < p < \infty$ if $W$ is a matrix $A_p$ weight. However, the sharp dependence of the norm of a Calderón-Zygmund operator on the $A_2$ characteristic of $W$ is unknown even for the martingale transform. In a recent paper, K. Bickel, S. Petermichl and B. Wick [2] modified a scalar argument to obtain that for the Hilbert and martingale transforms this dependence is no worse than $[W]^{3/2}_{A_2} \log [W]_{A_2}$. This has very recently been improved to $[W]^{3/2}_{A_2}$, or more precisely, $[W]^{1/2}_{A_2} [W^{1/2} (W^{-1})^{1/2}]_{A_2}$, for all Calderón-Zygmund operators [16].

Even more recently, T. Hytönen, S. Petermichl and A. Volberg [11] proved the sharp linear upper bound $[W]_{A_2}$ for the matrix-weighted square function, which can be understood as an average of the matrix martingale transforms we consider. This raises the hope that the expected sharp linear bound for matrix martingale transforms in terms of $[W]_{A_2}$ may now come into reach.

2010 Mathematics Subject Classification. 42B20, 60G46, 46B09, 46B28, 26B25.

Key words and phrases. Calderón-Zygmund operator, matrix $A_2$ weights, weighted $L^2$ spaces, martingale transform, Bellman function, dyadic Haar shift, Carleson embedding theorem, Schur multiplier.
In this paper we prove that the norms of all Calderón-Zygmund singular integrals with cancellation have the same dependence on $[W]_{A_2}$ as the matrix martingale transforms (we denote this dependence by $N([W]_{A_2})$). The $A_2$ conjecture for matrix-weighted spaces is thus reduced to the case of dyadic martingale transforms and of the paraproducts. The proof follows S. Treil’s approach for the proof of the linear $A_2$ bound in the scalar case (see [23]). The main challenge here is the adaptation of the Bellman function to the matrix case, where convexity properties are much more difficult than in the scalar setting. Using Hytönen’s representation of a Calderón-Zygmund operator, it is enough to obtain the right estimate for the dyadic Haar shift operators. Since we want to obtain the same bound in terms of $[W]_{A_2}$ for the norm of dyadic Haar shifts, we have to use the martingale transform only once. We will decompose a dyadic Haar shift of complexity $k$ into $k$ “slices” that can be seen as martingale transforms. The main idea is to linearize the norm of these slices and then use the Bellman function to estimate each summand. In order to do this, we start with a standard dyadic martingale of points from the domain of the Bellman function, where at each point we have two choices with equal probability. We will then modify the martingale, preserving the initial point and the endpoints, and probabilities. From the starting point, instead of going to the next level in the standard martingale, we move with probabilities $1/2$ to two new points that are “far enough” from the initial point, but also “almost averages” of the endpoints. We can still move from these new points to the endpoints, this time using a modified dyadic martingale, where at each point we have two choices with “almost equal” probability. This new martingale is constructed in such a way that the probabilities of moving from the starting point to the endpoints are still equal, as in the case of the standard martingale. Although we have used probabilistic terms, the formal proof involving the Bellman function is elementary.

The paper is organized as follows: in Section 2 we recall the necessary definitions and results that we are using. Then we state our main result (Theorem 2.2) and show that it is enough to obtain a corresponding estimate for dyadic Haar shift operators, which is the content of Theorem 2.3. In Section 3 we use the boundedness of the martingale transform to relate the norm of a dyadic Haar shift to an expression that will be controlled by the Bellman function. Section 4 contains the definition of the Bellman function associated to our problem and the description of its properties. In Section 5 we formulate and prove the main technical result of the paper, which is inspired by [23]. In Section 6 we show how the main estimate from the previous section is used to conclude the proof of Theorem 2.3. In the following section we prove a similar result for Calderón-Zygmund singular integrals with even kernel, this time using the same martingale transform as in the scalar case. We finish with a further application of our Bellman function argument, namely a matrix-weighted Carleson Embedding Theorem which holds with constants independent of the dimension and the weight. This is, however, not the simple generalization of the usual weighted Carleson Embedding Theorem in [15].

2. Definitions and statement of the main results

In this section, we recall some well-known notions and results that we are going to use later on.

2.1. Calderón-Zygmund operators. Let $\Delta = \{(x, x) : x \in \mathbb{R}^p\}$ be the diagonal of $\mathbb{R}^p \times \mathbb{R}^p$. We say that a function $K : \mathbb{R}^p \times \mathbb{R}^p \setminus \Delta \to \mathbb{C}$ is a standard Calderón-Zygmund kernel if there exists $\delta > 0$ such that

$$|K(x, y)| \leq \frac{C}{|x - y|^p},$$

$$|K(x, y) - K(x, z)| + |K(y, x) - K(z, x)| \leq C_\delta \frac{|y - z|^\delta}{|x - y|^{p+\delta}},$$

for all $x, y, z \in \mathbb{R}^p$ with $|x - y| > 2|y - z|$.

An operator $T$, defined on the class of step functions (which is dense in $L^2(\mathbb{R}^p)$), is called a Calderón-Zygmund operator on $\mathbb{R}^p$ associated to $K$, if it satisfies the kernel representation

$$Tf(x) = \int_{\mathbb{R}^p} K(x, y)f(y) \, dy, \quad x \notin \text{supp} \, f.$$
2.2. **Matrix $A_2$ weights.** For $d \geq 1$, the non-weighted Lebesgue space $L^2(\mathbb{R}^p)$ consists of all measurable functions $f : \mathbb{R}^p \to \mathbb{C}^d$ such that

$$
\|f\|_{L^2(\mathbb{R}^p)} := \left( \int_{\mathbb{R}^p} \|f(t)\|^2_{\mathbb{C}^d} \, dt \right)^{1/2} < \infty.
$$

We will also use the space $C^1_c(\mathbb{R}^p)$ of compactly supported, continuously differentiable functions $f : \mathbb{R}^p \to \mathbb{C}^d$.

Let $\mathcal{M}_d(\mathbb{C})$ be the space of $d \times d$ complex matrices. A matrix weight on $\mathbb{R}^p$ is a measurable locally integrable function $W : \mathbb{R}^p \to \mathcal{M}_d(\mathbb{C})$ whose values are almost everywhere positive definite. We define $L^2(W)$ to be the space of measurable functions $f : \mathbb{R}^p \to \mathbb{C}^d$ with norm

$$
\|f\|^2_{L^2(W)} = \int_{\mathbb{R}^p} \|W^{1/2}(t)f(t)\|^2_{\mathbb{C}^d} \, dt = \int_{\mathbb{R}^p} \langle W(t)f(t), f(t) \rangle \, dt < \infty.
$$

It is well-known that the dual of $L^2(W)$ can be identified with $L^2(W^{-1})$, where the duality between these two spaces is given by the unweighted standard inner product.

We say that a matrix weight $W$ satisfies the matrix $A_2$ Muckenhoupt condition if

$$
[W]_{A_2} := \sup_Q \left\| \frac{1}{|Q|} \int_Q W(t) \, dt \right\|^{1/2} \left( \frac{1}{|Q|} \int_Q W^{-1}(t) \, dt \right)^{1/2} < \infty,
$$

where the supremum is taken over all cubes $Q \subset \mathbb{R}^p$, and $\| \cdot \|$ denotes the norm of the matrix acting on $\mathbb{C}^d$. The number $[W]_{A_2}$ is called the $A_2$ characteristic of the weight $W$. We say that a matrix weight $W$ satisfies the dyadic matrix Muckenhoupt condition $A^2_d$ on $\mathbb{R}^p$ or $\mathbb{R}$, if (1) is satisfied, but with the supremum now being taken only over dyadic cubes or intervals, respectively (see [24]).

2.3. **Dyadic setting.** Since we will reduce the proof of our main result to the case of functions defined on $\mathbb{R}$, we will only introduce the required notions in this setting. For the analogous definitions in the case of functions on $\mathbb{R}^p$, we refer the readers to [9].

The standard dyadic system in $\mathbb{R}$ is

$$
\mathcal{D}^0 := \bigcup_{j \in \mathbb{Z}} \mathcal{D}^0_j, \quad \mathcal{D}^0_j := \{2^{-j}([0,1) + k) : k \in \mathbb{Z} \}.
$$

Given a binary sequence $\omega = (\omega_j)_{j \in \mathbb{Z}} \in \{(0,1)\}^\mathbb{Z}$, a general dyadic system on $\mathbb{R}$ is defined by

$$
\mathcal{D}^\omega := \bigcup_{j \in \mathbb{Z}} \mathcal{D}^\omega_j, \quad \mathcal{D}^\omega_j := \mathcal{D}^0_j + \sum_{i > j} 2^{-i} \omega_i.
$$

When the particular choice of $\omega$ is not important, we will use the notation $\mathcal{D}$ for a generic dyadic system. We equip the set $\Omega := \{(0,1)\}^\mathbb{Z}$ with the canonical product probability measure $\mathbb{P}_\Omega$ which makes the coordinates $\omega_j$ independent and identically distributed with $\mathbb{P}_\Omega(\omega_j = 0) = \mathbb{P}_\Omega(\omega_j = 1) = 1/2$. We denote by $\mathbb{E}_\Omega$ the expectation over the random variables $\omega_j, j \in \mathbb{Z}$.

For an interval $I \in \mathcal{D}$, let $I^+$ and $I^-$ be the left and right children of $I$. The parent of $I$ will be denoted by $\bar{I}$. We will also use the notation $\mathcal{D}_n(I) := \{ J \in \mathcal{D} : J \subset I, |J| = 2^{-n}|I| \}$ for the collection of $n$-th generation children of $I$, where $|J|$ stands for the length of the interval $J$.

For any interval $I \in \mathcal{D}$, there is an associated Haar function defined by

$$
h_I = |I|^{-1/2}(\chi_{I^+} - \chi_{I^-}),
$$

where $\chi_I$ is the characteristic function of $I$.

For an arbitrary dyadic system $\mathcal{D}$, the Haar functions form an orthogonal basis of $L^2(\mathbb{R})$. Hence any function $f \in L^2(\mathbb{R})$ admits the orthogonal expansion

$$
f = \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_I.
$$

We denote the average of a locally integrable function $f$ on the interval $I$ by $\langle f \rangle_I := |I|^{-1} \int_I f(t) \, dt$.

Let $W$ be a matrix weight. For a sequence of $d \times d$ matrices $\sigma = \{\sigma_I\}_{I \in \mathcal{D}}$, we introduce the notation $\|\sigma\|_{\infty,W} = \sup_{I \in \mathcal{D}} \| \langle W \rangle_I^{1/2} \sigma_I \langle W \rangle_I^{-1/2} \|$.
For a sequence $\sigma$ such that $\|\sigma\|_{\infty,W} < \infty$, we define the martingale transform operator $T_\sigma$ by

$$T_\sigma f = \sum_{I \in \mathcal{D}} \sigma_I \langle f, h_I \rangle h_I.$$ 

If $W$ is a matrix $A_2$ weight, then the condition $\|\sigma\|_{\infty,W} < \infty$ is equivalent to the boundedness of $T_\sigma$ on $L^2(W)$ (see, e.g. Theorem 5.2 in [2] for an explicit statement; it is also contained in [24]). Such martingale transforms are considered a good model for Calderón-Zygmund singular integral operators.

A (cancellative) dyadic Haar shift on $\mathbb{R}$ of parameters $(m,n)$, with $m, n \in \mathbb{N}_0$, is an operator of the form

$$S f = \sum_{L \in \mathcal{D}} \sum_{I \in \mathcal{D}_n(L)} c^L_{I,J} \langle f, h_I \rangle h_J,$$

where $|c^L_{I,J}| \leq \sqrt{|I||J|} = 2^{-(m+n)/2}$ and $f$ is any locally integrable function. The number $k := \max(m,n) + 1$ is called the complexity of the Haar shift.

For $0 \leq j \leq k - 1$ we introduce the notation $\mathcal{L}_j := \{ I \in \mathcal{D} : |I| = 2^{j+k}, t \in \mathbb{Z} \}$, and define the slice $S_j$ by

$$S_j f = \sum_{L \in \mathcal{L}_j} \sum_{I \in \mathcal{D}_n(L)} c^L_{I,J} \langle f, h_I \rangle h_J.$$ 

We can thus decompose $S$ as $S = \sum_{j=0}^{k-1} S_j$. The key point is now that the operators $S_j$ can be seen as martingale transforms when we are moving $k$ units of time at once, so it is possible to apply the Bellman function for dyadic martingale transforms.

Following the approach in [24], one can show that it is enough to consider only dyadic Haar shifts on a dyadic system in $\mathbb{R}$. The following construction works for general dyadic systems, but for convenience we will assume that we are dealing with the standard one. This reduction is obtained by “arranging” the dyadic cubes on the real line.

More precisely, for a dyadic cube $Q$ in $\mathbb{R}^p$, we choose a dyadic interval $I$ such that $|I| = |Q|$ (this interval $I$ will correspond to the cube $Q$). We then split $Q$ into two congruent parallelepipeds by dividing one of its sides into two parts, and then pick a bijection between these two parallelepipeds and the children of $I$. By dividing a long side, we split each parallelepiped into two congruent ones, and then choose a bijection between the four parallelepipeds and the children of the two intervals from the previous step. After $p$ divisions we obtain a bijection between the children of $Q$ and the intervals $J \in \mathcal{D}_p(I)$. The intervals $J \in \mathcal{D}_n(I), 1 \leq n < p$, correspond to some “almost children” $R$ of $Q$, where by an “almost child” of $Q$ we mean a parallelepiped with some sides coinciding with the sides of $Q$, and the other sides being half of the corresponding sides of $Q$.

This construction can also be done in the opposite direction. If $I$ is the parent of the interval $I$, and $Q$ is the grandparent of $Q$ of order $p$, by the above method we obtain a bijection $\Phi$ between the children and “almost children” of $Q$, and the intervals $J \in \mathcal{D}_n(I), 1 \leq n < p$, such that $\Phi(Q) = I$. To make sure that $\Phi(Q) = I$, at each division we have to assign to the "almost child" containing $Q$ the dyadic interval of appropriate length that contains $I$.

A locally integrable function $f$ on $\mathbb{R}^p$ will thus be transferred to a locally integrable function $g$ on $\mathbb{R}$ such that $\langle f, Q \rangle = \langle g, I \rangle$, for all $Q$ and $I$ with $\Phi(Q) = I$.

We now look at the differences that arise when using this reduction. If $S$ is a dyadic Haar shift (or one of its slices) of complexity $k$ in $\mathbb{R}^p$, then its model in $\mathbb{R}$ will be a Haar shift of complexity $kp$.

If $W$ is a matrix $A^2_2$ weight on $\mathbb{R}^p$, then the $A^2_2$ characteristic of the corresponding weight on $\mathbb{R}$ is $\sup_R \| W^{1/2} W^{-1/2} \|$, where the supremum is taken over all dyadic cubes in $\mathbb{R}^p$ and all their “almost children”. If $R$ is an “almost child” of a cube $Q$, then

$$\int_R W(t) \, dt \leq \int_Q W(t) \, dt, \quad \int_R W^{-1}(t) \, dt \leq \int_Q W^{-1}(t) \, dt,$$
and \(|R| \geq 2^{-p+1}|Q|\). We thus have
\[
\| (W)_{1,R}^{1/2} (W^{-1})_{1,R}^{1/2} \|^{2} = \| (W)_{1,R}^{1/2} R_{1,R}^{1/2} \| \leq \| (W)_{1,R}^{1/2} R_{1,R}^{1/2} (W^{-1})_{Q,R}^{1/2} \| \leq 2^{p-1} \| (W)_{R}^{1/2} (W^{-1})_{Q,R}^{1/2} \| \leq 2^{(p-1)} \| (W)_{R}^{1/2} (W^{-1})_{Q,R}^{1/2} \|^{2}.
\]
Thus, after the transfer to the real line, the \(A^{2}_{d}\) characteristic \([W]_{A^{2}_{d}}\) of the weight increases at most by a factor of \(2^{2(p-1)}\).

We are using the following representation of a Calderón-Zygmund operator in terms of dyadic Haar shifts.

**Theorem 2.1** (Hytönen [3]). Let \(T\) be a Calderón-Zygmund operator on \(\mathbb{R}^{p}\) which satisfies the standard kernel estimates, the weak boundedness property \(\| \langle T_{Q} \chi_{Q} \rangle \| \leq C|Q|\) for all cubes \(Q\), and the vanishing paraproduct conditions \(T(1) = T^{*}(1) = 0\). Then it has an expansion, say for \(f, g \in C^{1}_{c}(\mathbb{R}^{p})\),
\[
\langle T f, g \rangle_{L^{2}(\mathbb{R}^{p}), L^{2}(\mathbb{R}^{p})} = C \cdot E_{\Omega} \sum_{m,n=0}^{\infty} \tau(m,n) \langle S_{\omega}^{mn} f, g \rangle_{L^{2}(\mathbb{R}^{p}), L^{2}(\mathbb{R}^{p})},
\]
where \(C\) is a constant depending only on the constants in the standard estimates of the kernel \(K\) and the weak boundedness property, \(S_{\omega}^{mn}\) is a dyadic Haar shift in \(\mathbb{R}^{p}\) of parameters \((m,n)\) on the dyadic system \(D^{\omega}\), and \(\tau(m,n) \leq P(\text{max}\{m,n\})2^{-\delta \text{max}\{m,n\}}\), with \(P\) a polynomial.

We define the function \(N : [1, \infty) \to [1, \infty)\) by
\[
N(X) = \sup_{r} \| T_{r} \|_{L^{2}(W) \to L^{2}(W)},
\]
where the supremum is taken over all \(d \times d\) matrix \(A^{2}_{d}\) weights \(W\) with \([W]_{A^{2}_{d}} \leq X\) and all matrix sequences \(\{\sigma\}_{I \in D}\) with \(\|\sigma\|_{\infty,W} \leq 1\). It was shown in [2] that
\[
N(X) \leq (\log X)X^{3/2}.
\]

Here is our main result:

**Theorem 2.2.** Let \(W\) be a \(d \times d\) matrix \(A^{2}_{d}\) weight on \(\mathbb{R}^{p}\). Let \(K\) be a standard kernel and \(T\) be a Calderón-Zygmund operator on \(\mathbb{R}^{p}\) associated to \(K\). Suppose that \(T\) satisfies the weak boundedness property \(\| \langle T_{Q} \chi_{Q} \rangle \| \leq C|Q|\) for all cubes \(Q\), and the vanishing paraproduct conditions \(T(1) = T^{*}(1) = 0\). Then
\[
\| T \|_{L^{2}(W) \to L^{2}(W)} \leq C \cdot pdN(2^{2(p-1)}[W]_{A^{2}_{d}}) \leq C_{p} \cdot dN([W]_{A^{2}_{d}}),
\]
where \(C\) depends only on the constants in the standard estimates and the weak boundedness property, while \(C_{p}\) depends on \(C\) and \(p\).

The second inequality in the theorem is a simple consequence of (2), we therefore turn to the first inequality. It is enough to show a corresponding result for Haar shift operators and then use the representation theorem of T. Hytönen.

Let \(f, g \in C^{1}_{c}(\mathbb{R}^{p})\) (if \(W\) is a matrix \(A^{2}_{d}\) weight, this space is dense in both \(L^{2}(W)\) and \(L^{2}(W^{-1})\)). Since the duality between \(L^{2}(W)\) and \(L^{2}(W^{-1})\) is the same as the standard duality on \(L^{2}(\mathbb{R}^{p})\), by Theorem 2.1 we have the representation
\[
\langle T f, g \rangle_{L^{2}(W), L^{2}(W^{-1})} = C \cdot E_{\Omega} \sum_{m,n=0}^{\infty} \tau(m,n) \langle S_{\omega}^{mn} f, g \rangle_{L^{2}(W), L^{2}(W^{-1})}
\]
and therefore
\[
\| T \|_{L^{2}(W) \to L^{2}(W)} \leq C \sum_{m,n=0}^{\infty} \tau(m,n) \| S_{\omega}^{mn} \|_{L^{2}(W) \to L^{2}(W)}. \]
We will show the estimate
\[
\| S_{\omega}^{mn} \|_{L^{2}(W) \to L^{2}(W)} \lesssim (\text{max}\{m,n\}+1)pdN(2^{2(p-1)}[W]_{A^{2}_{d}}).
\]
for all dyadic Haar shifts $S_{mn}^n$ on $\mathbb{R}^p$ with parameters $(m, n)$, which ensures the convergence of the series and completes the proof of Theorem 2.2. Using the above transference result, we can restrict ourselves to Haar shifts in $\mathbb{R}$. This is the content of the following theorem.

**Theorem 2.3.** Let $S$ be a dyadic Haar shift on $\mathbb{R}$ of complexity $k \geq 1$ and $W$ be a matrix $A_d^d$ weight. Then

$$
\|S\|_{L^2(W) \rightarrow L^2(W)} \leq c \cdot kdN([W]_{A_d^d}),
$$

where $c$ is an absolute, positive constant.

### 3. Reduction of the Proof of Theorem 2.3

Let $W$ be a $d \times d$ matrix $A_d^d$ weight on $\mathbb{R}$. For each $I \in \mathcal{D}$, choose an orthonormal basis of eigenvectors $B_I = \{e_{I,1}^I, e_{I,2}^I, \ldots, e_{I,d}^I\}$ of $(W)_I$, and let $P_I^j, 1 \leq i \leq d$, be the corresponding orthogonal projection onto the span of $e_{I,i}^I$.

Using the definition of the martingale transform operator $T_\sigma$ and the fact that each $(W)_I$ commutes with the $P_I^j$'s, we have for $f \in L^2(W), g \in L^2(W^{-1})$,

$$
\sum_{i=1}^d \sum_{I \in \mathcal{D}} \left| \langle P_I^j(f, h_I), P_I^j(g, h_I) \rangle_{C^d} \right|
\leq d \cdot \sup_{\sigma} \sum_{I \in \mathcal{D}} \left| \langle \sigma_I(f, h_I), \langle g, h_I \rangle \rangle_{C^d} \right|
\leq d \cdot \sup_{\sigma} \|T_\sigma\|_{L^2(W) \rightarrow L^2(W)} \|f\|_{L^2(W)} \|g\|_{L^2(W^{-1})}
\leq d \cdot N([W]_{A_d^d}) \|f\|_{L^2(W)} \|g\|_{L^2(W^{-1})},
$$

where the supremum is now taken over all matrix sequences $\sigma = \{\sigma_I\}_{I \in \mathcal{D}}$ such that $\|\sigma\|_{\infty, W} \leq 1$. Notice that it would suffice to just take the $\sigma_I$'s which are diagonal in the basis $B_I$.

We can thus rewrite the estimate (3) above as

$$
\sum_{i=1}^d \sum_{I \in \mathcal{D}} \left| \langle P_I^j((f)_I^+, (g)_I^-), P_I^j((f)_I^-, (g)_I^+) \rangle_{C^d} \right| \cdot |I|
\leq 4 \sum_{i=1}^d \sum_{I \in \mathcal{D}} \left| \langle P_I^j(f, h_I), P_I^j(g, h_I) \rangle_{C^d} \right| \leq 4 \cdot N([W]_{A_d^d}) \|f\|_{L^2(W)} \|g\|_{L^2(W^{-1})}
$$

for all $f \in L^2(W)$ and $g \in L^2(W^{-1})$.

Since $S$ is a Haar shift operator of complexity $k$, it has the form

$$
Sf = \sum_{L \in \mathcal{D}} \sum_{I \in \mathcal{D}_n(L)} \sum_{J \in \mathcal{D}_n(L)} c_{I,J}^L(f, h_I)h_J,
$$

where $|c_{I,J}^L| \leq \sqrt{|I|} \sqrt{|J|} = 2^{-(m+n)/2}$.
Let $f \in L^2(W)$, $g \in L^2(W^{-1})$ and $0 \leq j \leq k - 1$ be fixed. For the slice $S_j$, we can write
\begin{align*}
\langle S_j f, g \rangle_{L^2(W), L^2(W^{-1})} &= \left\langle \sum_{I, J \in D_n(L)} c_{I, J}^L \langle f, h_I \rangle h_J, \sum_{I' \in D} (g, h_{I'}) h_{I'} \right\rangle_{L^2(W), L^2(W^{-1})} \\
&= \sum_{I, J \in D_n(L)} \sum_{I' \in D} c_{I, I'}^L \langle f, h_I \rangle \langle g, h_{I'} \rangle_{D} (h_J, h_{I'})_{D(R), D(R)} \\
&= \sum_{I, J \in D_n(L)} \sum_{I' \in D} c_{I, I'}^L \langle f, h_I \rangle \langle g, h_{I'} \rangle_{D} = \sum_{I, J \in D_n(L)} c_{I, J}^L \langle P_L^i(f, h_I), P_L^i(g, h_J) \rangle_{D} \\
&= \sum_{I, J \in D_n(L)} \sum_{I' \in D} c_{I, I'}^L \sum_{P \in D_{n}(L)} \sum_{P \subset I^+} P_L^i((f)_{P} - (f)_{L}) + \sum_{P \in D_{n}(L)} \sum_{P \subset I^-} P_L^i((f)_{P} - (f)_{L}) \\
&\hspace{1cm} \sum_{Q \in D_{n}(L)} \sum_{Q \subset J^+} P_L^i((g)_{Q} - (g)_{L}) + \sum_{Q \in D_{n}(L)} \sum_{Q \subset J^-} P_L^i((g)_{Q} - (g)_{L})_{C^d}.
\end{align*}

We therefore have
\begin{equation}
\langle S_j f, g \rangle_{L^2(W), L^2(W^{-1})} \leq \sum_{I, J \in D_n(L)} |I| \sum_{i=1}^d \sum_{P, Q \in D_n(L)} \left\langle P_L^i\left(\frac{(f)_{P} - (f)_{L}}{2^k}\right), P_L^i\left(\frac{(g)_{Q} - (g)_{L}}{2^k}\right) \right\rangle_{C^d}.
\end{equation}

4. The Bellman function

We are now going to define the Bellman function associated to our problem. Let $X > 1$, fix a dyadic interval $I_0$, and for $f \in C^d$, $F \in R$, $U \in M_d(C)$, $g \in C^d$, $G \in R$, $V \in M_d(C)$ satisfying
\begin{equation}
U, V > 0, I_0 \leq V^{1/2} U V^{1/2} \leq X \cdot I_d, \|V^{-1/2} f\|_{C^d}^2 \leq F, \|U^{-1/2} g\|_{C^d}^2 \leq G,
\end{equation}
define the function $B_X = B_{X_0} : C^d \times R \times M_d(C) \times C^d \times R \times M_d(C)$ by
\begin{equation}
B_X(f, F, U, g, G, V) := |I_0|^{-1} \sup_{I \subseteq I_0} \left\langle \langle f \rangle_{I^+} - \langle f \rangle_{I^+}, \langle g \rangle_{I^+} - \langle g \rangle_{I^+} \right\rangle_{C^d} \cdot |I|,
\end{equation}
where the supremum is taken over all functions $f, g : R \to C^d$ and matrix $A_2$ weights $W$ on $I_0$ such that
\begin{equation}
\langle f \rangle_{I_0} = f \in C^d, \quad \langle \|W^{-1/2} f\|_{C^d}^2 \rangle_{I_0} = F \in R, \quad \langle g \rangle_{I_0} = g \in C^d, \quad \langle \|W^{-1/2} g\|_{C^d}^2 \rangle_{I_0} = G \in R,
\end{equation}
and all sequences of $d \times d$ matrices $\sigma = \{\sigma_I\} \in D$ with $\|\sigma\|_{\infty, W} \leq 1$.

The Bellman function $B_X$ has the following properties:

(i) (Domain) The domain $D_X := \text{Dom} B_X$ is given by \([\mathfrak{S}].\) This means that for every tuple $(f, F, U, g, G, V)$ that satisfies \([\mathfrak{S}],\) there exist functions $f, g$ and a matrix weight $W$ such that \([\mathfrak{S}]\) holds, so the supremum is not $-\infty$. Conversely, if the variables $(f, F, U, g, G, V)$ are the corresponding averages of some functions $f, g$ and $W$, then they must satisfy condition \([\mathfrak{S}].\) Since the set \{$(U, V) \in M_d(C) \times M_d(C) : U, V > 0, I_0 \leq V^{1/2} U V^{1/2} \leq X \cdot I_d\}$ is not convex, the domain $D_X$ is not convex either.

(ii) (Range) $0 \leq B_X(f, F, U, g, G, V) \leq 4N(X) F^{1/2} G^{1/2}$ for all $(f, F, U, g, G, V) \in D_X$. 

(iii) (Concavity condition) Consider all tuples $A = (f, F, U, g, G, V)$, $A_+ = (f_+, F_+, U_+, g_+, G_+, V_+)$ and $A_- = (f_-, F_-, U_-, g_-, G_-, V_-)$ in $\mathcal{D}_X$ such that $A = (A_+ + A_-)/2$. For all such tuples, we have the following concavity condition:

$$B_X(A) \geq \frac{B_X(A_+) + B_X(A_-)}{2} + \sup_{\|\tau\| \leq 1} |\langle \tau(f_+ - f_-), g_+ - g_- \rangle|_{\mathbb{C}^d}|.$$

Here, the supremum is taken over all $d \times d$ matrices $\tau$ with $\|\tau\|_F := \|U^{1/2} \tau U^{-1/2}\| \leq 1$.

Let us now explain these properties of the function $B_X$. For any matrix weight $W$ and any interval $I$ we have $\langle W^{-1/2} W(t) \rangle_I \langle W^{-1/2} \rangle_I \geq I_d$, so $V^{1/2} U V^{1/2} \geq I_d$. The inequality $V^{1/2} U V^{1/2} \leq X \cdot I_d$ follows from the definition of the matrix $A_2$ Muckenhoupt condition. Conversely, for any positive definite matrices $U, V$ such that $I_d \leq V^{1/2} U V^{1/2} \leq X \cdot I_d$, we can find a matrix weight $W$ that satisfies (9). To see this, we construct a matrix weight $W$ that is constant on the children of $I_0$.

Given two matrices $U$ and $V$ as above, we want to find two positive definite matrices $W_1$ and $W_2$, such that

$$U = \frac{1}{2} (W_1 + W_2) \quad \text{and} \quad V = \frac{1}{2} (W_1^{-1} + W_2^{-1}).$$

We have $U = W_1 V W_2 = W_2 V W_1$, thus $W_2^{-1} = W_1 U^{-1} = U^{-1} W_1 V$. Let $M := U^{-1/2} W_1 U^{-1/2}$, $N := U^{-1/2} V^{-1/2} U^{-1/2}$, and notice that $N \leq I_d$. Then the matrices $M$ and $N$ commute:

$$N^{-1} M = (U^{1/2} V U^{1/2})(U^{-1/2} W_1 U^{-1/2}) = U^{1/2} (W_1 U^{-1}) U^{1/2} = (U^{-1/2} W_1 U^{-1/2})(U^{1/2} V U^{1/2}) = M N^{-1}.$$

Furthermore, $U = \frac{1}{2} (W_1 + W_2) = \frac{1}{2} (W_1 + U W_1^{-1} V^{-1})$, so $W_1 = \frac{1}{2} (W_1 U^{-1} W_1 + V^{-1})$. It follows that

$$M = \frac{1}{2} (U^{-1/2} W_1 U^{-1/2} + U^{-1/2} V^{-1/2} U^{-1/2}) = \frac{1}{2} (I^2 + N),$$

hence $M$ satisfies the quadratic equation $(M^2 - 2M + I_d) - (I_d - N) = 0$. Choosing $M = I_d + (I_d - N)^{1/2}$, we obtain

$$W_1 = U^{1/2} M U^{1/2} = U^{1/2}(I_d + (I_d - N)^{1/2}) U^{1/2},$$

and

$$W_2 = 2U - W_1 = U^{1/2}(I_d - (I_d - N)^{1/2}) U^{1/2}.$$
On the other hand, given a tuple $(f, F, U, g, G, V) ∈ D$ and a matrix weight $W$ satisfying $[5]$, we can always find two functions $f, g$ satisfying $[5]$. We first choose a function $φ : R → C^d$ such that

$$\int_0^1 φ(t) \, dt = 0, \quad \int_0^1 W(t) φ(t) \, dt = 0, \quad \frac{1}{|I_0|} \int_{I_0} \|W^{1/2}(t)φ(t)\|^2_{C^d} \, dt = 1,$$

and then set $f(t) := W^{-1}(t)F + (F - \|V^{-1/2}f\|^2)^{1/2}φ(t)$. It can be easily checked that this function has the required properties. A similar argument allows us to construct the function $g$.

Property (ii) follows from the definition of $B_X$ and the inequality $[11]$.

To prove the concavity condition, we consider three tuples $A, A_+, A_- \in D_X$ such that $A = (A_+ + A_-)/2$ and choose two functions $f, g$ and a matrix weight $W$ on $I_0$ so that

$$A_+ = \left(⟨f⟩_{I_0^2}, \langle∥W^{1/2}f∥_{C^d}^2⟩_{I_0^2}, \langle W⟩_{I_0^2}, \langle g⟩_{I_0^2}, \langle∥W^{1/2}g∥_{C^d}^2⟩_{I_0^2}, ⟨W⟩_{I_0}^{−1}\right).$$

Then

$$A = \frac{A_+ + A_-}{2} = \left(⟨f⟩_{I_0}, \langle∥W^{1/2}f∥_{C^d}^2⟩_{I_0}, \langle W⟩_{I_0}, \langle g⟩_{I_0}, \langle∥W^{1/2}g∥_{C^d}^2⟩_{I_0}, ⟨W⟩_{I_0}^{−1}\right)$$

is the vector of corresponding averages over $I_0$. The expression in the definition of $B_X(f, F, U, g, G, V)$, before taking the supremum, can be split into the average of the corresponding expressions for $B_X(f_+, f_-, U_+, g_+, G_+, V_+)$ and $B_X(f_-, f_-, U_-, g_-, G_-, V_-)$, plus the term

$$\sup_{σ_{I_0} : ∥U^1∥_{C^d} = 1} \left| ⟨π_{I_0}(f_+ - f_-), (g_+ - g_-)⟩_{C^d} \right|.$$
where $(\cdot, \cdot)_{HS}$ denotes the Hilbert-Schmidt (trace) inner product.

Using the identities

$$(V_1 + V_2)^{-1} = V_1^{-1} - V_1^{-1}V_2(V_1 + V_2)^{-1} \quad \text{and} \quad (V_1 + V_2)^{-1} = V_2^{-1} - V_2^{-1}V_1(V_1 + V_2)^{-1},$$

we get that

$$T_1 = \langle (V_1 + V_2)^{-1}, f_1 \otimes f_1 \rangle_{HS} = \|V_1^{-1/2}f_1\|_{C^2}^2 - \langle V_1^{-1}V_2(V_1 + V_2)^{-1}, f_1 \otimes f_1 \rangle_{HS},$$

and

$$T_3 = \langle (V_1 + V_2)^{-1}, f_2 \otimes f_2 \rangle_{HS} = \|V_2^{-1/2}f_2\|_{C^2}^2 - \langle V_2^{-1}V_1(V_1 + V_2)^{-1}, f_2 \otimes f_2 \rangle_{HS}.$$

Noting that $V_1^{-1}V_2(V_1 + V_2)^{-1} = (V_1 + V_2)^{-1}V_2V_1^{-1} > 0$ and writing $\hat{f}_2 = V_1V_2^{-1}f_2$, we find that

$$T_1 + T_2 + T_3 \leq -\langle V_1^{-1}V_2(V_1 + V_2)^{-1}, f_1 \otimes f_1 \rangle_{HS} - \langle V_1^{-1}V_2(V_1 + V_2)^{-1}, \hat{f}_2 \otimes \hat{f}_2 \rangle_{HS}$$

$$+ \langle V_1^{-1}V_2(V_1 + V_2)^{-1}, f_1 \otimes \hat{f}_2 \rangle_{HS} + \langle V_1^{-1}V_2(V_1 + V_2)^{-1}, \hat{f}_2 \otimes f_1 \rangle_{HS} + \text{F}_1 + \text{F}_2$$

$$= -\langle V_1^{-1}V_2(V_1 + V_2)^{-1}, (f_1 - \hat{f}_2) \otimes (f_1 - \hat{f}_2) \rangle_{HS} + \text{F}_1 + \text{F}_2 \leq \text{F}_1 + \text{F}_2.$$

This concludes the proof of our claim.

We now check that the set $C_0 := \{(U, V) \in M_d(\mathbb{C}) \times M_d(\mathbb{C}) : U, V > 0, I_d \leq V^{1/2}UV^{1/2}\}$ is convex. As before, it is enough to show that it is midpoint convex.

Let $(U_1, V_1), (U_2, V_2) \in C_0$. We have to prove that

$$I_d \leq \left(\frac{V_1 + V_2}{2}\right)^{1/2} \left(\frac{U_1 + U_2}{2}\right)^{1/2},$$

which is equivalent to

$$(V_1 + V_2)(U_1 + U_2)(V_1 + V_2) \geq 4(V_1 + V_2).$$

Since $(U_1, V_1), (U_2, V_2) \in C_0$, we have $U_1 \geq V_1^{-1}$ and $U_2 \geq V_2^{-1}$, so

$$(V_1 + V_2)(U_1 + U_2)(V_1 + V_2) \geq (V_1 + V_2)(V_1^{-1} + V_2^{-1})(V_1 + V_2)$$

$$= 3V_1 + 3V_2 + V_1V_2^{-1}V_1 + V_2V_1^{-1}V_2.$$

It is therefore enough to check that

$$V_1V_2^{-1}V_1 + V_2V_1^{-1}V_2 - V_1 - V_2 \geq 0,$$

which is the same as showing that

$$V_1^{1/2}V_2^{-1}V_1^{1/2} + V_1^{-1/2}V_2V_1^{-1/2}V_2^{-1}V_1^{-1/2}V_2V_1^{-1/2} = I_d - V_1^{-1/2}V_2V_1^{-1/2} \geq 0.$$

Let $T := V_1^{1/2}V_2^{-1}V_1^{1/2} > 0$. The previous inequality becomes $T + T^{-2} - I_d - T^{-1} \geq 0$, which is equivalent to $T^3 + I_d - T^2 - T \geq 0$. But $T^3 + I_d - T^2 - T = (T - I_d)(T + I_d)(T - I_d)$, and this is a positive semidefinite matrix since $T + I_d \geq 0$. This concludes the proof of the convexity of $C_0$.

To finish the proof of the lemma, we have to show that if $(U, V), (U_+, V_+), (U_-, V_-)$ are in the set $C_X := \{(U, V) \in M_d(\mathbb{C}) \times M_d(\mathbb{C}) : U, V > 0, V^{1/2}UV^{1/2} \leq X \cdot I_d\}$ and $(U, V) = \frac{1}{2}[\langle U_+, V_+ \rangle + \langle U_-, V_- \rangle]$, then for all $\theta \in [0, 1]$, the points $(U_\theta, V_\theta) = (\theta U_+ + (1 - \theta) U_- , \theta V_+ + (1 - \theta) V_-)$ belong to the set $C_X$.

Since $\theta \in [0, 1]$, we have $\theta U_+ \leq U_+$ and $(1 - \theta) U_- \leq U_-$, so $U_\theta \leq U_+ + U_- = 2U$; we also have $V_\theta \leq 2V$. It is then sufficient to show that $V_\theta^{1/2}(U_+ + U_-)V_\theta^{1/2} \leq 4XI_d$. But this is equivalent to $\|V_\theta^{1/2}(U_+ + U_-)^{1/2}\| \leq 4X$. All matrices that appear are positive definite, so $\|V_\theta^{1/2}(U_+ + U_-)^{1/2}\| = \|(U_+ + U_-)^{1/2}V_\theta(U_+ + U_-)^{1/2}\|$. Then again $\|V_\theta(U_+ + U_-)^{1/2}\| \leq 4X$ if and only if $(U_+ + U_-)^{1/2}V_\theta(U_+ + U_-)^{1/2} \leq 4XI_d$. We finally have

$$(U_+ + U_-)^{1/2}V_\theta(U_+ + U_-)^{1/2} = 2U^{1/2}V_\theta U^{1/2} \leq 4U^{1/2}VU^{1/2} \leq 4XI_d,$$

since $(U, V)$, and thus also $(V, U)$, are in the set $C_X$, so the proof of the lemma is complete. □
The following result is the main tool in the proof of Theorem 2.3.

**Lemma 5.1.** Let $X > 1$ and $B_X$ be a function satisfying properties (i)-(iii) from Section 6. Fix $k \geq 1$ and a dyadic interval $I_0$. For all $I \in D_n(I_0)$, $0 \leq n \leq k$, let the points $A_I = (f_I, F_I, U_I, g_I, G_I, V_I) \in \mathcal{D}_X = \text{Dom} B_X$ be given. Assume that the points $A_I$ satisfy the dyadic martingale dynamics, i.e. $A = (A_I + A_{I-})/2$, where $I^+$ and $I^-$ are the children of $I$. Let $B_{I_0} = \{e_{I_0}^1, e_{I_0}^2, \ldots, e_{I_0}^d\}$ be an orthonormal basis of eigenvectors of $U_{I_0}$ and for $1 \leq i \leq d$, let $P_{I_0}^i$ be the orthogonal projection onto the span of $e_{I_0}^i$. For $1 \leq i \leq d$ and $K, L \in D_k(I_0)$, we define the coefficients $\lambda_{KL}^i$ by

$$\lambda_{KL}^i := \left\langle P_{I_0}^i \left( f_{K} - f_{I_0} \right) / 2^k, P_{I_0}^i \left( g_{L} - g_{I_0} \right) / 2^k \right\rangle_{L^2}.$$

Then

$$\sum_{i=1}^d \sum_{K, L \in D_k(I_0)} |\lambda_{KL}^i| \leq c \cdot d \left( B_X(A_{I_0}) - 2^{-k} \sum_{I \in D_k(I_0)} B_X(A_I) \right),$$

where $c$ is a positive absolute constant and $X' = \frac{100}{9} X$.

**Proof.** For $1 \leq i \leq d$, we introduce the notation $\Lambda^i := (\lambda_{KL}^i)_{K, L \in D_k(I_0)}$. Assume for the moment that for each $i$, we can find a sequence $\{\alpha_{lI}^i\}_{I \in D_k(I_0)}$ such that $|\alpha_{lI}^i| \leq 1/4$ for all $I \in D_k(I_0)$, $\sum_{I \in D_k(I_0)} \alpha_{lI}^i = 0$, and

$$\left| \sum_{K, L \in D_k(I_0)} \alpha_{lI}^i \lambda_{KL}^i \right| \geq c \sum_{K, L \in D_k(I_0)} |\lambda_{KL}^i|.$$

For each $i$, we define $A_{I_0}^{\pm} = (f_{I_0}^{\pm}, F_{I_0}^{\pm}, U_{I_0}^{\pm}, g_{I_0}^{\pm}, G_{I_0}^{\pm}, V_{I_0}^{\pm})$ by

$$A_{I_0}^{\pm} := 2^{-k} \sum_{I \in D_k(I_0)} (1 \pm \alpha_{lI}^i) A_I = A_{I_0} \pm 2^{-k} \sum_{I \in D_k(I_0)} \alpha_{lI}^i A_I,$$

so $A_{I_0} = (A_{I_0}^{+} + A_{I_0}^{-})/2$.

The following notations and computations hold for every $1 \leq i \leq d$, so we fix such an $i$. For simplicity, we also drop the $i$ superscript until further notice.

For each $I \in D_k(I_0)$, let $a_{lI}^i := 1 \pm \alpha_{lI}^i$ and note that $3/4 \leq a_{lI}^i \leq 5/4$.

For $I \in D_n(I_0)$, $1 \leq n \leq k$, let us define

$$A_I^{\pm} := \left( \sum_{J \in D_k(I_0) \backslash J \subseteq I} a_{lJ}^i A_J \right) \left( \sum_{J \in D_k(I_0) \backslash J \subseteq I} a_{lJ}^i \right)^{-1}.$$

If $I \in D_k(I_0)$ we have $A_I^{+} = A_I^{-} = A_I$, where the $A_I$'s are the points from the statement of the lemma. The points $A_I^{\pm}$ are in the convex hull of the set $\{A_J : J \in D_k(I_0), J \subseteq I\}$. To address the lack of convexity of $\mathcal{D}_X$, we need an additional lemma:

**Lemma 5.2.** $A_I^{\pm} \in \mathcal{D}_{25/9X}$ for all $I \in D_n(I_0)$, $1 \leq n \leq k$.

**Proof.** of Lemma 5.2 Since the points $A_I^{\pm}$ are in the convex hull of the set $\{A_J : J \in D_k(I_0) \subseteq \mathcal{D}_{25/9X}\}$, and among the conditions that define $\mathcal{D}_{25/9X}$ only the constraint $V^{1/2} U V^{1/2} \leq \frac{25}{9} X \cdot I_d$ is not convex, we just have to check this condition.

Let us consider the $U$-coordinate of the points $A_I$. The maximal numerator is obtained when all coefficients $a_{lJ}^i$ are equal to $5/4$, and the minimal denominator is attained when $a_{lJ}^i = 3/4$ for all $J \in D_k(I_0)$, $J \subseteq I$. This implies that $U_I^{\pm} \leq \frac{1}{2} (\frac{5}{3})^{-1} U_I = \frac{3}{5} U_I$. Similarly, we also have $V_I^{\pm} \leq \frac{3}{5} V_I$. Using elementary properties of positive definite matrices, it follows that

$$\left( V_I^{\pm} \right)^{1/2} U_I^{\pm} \left( V_I^{\pm} \right)^{1/2} \leq \frac{5}{3} \left( V_I^{\pm} \right)^{1/2} U_I \left( V_I^{\pm} \right)^{1/2}.$$
and
\[
\| (V_I^\pm)^{1/2} U_I^\pm (V_I^\pm)^{1/2} \| \leq \frac{5}{3} \| (V_I^\pm)^{1/2} U_I (V_I^\pm)^{1/2} \| \leq \frac{5}{3} \| U_I^{1/2} V_I U_I^{1/2} \| \leq \frac{\sqrt{25}}{9} X,
\]
hence \((V_I^\pm)^{1/2} U_I^\pm (V_I^\pm)^{1/2} \leq \frac{\sqrt{25}}{9} X \cdot I_d\). This means that the points \( A_I^\pm \) belong to \( D_{25/9X} \).

Let \( \tilde{A}_I^\pm \) be the midpoints of the line segments with endpoints \( A_I^\pm \) and \( A_{I^-}^\pm \). We prove that \( \tilde{A}_I^\pm \in D_{25/9X} \).

As before, we have \( U_I^\pm \leq \frac{5}{3} U_I \) and \( V_I^\pm \leq \frac{5}{3} V_I \). Therefore,
\[
\tilde{U}_I^\pm = \frac{U_I^\pm + U_I^-}{2} \leq \frac{5}{3} \frac{U_I + U_I^-}{2} = \frac{5}{3} U_I
\]
and
\[
\tilde{V}_I^\pm \leq \frac{5}{3} V_I.
\]
It follows that \((V_I^\pm)^{1/2} \tilde{U}_I^\pm (V_I^\pm)^{1/2} \leq \frac{\sqrt{25}}{9} X \cdot I_d\), so the points \( \tilde{A}_I^\pm \) belong to \( D_{25/9X} \).

Applying Lemma 4.2, we conclude that the line segments with endpoints \( A_I^\pm \) and \( A_{I^-}^\pm \) are in \( D_{X'} \), where \( X' = 4 \frac{25}{9} X = \frac{100}{9} X \). This finishes the proof of Lemma 5.2

We continue with the proof of Lemma 5.1. For \( I \in D_n(I_0) \), \( 1 \leq n \leq k \), we define
\[
\theta_I^\pm := \left( \sum_{J \subseteq I} a_J^\pm \right)^{-1} \left( \sum_{J \subseteq I} a_J^\pm \right).
\]

It is easy to see that \( 3/10 \leq \theta_I^\pm \leq 5/6 \) and
\[
\theta_I^+, + \theta_I^- = 1, \quad A_I^\pm = A_I^+, A_{I^+} + A_{I^-}.
\]

The last equality means that the point \( A_I^\pm \) is on the line segment with endpoints \( A_I^+ \) and \( A_{I^-} \), and similarly for \( A_{I^-} \). \( \theta_I^+ \) and \( \theta_I^- \) represent the probabilities of moving from the points \( A_I^\pm \) to \( A_I^+ \) and \( A_{I^-} \), respectively.

Since \( 13 \)
\[
| \langle P_{I_0} (f^+ - f^-), P_{I_0} (g^+ - g^-) \rangle \rangle | = \left| \left( 2 \cdot 2^{-k} \sum_{I \in D_n(I_0)} \alpha_I P_{I_0} (f_I) \right) \cdot 2 \cdot 2^{-k} \sum_{I \in D_n(I_0)} \alpha_I P_{I_0} (g_I) \right| \leq 4 \left| \left< P_{I_0} (f^+ - f_{I_0}), P_{I_0} (g^+ - g_{I_0}) \right> \right| \leq \frac{1}{4} \left( B_{X'}(A_{I_0}) - \frac{B_{X'}(A_{I_0}^+) + B_{X'}(A_{I_0}^-)}{2} \right).
\]

From the concavity of the function \( B_{X'} \), and \( 14 \) it follows that
\[
B_{X'}(A_{I_0}^+) \geq \theta_I^+, B_{X'}(A_{I_0}^+) + \theta_I^- B_{X'}(A_{I_0}^-).
\]

Applying now this inequality to \( I \in D_n(I_0) \), \( 0 \leq n \leq k - 1 \), and taking into account that
\[
\prod_{I \in D_n(I_0)} \theta_I^\pm = a_I^\pm \left( \sum_{J \subseteq I} a_J^\pm \right)^{-1} = 2^{-k} a_I^\pm
\]
for all \( I \in D_k(I_0) \), we obtain the estimate
\[
B_{X'}(A_{I_0}^+) \geq 2^{-k} \sum_{I \in D_n(I_0)} a_I^\pm B_{X'}(A_I).
\]
Since $a_i^+ + a_i^- = 2$ when $I \in \mathcal{D}_k(I_0)$, substituting the previous inequality in (15) gives

\[
\left| \langle P_{I_0}(f^\pm - f_{I_0}), P_{I_0}(g^\pm - g_{I_0}) \rangle \right| \leq \frac{1}{4} \left( B_{X^*}(A_{I_0}) - 2^{-k} \sum_{I \in \mathcal{D}_k(I_0)} B_{X^*}(A_I) \right).
\]

We are now ready to obtain the conclusion of the lemma. By (12) and the fact that \( \sum_{I \in \mathcal{D}_k(I_0)} \alpha_I = 0 \), we have the estimate

\[
eq \frac{d}{4} \left( B_{X^*}(A_{I_0}) - 2^{-k} \sum_{I \in \mathcal{D}_k(I_0)} B_{X^*}(A_I) \right),
\]

where the last inequality follows from (16). This completes the proof of Lemma \ref{lemma:matrix} under the assumption (12).

For each $1 \leq i \leq d$, the matrix $A^i$ has complex rank 1. Dropping again the $i$ superscript, there exist $m = (m_K), n = (n_L) \in \mathbb{C}^{2^k}$ such that $\lambda_{KL} = m_K n_L = (m_{KL}^1 + im_{KL}^2)(n_{KL}^1 + in_{KL}^2)$, for every $K, L \in \mathcal{D}_k(I_0)$. We then have

\[
\sum_{K, L \in \mathcal{D}_k(I_0)} |\lambda_{KL}| \leq \sum_{K, L \in \mathcal{D}_k(I_0)} (|m_{KL}^1|^2 + |m_{KL}^2|^2)(|n_{KL}^1|^2 + |n_{KL}^2|^2).
\]

Without loss of generality, we may assume that $\sum_{K \in \mathcal{D}_k(I_0)} |m_{KL}^1|^2$ is the maximum of the four sums in the above right hand side. By an application of K. Ball’s “multiple Hahn-Banach Theorem” (II, Theorem 7), or alternatively an elementary functional analysis argument (see \cite{23}, Theorem 6.2 and Lemma 6.3), we can find a real-valued sequence $\{\alpha_I\}_{I \in \mathcal{D}_k(I_0)}$ such that $|\alpha_I| \leq 1/4$ for all $I \in \mathcal{D}_k(I_0), \sum_{I \in \mathcal{D}_k(I_0)} \alpha_I = 0$, and

\[
\left| \sum_{K \in \mathcal{D}_k(I_0)} \alpha_K m_{KL}^1 \right| \geq \frac{1}{16} \sum_{K \in \mathcal{D}_k(I_0)} |m_{KL}^1|, \quad \left| \sum_{L \in \mathcal{D}_k(I_0)} \alpha_L n_{KL}^1 \right| \geq \frac{1}{16} \sum_{L \in \mathcal{D}_k(I_0)} |n_{KL}^1|.
\]

It follows that

\[
\sum_{K, L \in \mathcal{D}_k(I_0)} |\lambda_{KL}| \leq 4 \sum_{K \in \mathcal{D}_k(I_0)} |m_{KL}^1|^2 \leq 4^5 \sum_{K \in \mathcal{D}_k(I_0)} \alpha_K m_{KL}^1 \cdot \sum_{L \in \mathcal{D}_k(I_0)} |n_{KL}^1| \\
\leq 4^5 \sum_{K \in \mathcal{D}_k(I_0)} \alpha_K (m_{KL}^1 + im_{KL}^2) \cdot \sum_{L \in \mathcal{D}_k(I_0)} |n_{KL}^1 + in_{KL}^2| \\
= 4^5 \sum_{K \in \mathcal{D}_k(I_0)} \alpha_K m_{KL} \cdot \sum_{L \in \mathcal{D}_k(I_0)} \alpha_L n_{KL} = 4^5 \sum_{K, L \in \mathcal{D}_k(I_0)} \alpha_K \alpha_L |\lambda_{KL}|,
\]

which is what we wanted to show. Therefore, the proof of the lemma is complete, with $c = 4^5$. \qed
6. CONCLUSION OF THE PROOF OF THEOREM 2.3

We are now ready to finish the proof of Theorem 2.3. Recall that for all slices $S_j$ of $S$ we have

$$\left| \langle S_j f, g \rangle_{L^2(W), L^2(W^{-1})} \right| \leq \sum_{L \in \mathcal{L}_j} \sum_{1 \leq i \leq d} \left| P^j_i \left( \frac{(f)_j - (f)L}{2^k} \right), P^j_i \left( \frac{(g)_j - (g)L}{2^k} \right) \right|_{\mathbb{C}^d}.$$  

Let $X := [W]_{A_j}$; fix $0 \leq j \leq k - 1$ and for all $I \in \mathcal{L}_j$ define

$$A_I := \left( \langle f \rangle_I, \|W^{1/2}f\|_{L^2}^2 \right)_I, (W)_I, (g)_I, \left( \|W^{-1/2}g\|_{L^2}^2 \right)_I, (W^{-1})_I.$$  

Notice that all these points are in Dom $\mathcal{B}_X = \mathcal{D}_X$. Lemma 5.1 says that

$$|I| \sum_{i=1}^d \sum_{P, Q \in \mathcal{D}_k} \left| \left( P^j_i \left( \frac{(f)_j - (f)L}{2^k} \right), P^j_i \left( \frac{(g)_j - (g)L}{2^k} \right) \right) \right|_{\mathbb{C}^d} \leq c \cdot d \left( |I| \mathcal{B}_X(A_L) - \sum_{I \in \mathcal{D}_k} |I| \mathcal{B}_X(A_I) \right),$$

for all $L \in \mathcal{L}_j$. We write this estimate for each $I \in \mathcal{D}_k(L)$ and then iterate the procedure $\ell$ times to obtain

$$\sum_{I \in \mathcal{L}_j} |I| \sum_{i=1}^d \sum_{P, Q \in \mathcal{D}_k(I)} \left| \left( P^j_i \left( \frac{(f)_j - (f)_L}{2^k} \right), P^j_i \left( \frac{(g)_j - (g)_I}{2^k} \right) \right) \right|_{\mathbb{C}^d} \leq c \cdot d \left( |I| \mathcal{B}_X(A_L) - \sum_{I \in \mathcal{D}_k(L)} |I| \mathcal{B}_X(A_I) \right) \leq c \cdot d |L| \mathcal{B}_X(A_L) \leq c \cdot dN(X')|L| \left( \|W^{1/2}f\|_{L^2}^2 \right)_L \left( \|W^{-1/2}g\|_{L^2}^2 \right)_L \leq c \cdot dN(X')\|f\|_{L^2(W)}\|g\|_{L^2(W^{-1})},$$

where the second inequality follows from property (ii) of the Bellman function.

Letting $\ell \to \infty$, we have

$$\sum_{I \in \mathcal{L}_j} |I| \sum_{i=1}^d \sum_{P, Q \in \mathcal{D}_k(I)} \left| \left( P^j_i \left( \frac{(f)_j - (f)_I}{2^k} \right), P^j_i \left( \frac{(g)_j - (g)_I}{2^k} \right) \right) \right|_{\mathbb{C}^d} \leq c \cdot dN(X')\|f\|_{L^2(W)}\|g\|_{L^2(W^{-1})}.$$

We now cover the real line with intervals $L \in \mathcal{L}_j$ of length $2^M$ and apply the last inequality to each $L$ to obtain that

$$\sum_{I \in \mathcal{L}_j} |I| \sum_{i=1}^d \sum_{P, Q \in \mathcal{D}_k(I)} \left| \left( P^j_i \left( \frac{(f)_j - (f)_I}{2^k} \right), P^j_i \left( \frac{(g)_j - (g)_I}{2^k} \right) \right) \right|_{\mathbb{C}^d} \leq c \cdot dN(X')\|f\|_{L^2(W)}\|g\|_{L^2(W^{-1})}.$$  

For $M \to \infty$, we get that the norm of $S_j$ is bounded by $c \cdot dN(X)$. Since $S$ was decomposed into $k$ slices, it follows that the operator norm of $S$ is bounded by $c \cdot kdN([W]_{A_j})$, and therefore the proof of Theorem 2.3 is complete.

Using the bound for matrix-weighted dyadic martingale transforms proved in [2] and the bound for matrix-weighted paraproducts in [13], page 7, together with Hytönen's representation theorem in [9], we obtain the following consequence of Theorem 2.2.
Theorem 6.1. Let \( W \) be a \( d \times d \) matrix \( A_2 \) weight on \( \mathbb{R}^p \). Let \( K \) be a standard kernel and \( T \) be a Calderón-Zygmund operator on \( \mathbb{R}^p \) associated to \( K \). Suppose that \( T \) satisfies the weak boundedness property \( \|T(\chi_Q \cdot \chi_Q)\| \leq C|Q| \) for all cubes \( Q \). Then
\[
\|T\|_{L^2(W) \to L^2(W)} \leq C \cdot 2^{3(p-1)}p([W]_{A_2})^{3/2}(2(p-1) + \log([W]_{A_2})),
\]
where \( C \) depends only on the constants in the standard estimates and the weak boundedness property, and the dimension \( d \).

Remark 6.2. Obviously, we have not used the full power of the Bellman function here - the supremum in the Bellman function is taken over all \( \tau \) with \( \|U^{1/2\tau}U^{-1/2}\| \leq 1 \), while we have only used the projections on the eigenspaces of \( U \). The setup actually allows to treat matrix-valued kernels as well, using the recent representation theorem for Calderón-Zygmund operators with operator-valued kernels in [11], which again gives a decomposition into dyadic shifts. However, in the matrix-weighted setting, one needs to adapt the decay conditions on the Calderón-Zygmund operator to the matrix weight \( W \) (see [12], page 3). This approach is the subject of the paper [21].

Remark 6.3. Following Remark 6.2, we could also have used a smaller version of the function \( N \) by choosing a smaller class of martingale transforms for our proof, namely for example
\[
N_1(X) = \sup \|T_\sigma\|_{L^2(W) \to L^2(W)},
\]
where the supremum is taken over all \( d \times d \) matrix \( A_2 \) weights \( W \) with \([W]_{A_2} \leq X \) and all sequences of \( d \times d \) matrices \( \sigma = \{\sigma_I\}_{I \in D} \) with \( \|\sigma_I\| \leq 1 \) and \( \sigma_I \) commuting with \([W]_I \) for all \( I \in D \).

One can then define the Bellman function with the projections \( P_I^I \) from Lemma 5.1 instead of the \( \tau \), running exactly the same proof. The reason we used the more general class of martingale transforms is that for both classes of \( \sigma \)'s, we have the pointwise estimate
\[
S_W(T_\sigma f)(t) \leq S_W f(t),
\]
where \( S_W \) is the matrix-weighted square function (see [2], [13]). Our expectation here was that the norm growth of the matrix-weighted square function controls the norm growth of the matrix-weighted martingale transforms in terms of \([W]_{A_2}\), and that both bounds are linear in \([W]_{A_2}\). This would, by Theorem 2.2, imply the linear bound in \([W]_{A_2}\) for general Calderón-Zygmund operators with cancellation. Indeed, the linear bound of the matrix-weighted square function has been proved after this paper was refereed [11]. The linear bound for martingale transforms remains currently open. An account on possible strategies and some of the obstacles can be found in Section 6 of [2].

7. More about Calderón-Zygmund operators with even kernel

One of the key aspects of the definition of the martingale transform operator in Section 2.2 is that the matrices \( \sigma_I \) interact well with the weight \( W \) (for the proof of our main result, we have essentially used the special case where the \( \sigma_I \)'s are diagonal in some basis).

In the scalar-valued case, the definition of the martingale transform is simpler. More precisely, for a real sequence \( \sigma = \{\sigma_I\}_{I \in D} \), \( \sigma_I = \pm 1 \), we define the martingale transform operator \( \tilde{T}_\sigma \) by
\[
\tilde{T}_\sigma f = \sum_{I \in D} \sigma_I \langle f, h_I \rangle h_I.
\]

Allowing this operator to act on vector-valued functions, we can prove a similar result to Theorem 2.2, but this time, the bound will only apply to Calderón-Zygmund operator with even kernels and sufficient smoothness of the kernel. For this, we define the function \( \tilde{N} : [1, \infty) \to [1, \infty) \) by
\[
\tilde{N}(X) = \sup \|\tilde{T}_\sigma\|_{L^2(W) \to L^2(W)},
\]
where the supremum is taken over all real sequences \( \sigma \) as above and all \( d \times d \) matrix \( A_2^2 \) weights \( W \) on \( \mathbb{R} \) with \([W]_{A_2^2} \leq X \).

Theorem 7.1. Let \( W \) be a \( d \times d \) matrix \( A_2 \) weight on \( \mathbb{R}^p \). Let \( \tilde{K} \) be an even standard kernel with smoothness \( \delta > 1/2 \) and \( T \) be a Calderón-Zygmund operator on \( \mathbb{R}^p \) associated to \( \tilde{K} \). Suppose that
T satisfies the weak boundedness property $|\langle T\chi_Q, \chi_Q \rangle| \leq C|Q|$ for all cubes $Q$, and the vanishing paraproduct conditions $T(1) = T^*(1) = 0$. Then

$$\|T\|_{L^2(W) \to L^2(W)} \leq C \cdot p d N(2^{p-1}|W|_{A_2}) \leq C_p \cdot d N(|W|_{A_2}),$$

where $C$ depends only on the constants in the standard estimates and the weak boundedness property, while $C_p$ depends on $C$ and $p$.

As before, the proof of this result follows from a corresponding inequality for self-adjoint Haar shift operators. More precisely, we will show the estimate

$$\|S^{mn}\|_{L^2(W) \to L^2(W)} \lesssim (\max\{m, n\} + 1)^{2^{\max\{m, n\}/2}} N(|W|_{A_2})$$

for all dyadic Haar shifts $S^{mn}$ of parameters $(m, n)$, which ensures the convergence of the series (since $\delta > 1/2$) in the representation theorem. This is the content of the following theorem.

**Theorem 7.2.** Let $S$ be a self-adjoint dyadic Haar shift on $\mathbb{R}$ of complexity $k \geq 1$ and $W$ be a matrix $A_2^d$ weight. Then

$$\|S\|_{L^2(W) \to L^2(W)} \leq c \cdot k 2^{k/2} N(|W|_{A_2^d}),$$

where $c$ is an absolute, positive constant.

The reduction of the proof follows almost like in Section 3, except that the orthogonal projection operators $P_i$ don’t appear. Since the dyadic Haar shift $S$ is self-adjoint, we obtain the following estimate:

$$2 \left| \langle S_j f, g \rangle_{L^2(W)} \right| \leq \left| \langle S_j f, g \rangle_{L^2(W)} \right| \leq \sum_{L \in \mathcal{L}} \sum_{P, Q \in \mathcal{D}_N(L)} \left\langle \frac{\langle f \rangle_p - \langle f \rangle_L, \frac{\langle g \rangle - \langle g \rangle_L}{2^k}}{2^k}, \frac{\langle g \rangle - \langle g \rangle_L}{2^k} \right\rangle_{C^d} + \left\langle \frac{\langle f \rangle_p - \langle f \rangle_L, \frac{\langle g \rangle - \langle g \rangle_L}{2^k}}{2^k}, \frac{\langle g \rangle - \langle g \rangle_L}{2^k} \right\rangle_{C^d} \right|.$$

With the same notations as in Section 4, the Bellman function $B_{X}$ is defined by

$$B_{X}(f, F, U, g, G, V) := |I_0|^{-1} \sup_{I \subseteq I_0} \left| \langle \langle f \rangle_{I^+} - \langle f \rangle_{I^-}, \langle g \rangle_{I^+} - \langle g \rangle_{I^-} \rangle_{C^d} \right| \cdot |I|.$$

The only differences between the properties of this function and those of the old Bellman function $\mathcal{N}(X)$ are the replacement of $\mathcal{N}(X)$ by $\tilde{N}(X)$ in (ii) and the absence of the operators $P_i$ in (iii).

The proof of Theorem 7.2 is based on the following result, which is a similar variant of Lemma 2.3.

**Lemma 7.3.** Let $X > 1$ and $B_{X}$ be a function satisfying properties (i)-(iii) from Section 6. Fix $k \geq 1$ and a dyadic interval $I_0$. For all $I \in \mathcal{D}_N(I_0)$, $0 \leq n \leq k$, let the points $A_I = (f_I, F_I, U_I, g_I, G_I, V_I) \in \mathcal{D}_X = \text{Dom} B_{X}$ be given. Assume that the points $A_I$ satisfy the dyadic martingale dynamics, i.e. $A = (A_I + A_{I^-})/2$, where $I^+$ and $I^-$ are the children of $I$. For $K, L \in \mathcal{D}_k(I_0)$, we define the coefficients $\lambda_{KL}$ by

$$\lambda_{KL} := \left\langle \frac{f_K - f_{I_0}, g_L - g_{I_0}}{2^k}, \frac{g_K - g_{I_0}}{2^k} \right\rangle_{C^d} + \left\langle \frac{f_K - f_{I_0}, g_K - g_{I_0}}{2^k}, \frac{g_L - g_{I_0}}{2^k} \right\rangle_{C^d}.$$

Then

$$\sum_{K, L \in \mathcal{D}_k(I_0)} |\lambda_{KL}| \leq c \cdot 2^{k/2} \left( B_{X'}(A_{I_0}) - 2^{-k} \sum_{I \in \mathcal{D}_k(I_0)} B_{X'}(A_I) \right),$$

where $c$ is a positive absolute constant and $X' = \frac{100}{9} X$.

The only difference between the proof of this result and that of Lemma 2.3 is the way to obtain the existence of the real sequence $\{\alpha_t\}_{t \in \mathcal{D}_k(I_0)}$ such that $|\alpha_t| \leq 1/4$ for all $t \in \mathcal{D}_k(I_0)$, $\sum_{t \in \mathcal{D}_k(I_0)} \alpha_t = 0$, and

$$\sum_{K, L \in \mathcal{D}_k(I_0)} \alpha_{KL} \lambda_{KL} \geq c \cdot 2^{-k/2} \sum_{K, L \in \mathcal{D}_k(I_0)} |\lambda_{KL}|.$$
We will use again the notation \( \Lambda := (\lambda_{KL})_{K,L \in \mathcal{D}_k(I_0)} \). Let us now define

\[
\|\Lambda\|_1 := \sup_{\alpha} \left| \sum_{K,L \in \mathcal{D}_k(I_0)} \alpha_K \alpha_L \lambda_{KL} \right|
\]

where the supremum is taken over all real sequences \( \alpha = \{\alpha_I\}_{I \in \mathcal{D}_k(I_0)} \) with \( \|\alpha\|_\infty \leq 1/4 \) and \( \sum_{I \in \mathcal{D}_k(I_0)} \alpha_I = 0 \). Since we are in a finite-dimensional space, we can find a sequence \( \alpha \) with \( |\alpha_I| \leq 1/4, \ I \in I_0 \), and \( \sum_{I \in \mathcal{D}_k(I_0)} \alpha_I = 0 \), such that

\[
\|\Lambda\|_1 \leq \sup_{\alpha} \left| \sum_{K,L \in \mathcal{D}_k(I_0)} \alpha_K \alpha_L \lambda_{KL} \right| = \|\Lambda\|_1.
\]

Using the symmetry of \( \Lambda \) and the fact that its row and column sums are all zero, it is easy to see that \( \|\Lambda\|_1 \) is equivalent to

\[
\|\Lambda\|_2 := \sup_{\|\alpha\|_\infty \leq 1} \left| \sum_{K,L \in \mathcal{D}_k(I_0)} \alpha_K \beta_L \lambda_{KL} \right|
\]

where we take the supremum over all real sequences \( \alpha = \{\alpha_I\}_{I \in \mathcal{D}_k(I_0)} \) and \( \beta = \{\beta_I\}_{I \in \mathcal{D}_k(I_0)} \). More precisely, we have \( 64\|\Lambda\|_1 \leq \|\Lambda\|_2 \leq 192\|\Lambda\|_1 \). Since we may assume that \( \Lambda \) is not the zero matrix (otherwise the lemma becomes trivially true), \( \|\Lambda\|_1 \), and hence \( \|\Lambda\|_2 \), are not 0.

We also need the notion of Schur multiplier. If \( A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R}) \), the Schur multiplier is the bounded operator \( S_A : \mathcal{M}_n(\mathbb{R}) \to \mathcal{M}_n(\mathbb{R}) \) that acts on a matrix \( M = (m_{ij}) \) by Schur multiplication: \( S_A(M) = (a_{ij}m_{ij}) \). The Schur multiplier norm is

\[
\|A\|_m := \sup_{M \in \mathcal{M}_n(\mathbb{C})} \frac{\|S_A(M)\|_{op}}{\|M\|_{op}},
\]

where \( \|M\|_{op} \) is the operator norm of the matrix \( M \) on \( \ell^2(\{1,\ldots,n\}) \). If \( A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R}) \) is of the form \( a_{ij} = s_i t_j \), then \( A \) is called a rank one Schur multiplier. It is easy to see that if \( A \) is a rank one Schur multiplier, then \( \|A\|_m \leq \|s_i\|_{\infty} \|t_j\|_{\infty} = \|A\|_1 \). A classical result due to A. Grothendieck says that the converse is essentially true (up to a constant called Grothendieck constant).

**Theorem 7.4 (5 Theorem 1.2, 24 Theorem 3.2).** The closure of the convex hull of the rank one Schur multipliers of norm one in the topology of pointwise convergence contains the ball of all Schur multipliers of norm at most \( K_G^{-1} \), where \( K_G \) is a universal constant.

If \( \alpha \) and \( \beta \) are two real sequences as above, the matrix \( \Phi = (\Phi_{KL}) = (\alpha_K \beta_L) \) is a rank one Schur multiplier of norm \( \|\Phi\|_m \) at most 1. The inequality \( \|\Lambda\|_2 \leq 192\|\Lambda\|_1 \) can thus be rewritten as

\[
\sup |\langle \Phi, \Lambda \rangle_{HS}| \leq 192\|\Lambda\|_1,
\]

where \( \langle \cdot, \cdot \rangle_{HS} \) denotes inner product on the Hilbert-Schmidt class and the supremum is taken over all rank one Schur multipliers \( \Phi \) of norm at most 1.

Using Theorem 7.4 we obtain that

\[
\sup |\langle M, \Lambda \rangle_{HS}| \leq 192K_G\|\Lambda\|_1,
\]

where the supremum is now taken over all Schur multipliers \( M \) of norm at most 1, and \( K_G \) is the (real) Grothendieck constant.

By choosing either the real or the imaginary part of the matrix \( \Lambda \) (the one with greater \( \ell^1 \)-norm), we have \( \sum_{K,L \in \mathcal{D}_k(I_0)} |\lambda_{KL}| \leq 2 \sup |\langle M, \Lambda \rangle_{HS}| \), where the supremum is taken over all matrices \( M \in \mathcal{M}_{2k}(\mathbb{R}) \) with entries \( \pm 1 \). For such a matrix \( M \) we have \( \|M\|_m \leq 2^{k/2} \), see 5, Lemma 2.5. Putting everything together, we get the inequality

\[
\sum_{K,L \in \mathcal{D}_k(I_0)} |\lambda_{KL}| \leq 384K_G2^{k/2}\|\Lambda\|_1.
\]
Using (20) and (21), it follows that there exists a real sequence \( \alpha \) with \( |\alpha| \leq 1/4 \) for all \( i \in \mathcal{D}_k(I_0) \), and \( \sum_{i \in \mathcal{D}_k(I_0)} \alpha_i = 0 \), such that

\[
\sum_{K,L \in \mathcal{D}_k(I_0)} |\lambda_{KL}| \leq 384 K G^{2k/2} \|A\|_1 = 384 K G^{2k/2} \sum_{K,L \in \mathcal{D}_k(I_0)} \alpha_K \alpha_L \lambda_{KL},
\]

which is what we wanted to show. Since the other arguments are the same as in Lemma 2.3, this completes the proof of Lemma 7.3.

The inequality in Theorem 7.2 is now obtained as in Section 6.

8. A MATRIX VERSION OF THE WEIGHTED CARLESON EMBEDDING THEOREM

In this section we will prove a version of the matrix-weighted Carleson Embedding Theorem. In the weighted setting, contrary to the unweighted case, the scalar-valued Carleson Embedding Theorem cannot be used to obtain the matrix version of the theorem. Here is the main result of this section.

Theorem 8.1 (Matrix Carleson Embedding Theorem). Let \( W \) be a \( d \times d \) matrix weight, and \{\( A_i \}_{i \in \mathcal{D}} \) be a sequence of \( d \times d \) positive definite matrices. Then for \( 0 < t \leq 1 \),

\[
t \sum_{i \in \mathcal{D}} \langle (I_d + t(W_i^{-1})\tilde{M}_i)^{-1} A_i (I_d + t\tilde{M}_i(W_i)^{-1})^{-1} (W_i^{1/2} f)_I, (W_i^{1/2} f)_I \rangle_{\mathcal{L}^d} \leq 8 \|f\|_{L^2(\mathbb{R})}^2
\]

if

\[
\frac{1}{|I|} \sum_{j \in I} \langle W_j A_j W_j \rangle_{J} \leq \langle W \rangle_I, \quad \text{for all } i \in \mathcal{D},
\]

where \( \tilde{M}_i = \frac{1}{|I|} \sum_{K \subseteq J} \langle W_k A_K \rangle_{\mathcal{L}} \) and \( I_d \) is the \( d \times d \) identity matrix.

As we have said earlier, this version is not the simple generalization of the usual weighted Carleson Embedding Theorem in [15]. This is due to the extra factor \( (I_d + t\tilde{M}_i(W_i)^{-1})^{-1} \) that appears (twice) in the left-hand side of the conclusion. However, the constants that appear in the theorem don’t depend on the dimension \( d \) or on the weight \( W \). The proof of the result also uses arguments that were previously discussed in Section 4.

Proof. Let \( t = 1 \). We first have to introduce the Bellman function associated to the problem. For \( f \in \mathbb{C}^d, F \in \mathbb{R}, W \in \mathcal{M}_d(\mathbb{C}), M \in \mathcal{M}_d(\mathbb{C}) \) satisfying

\[
(W^{-1} f, f)_{\mathcal{L}^d} \leq F \quad \text{and} \quad M \leq W,
\]

deﬁne the function \( \mathcal{B} : \mathbb{C}^d \times \mathbb{R} \times \mathcal{M}_d(\mathbb{C}) \times \mathcal{M}_d(\mathbb{C}) \) by

\[
\mathcal{B}(f, F, W, M) := 4(F - \langle (W + M)^{-1} f, f \rangle_{\mathcal{L}^d}).
\]

The Bellman function \( \mathcal{B} \) has the following properties:

(i) (Domain) The domain \( \mathcal{D} := \text{Dom} \mathcal{B} \) is given by (22).

(ii) (Range) \( 0 \leq \mathcal{B}(f, F, W, M) \leq 4F \) for all \( (f, F, W, M) \in \mathcal{D} \).

(iii) (Concavity condition) Consider all tuples \( A = (f, F, W, M), A_+ = (f_+, F_+, W_+, M_+) \) and \( A_- = (f_-, F_-, W_-, M_-) \) in \( \mathcal{D} \) such that \( f = (f_+ + f_-)/2, F = (F_+ + F_-)/2, W = (W_+ + W_-)/2, \) and \( M = m + (M_+ + M_-)/2 = m + \tilde{M} \), where \( m \) is a positive definite matrix. For all such tuples, we have the following concavity condition:

\[
\mathcal{B}(A) - \frac{\mathcal{B}(A_+) + \mathcal{B}(A_-)}{2} \geq \frac{1}{2} \langle (W + \tilde{M})^{-1} m (W + \tilde{M})^{-1} f, f \rangle_{\mathcal{L}^d}.
\]

Let us now explain these properties of the function \( \mathcal{B} \).

The inequality \( \|W^{-1/2} f\|^2_{L^2(\mathbb{R})} \leq F \) follows from the Cauchy-Schwarz Inequality. The other inequality in (6) is related to the Carleson condition.

Property (ii) follows trivially from the deﬁnition of \( \mathcal{B} \).

To prove the concavity condition, we consider three tuples \( A, A_+, A_- \in \mathcal{D} \) such that \( f = (f_+ + f_-)/2, F = (F_+ + F_-)/2, W = (W_+ + W_-)/2, \) and \( M = m + (M_+ + M_-)/2 = m + \tilde{M} \). Let
$$\hat{A} = (f, F, W, \tilde{M})$$. We prove the inequality in (iii) by splitting it into two inequalities. The first one,

$$B(\hat{A}) - \frac{1}{2}(B(A_+) + B(A_-)) \geq 0,$$

follows from the convexity of the first inequality in (ii) (like in the first part of the proof of Lemma 4.2). The second inequality,

$$B(A) - B(\hat{A}) = \frac{1}{2}((W + \tilde{M})^{-1}m(W + \tilde{M})^{-1}f, f),$$

is obtained by showing that

$$(W + \tilde{M})^{-1} - (W + \tilde{M} + m)^{-1} \geq \frac{1}{2}(W + \tilde{M})^{-1}m(W + \tilde{M})^{-1}.$$  

To see this, notice that the left-hand side of this inequality can be written as

$$(W + \tilde{M})^{-1/2} \left( I_d - \left( I_d + (W + \tilde{M})^{-1/2}m(W + \tilde{M})^{-1/2} \right)^{-1} \right) (W + \tilde{M})^{-1/2}.$$  

If $E := (W + \tilde{M})^{-1/2}m(W + \tilde{M})^{-1/2}$, we have that $0 < E \leq I_d$, since $m \leq W \leq W + \tilde{M}$. Then the inequality $I_d - (I_d + E)^{-1} \geq \frac{1}{2}E$ is equivalent to $I_d + E - I_d \geq \frac{1}{2}E(I_d + E)$, which can be rewritten as $E \geq E^2$. This last inequality is clearly true since $E \leq I_d$. It follows that

$$(W + \tilde{M})^{-1} - (W + \tilde{M} + m)^{-1} = (W + \tilde{M})^{-1/2}(I_d - (I_d + E)^{-1})(W + \tilde{M})^{-1/2} \geq \frac{1}{2}(W + \tilde{M})^{-1/2}E(W + \tilde{M})^{-1/2} = \frac{1}{2}(W + \tilde{M})^{-1}m(W + \tilde{M})^{-1},$$

which is the desired inequality.

To prove Theorem 5.1, let $W$ be a matrix weight, $f \in L^2(\mathbb{R})$ and $\{A_I\}_I$ be a sequence of $d \times d$ positive definite matrices. For any $I \in D$, let

$$f_I = \langle W^{1/2}f \rangle_I \in \mathbb{C}^d, \quad F_I = \langle |f|^2 \rangle_I \in \mathbb{R},$$

$$W_I = \langle W \rangle_I \in \mathcal{M}_d(\mathbb{C}), \quad M_I = \frac{1}{|I|} \sum_{J \subseteq I} \langle W \rangle_J A_J \langle W \rangle_J \in \mathcal{M}_d(\mathbb{C}).$$

Then

$$m_I = \frac{1}{|I|} W_I A_I W_I$$

and

$$\tilde{M}_I = \frac{1}{|I|} \sum_{J \subseteq I} \langle W \rangle_J A_J \langle W \rangle_J.$$

For the interval $I$, the concavity condition (iii) implies that

$$\frac{|I|}{2} \langle (W_I + \tilde{M}_I)^{-1}m_I(W_I + \tilde{M}_I)^{-1}f, f \rangle \leq |I|B(f_I, F_I, W_I, M_I) - |I^+|B(f_{I^+}, F_{I^+}, W_{I^+}, M_{I^+}) - |I^-|B(f_{I^-}, F_{I^-}, W_{I^-}, M_{I^-}).$$

Iterating this inequality $k$ times, we obtain

$$\frac{|I|}{2} \sum_{|J| > 2^{-k}|I|} \langle (W_J + \tilde{M}_J)^{-1}m_J(W_J + \tilde{M}_J)^{-1}f_J, f_J \rangle \leq |I|B(f_I, F_I, W_I, M_I) - \sum_{|J| = 2^{-k}|I|} |J|B(f_J, F_J, W_J, M_J) \leq |I|B(f_I, F_I, W_I, M_I) \leq 4|I|F_I.$$  

Using that

$$\langle (W_I + \tilde{M}_I)^{-1}m_I(W_I + \tilde{M}_I)^{-1}f, f \rangle = \frac{1}{|I|} \langle (I_d + \langle W \rangle_I^{-1}\tilde{M}_I)^{-1}A_I(I_d + \langle W \rangle_I^{-1}\tilde{M}_I)^{-1} \langle W^{1/2}f \rangle_I, \langle W^{1/2}f \rangle_I \rangle,$$
and letting $k \to \infty$, we get
\[
\sum_{J \subseteq I} \langle (I_d + (W)^{-1}M_J)^{-1}A_J(I_d + \tilde{M}_J(W)^{-1})^{-1}(W^{1/2}f)_J, (W^{1/2}f)_J \rangle_{C^d} \leq 8|I|\|f\|^2_{L^2(R)}
\]
which is our desired conclusion for $t = 1$.

For $0 < t < 1$, just replace $A_J$ by $tA_J$ and apply the inequality which we have just proved.

\[\square\]

**Remark 8.2.** While this paper was prepared for publication, A. Culiuc and S. Treil posted a result which appears to be the correct generalization of the scalar weighted Carleson Embedding Theorem to matrix weights in finite dimension $d$ (see [4]). In the notation of Theorem 8.1, it says that
\[
\sum_{J \subseteq I} \langle A_J(W^{1/2}f)_I, (W^{1/2}f)_I \rangle_{C^d} \leq C(d)\|f\|^2_{L^2(R)}
\]
if
\[
\frac{1}{|I|} \sum_{J \subseteq I} \langle W^{1/2}f_J, W^{1/2}f_J \rangle \leq \langle W \rangle_I
\]
for all $I \in \mathcal{D}$.

(Theorem 1.2 in [4]). An important step in their proof, the estimate (2.5) in [4], is essentially identical with our Theorem 8.1 obtained with a different proof.

**References**

[1] K. Ball, The plank problem for symmetric bodies, Invent. Math. 104 (1991), no. 3, 535-543.
[2] K. Bickel, S. Petermichl, and B. Wick, Bounds for the Hilbert transform with matrix $A_2$ weights, J. Funct. Anal. 270 (2016), no. 5, 1719-1743.
[3] R. R. Coifman and C. Fefferman, Weighted norm inequalities for maximal functions and singular integrals, Studia Math. 51 (1974), 241-250.
[4] A. Culiuc and S. Treil, The Carleson Embedding Theorem with Matrix Weights Preprint, arXiv:1508.01716 (2015).
[5] K. R. Davidson and A. P. Donsig, Norms of Schur multipliers, Illinois J. Math. 51 (2007), no. 3, 743-766.
[6] M. Goldberg, Matrix $A_p$ weights via maximal functions, Pacific J. Math. 211 (2003), no. 2, 201-220.
[7] T. S. Hämminen and T. P. Hytönen, Operator-valued dyadic shifts and the $T(1)$ theorem, Monatsh. Math. 180 (2016), no. 2, 213 – 253.
[8] R. A. Hunt, B. Muckenhoupt, and R. L. Wheeden, Weighted norm inequalities for the conjugate function and Hilbert transform, Trans. Amer. Math. Soc. 176 (1973), 227-251.
[9] T. P. Hytönen, Representation of singular integrals by dyadic operators, and the $A_2$ theorem, Lecture notes of an intensive course at Universidad de Sevilla, Summer 2011 Preprint, arXiv:1108.5110 (2011).
[10] , The sharp weighted bound for general Calderón-Zygmund operators, Ann. of Math. (2) 175 (2012), no. 3, 1473-1506.
[11] T. Hytönen, S. Petermichl, and A. Volberg, The sharp square function estimate with matrix weight Preprint, arXiv:1702.04569 (2017).
[12] J. Isralowitz, A matrix weighted $T_1$ theorem for matrix kernelled CZOs and a matrix weighted John-Nirenberg theorem Preprint, arXiv:1508.02474 (2015).
[13] J. Isralowitz, H.-K. Kwon, and S. Pott, Matrix-weighted norm inequalities for commutators and paraproducts with matrix symbols, to appear in J. London Math. Soc. Preprint, arXiv:1507.04032 (2015).
[14] F. Nazarov and S. Treil, The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis, St. Petersburg Math. J. 8 (1997), no. 5, 721–824.
[15] F. Nazarov, S. Treil, and A. Volberg, The Bellman functions and two-weight inequalities for Haar multipliers, J. Amer. Math. Soc. 12 (1999), no. 4, 909-928.
[16] F. Nazarov, S. Petermichl, S. Treil, and A. Volberg, Convex body domination and weighted estimates with matrix weights arXiv:1701.01907 (2017).
[17] S. Petermichl, The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical $A_p$ characteristic, Amer. J. Math. 129 (2007), no. 5, 1355-1375.
[18] S. Petermichl and S. Pott, An estimate for weighted Hilbert transform via square functions, Trans. Amer. Math. Soc. 354 (2002), no. 4, 1699-1703.
[19] S. Petermichl and A. Volberg, Heating of the Ahlfors-Beurling operator: weakly quasiregular maps on the plane are quasiregular, Duke Math. J. 112 (2002), no. 2, 281-305.
[20] G. Feier, Grothendieck’s theorem, past and present, Bull. Amer. Math. Soc. (N.S.) 49 (2012), no. 2, 237-323.
[21] S. Pott and A. Stoica, Sharp bounds and $T_1$ theorem for Calderón-Zygmund operators with matrix kernel on matrix weighted spaces Preprint, arXiv:1705.06105 (2017).
[22] A. W. Roberts and D. E. Varberg, *Convex functions*, Pure and Applied Mathematics, vol. 57, Academic Press, New York-London, 1973.

[23] S. Treil, *Sharp $A_2$ estimates of Haar shifts via Bellman function*, Recent trends in Analysis, Theta Ser. Adv. Math., 2013, pp. 187–208. arXiv:1105.2252.

[24] S. Treil and A. Volberg, *Wavelets and the Angle between Past and Future*, J. Funct. Anal. **143** (1997), no. 2, 269-308.

[25] A. Volberg, *Matrix $A_p$ weights via $S$-functions*, J. Amer. Math. Soc. **10** (1997), no. 2, 445-466.

[26] J. Wittwer, *A sharp estimate on the norm of the martingale transform*, Math. Res. Lett. **7** (2000), no. 1, 1-12.

Centre for Mathematical Sciences, University of Lund, P.O. Box 118, SE-221 00 Lund, Sweden

*E-mail address:* sandra@maths.lth.se

Centre for Mathematical Sciences, University of Lund, P.O. Box 118, SE-221 00 Lund, Sweden

*E-mail address:* andrei@maths.lth.se