

The matrix A_2 conjecture

Stefanie Petermichl

Abstract. This survey begins with the history of modern weighted theory and its consequences for the Hilbert transform with scalar weight. It proceeds through further development in the area that leads to the counterexample to the famous matrix A_2 conjecture.

1. The Hilbert transform

Defined as a singular integral operator on the real line, the Hilbert transform $\mathcal{H}$ is described as convolution against $1/\pi t$ and takes the form

$$\mathcal{H}f(x) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{f(t)}{x-t} dt,$$

understood as a principal value integral. It has a particularly simple form as a Fourier multiplier

$$\widehat{\mathcal{H}f}(\xi) = -i \operatorname{sign}(\xi) \widehat{f}(\xi).$$

There exists a related version on the circle $\mathbb{T}$, where the singular convolution kernel becomes

$$\frac{1}{2\pi} \cot\left(\frac{t}{2}\right)$$

and the Fourier transform is defined on integers. It follows immediately that $\mathcal{H}$ gives access to the boundary values of complex conjugate functions and that the analytic projection $\mathcal{P}_+$ can be written as $\mathcal{P}_+ = 1/2(\mathcal{I} + i\mathcal{H})$, where $\mathcal{I}$ is the identity.

The Hilbert transform itself is central to numerous branches of mathematics and certainly lies at the core of real and complex analysis as well as operator theory, but it has also long been applied as a frequency filter in AM/FM antennas. Through its multiplier form on the Fourier transform side, its formal application to $e^{imt}e^{int}$ with $|m| \gg |n|$ only sees the sign of m as $\operatorname{sign}(m+n) = \operatorname{sign}(m)$. This makes the

Hilbert transform a frequency filter, one of several crucial features when trying to filter a signal from its carrier wave. The operation of $\mathcal{H}$ is achieved by basic electrical components.

The Hilbert transform $\mathcal{H}$ is trivially bounded in L^2 by the Plancherel theorem, and using some standard methods, a weak bound at the endpoint $p = 1$, interpolation and duality, one obtains L^p boundedness for $1 < p < \infty$ (in addition to the weak-type bound in L^1 and the trivial L^2 bound). There are also many direct proofs of this fact, not using interpolation, one of which gives the sharp estimate for $\mathcal{H}$ in L^p . The proof using the weak-type bound and interpolation is the one that generalizes most easily and is usually the one found in textbooks on harmonic analysis. If a non-negative measure μ enters the picture, the boundedness in L^2_μ is no longer obvious and is false in many cases. One can show easily that it is necessary that μ be absolutely continuous with respect to Lebesgue measure and is thus described via a non-negative density w called a weight. In addition, there is a simple testing condition on the weight w that characterizes boundedness (see below).

The wish for a weighted theory of the Hilbert transform is well motivated within mathematics; it stems from certain PDE problems, stationary stochastic processes and the Riemann Hilbert problem in conjunction with Toeplitz operators, to name a few.

In order to achieve boundedness in $L^2_w(\mathbb{R})$ (in this text this means that f takes values in $\mathbb{R}$; we do not care much if our functions are defined on $\mathbb{T}$ or $\mathbb{R}$) with norm

$$\|f\|_{L^2_w(\mathbb{R})} = \left(\int_{\mathbb{R}} |f(t)|^2 w(t) dt \right)^{1/2},$$

one needs the necessary and sufficient Muckenhoupt condition [14]

$$[w]_{A_2} = \sup_I \langle w \rangle_I \langle w^{-1} \rangle_I < \infty.$$

Here, the supremum runs over all intervals I , and $\langle w \rangle_I = 1/|I| \int_I w(t) dt$ denotes the average of w over I . It is easy to check that $\langle w \rangle_I \langle w^{-1} \rangle_I \geq 1$ always holds. Examples of A_2 weights are power weights such as $w(x) = |x|^\alpha$ with $|\alpha| < 1$.

Theorem 1.1 (Hunt–Muckenaupt–Wheeden). *There exists a constant Φ that only depends on $[w]_{A_2}$ such that*

$$\|\mathcal{H}\|_{L^2_w(\mathbb{R}) \rightarrow L^2_w(\mathbb{R})} \leq \Phi([w]_{A_2}).$$

The condition $[w]_{A_2} < \infty$ is necessary for the boundedness.

When the Hilbert transform acts on vector functions $f : \mathbb{R} \rightarrow \mathbb{R}^d$ in an $L^2_W(\mathbb{R}^d)$ (functions with values in $\mathbb{R}^d$) with norms written as

$$\|f\|_{L^2_W(\mathbb{R}^d)} = \left(\int_{\mathbb{R}} \langle W(t) f(t), f(t) \rangle_{\mathbb{R}^d} dt \right)^{1/2},$$

the weight is a positive semidefinite symmetric matrix function W , and the corresponding Muckenhoupt condition [34] becomes

$$[W]_{A_2} = \sup_I \|\langle W \rangle_I^{1/2} \langle W^{-1} \rangle_I^{1/2}\|^2 < \infty,$$

where one can again show that the operator $\langle W \rangle_I^{1/2} \langle W^{-1} \rangle_I \langle W \rangle_I^{1/2}$ is expanding so that all terms in the above supremum are at least 1. The condition $[W]_{A_2} < \infty$ is again necessary and sufficient for the boundedness of $\mathcal{H} : L^2_W(\mathbb{R}^d) \rightarrow L^2_W(\mathbb{R}^d)$, where $\mathcal{H}$ acts component wise on the vector function f .

Theorem 1.2 (Treil–Volberg). *There exists a constant Φ only depending on $[W]_{A_2}$ and the dimension d such that*

$$\|\mathcal{H}\|_{L^2_W(\mathbb{R}^d) \rightarrow L^2_W(\mathbb{R}^d)} \leq \Phi([W]_{A_2}).$$

The condition $[W]_{A_2} < \infty$ is necessary for the boundedness.

In this text, we are concerned with sharp weighted estimates, which means we are interested in bounds such as

$$\|\mathcal{H}\|_{L^2_w(\mathbb{R}) \rightarrow L^2_w(\mathbb{R})} \lesssim \Phi([w]_{A_2}), \quad \|\mathcal{H}\|_{L^2_W(\mathbb{R}^d) \rightarrow L^2_W(\mathbb{R}^d)} \lesssim \Phi([W]_{A_2}),$$

where we care about the exact growth near infinity of the positive functions Φ and $\tilde{\Phi}$ that are defined on $[1, \infty)$. We say that an estimate is sharp if for any $\tilde{\Phi}$ or $\tilde{\Phi}$ that satisfy the inequalities above there holds $\Phi = O(\tilde{\Phi})$ and $\tilde{\Phi} = O(\Phi)$. This text discusses on the one hand that one can choose $\Phi(t) = t$ for scalar weights and no better than this, a well-known theorem by the author [27]. On the other hand, we highlight our recent breakthrough, Domelevo–Petermichl–Treil–Volberg [10], showing that the estimate with $\Phi(t) = t^{3/2}$, previously obtained in Nazarov–Petermichl–Treil–Volberg [23], is sharp.

Theorem 1.3 (Petermichl). *There exists $C > 0$ such that*

$$\|\mathcal{H}\|_{L^2_w(\mathbb{R}) \rightarrow L^2_w(\mathbb{R})} \leq C[w]_{A_2} \quad \forall w \in A_2.$$

There exists $c > 0$ such that for all sufficiently large Q , there exists a weight w with $[w]_{A_2} \leq Q$ and a function $0 \neq g \in L^2_w(\mathbb{R})$ such that

$$\|\mathcal{H}g\|_{L^2_w(\mathbb{R})} \geq c[w]_{A_2} \|g\|_{L^2_w(\mathbb{R})}.$$

Theorem 1.4 (Nazarov–Petermichl–Treil–Volberg). *There exists $C > 0$ only depending on d such that*

$$\|\mathcal{H}\|_{L^2_W(\mathbb{R}^d) \rightarrow L^2_W(\mathbb{R}^d)} \leq C[W]_{A_2}^{3/2} \quad \forall W \in A_2.$$

Theorem 1.5 (Domelevo–Petermichl–Treil–Volberg). *There exists $c > 0$ such that for all sufficiently large Q , there exists a 2×2 matrix weight W with $[W]_{\mathbf{A}_2} \leq Q$ and a function $0 \neq g \in L^2_W(\mathbb{R}^2)$ such that*

$$\|\mathcal{H}g\|_{L^2_W(\mathbb{R}^2)} \geq c[W]_{\mathbf{A}_2}^{3/2} \|g\|_{L^2_W(\mathbb{R}^2)}.$$

One should mention that it had been widely believed that $\Phi(t) = t^{3/2}$ even before such an estimate with $\Phi(t) = t^{3/2}$ was found by Petermichl–Pott [28] and of course before it was then again improved to the linear and sharp bound $\Phi(t) = t$ [27]. For the matrix counterpart, it had been shown by Nazarov–Petermichl–Treil–Volberg that one could choose $\Phi(t) = t^{3/2}$ [23], but it was in turn believed that the growth would again be linear. Surprisingly, in the matrix case, the growth exponent remains $3/2$, as demonstrated by Domelevo–Petermichl–Treil–Volberg [10] with an involved example. The reason for this is, of course, the non-commutativity of the matrix multiplication. In the scalar case, we have an estimate for $\langle w \rangle_I \langle w^{-1} \rangle_I$, but in the matrix case for $\langle W^{-1} \rangle_I^{1/2} \langle W \rangle_I \langle W^{-1} \rangle_I^{1/2}$ and not $\langle W \rangle_I \langle W^{-1} \rangle_I$.

We give the part of the scalar weighted history that we deem important for the estimate of the Hilbert transform as well as those results and techniques that influenced the matrix weighted case. The recent scalar weighted history is much broader, with many more interesting results. Likewise, there have been a number of advances in the matrix weighted case that we do not mention. Again, we focus on the results targeting our recent counterexample to the matrix $\mathbf{A}_2$ conjecture. To keep the flow simple and concise, we often do not state theorems in full generality or even state the results qualitatively. The reader is invited to find the precise results in the provided references.

2. First significant results

The Fourier system is the best system to describe the action of the Hilbert transform, but it is not well adapted to working with weights – this is a central difficulty of these estimates. Analysts use local systems for such tasks, and a dyadic model approach in this context was picked up by Nazarov–Treil–Volberg through the use of the Haar system in conjunction with the model operator $\mathcal{T}_\sigma$. To this end, we define the dyadic grid $\mathcal{D}$ as the following collection of dyadic intervals:

$$\mathcal{D} = \{2^k([0, 1) + n) : k, n \in \mathbb{Z}\}.$$

The Haar system is a basic wavelet system

$$\{h_I : I \in \mathcal{D}\}, h_I = \frac{1}{\sqrt{|I|}}(\chi_{I_+} - \chi_{I_-}).$$

The h_I are an orthonormal system of L^2 . It is remarkable that step functions were used – they are not smooth, and the Hilbert transform produces mild singularities and does not have good behavior on these basis functions – but they give clarity in other ways, have a direct link to powerful tools and inherit a probabilistic structure. But to achieve anything, the use of a model operator was often needed. It replaces the Hilbert transform in part of the proof or at least models how to proceed. In the early proofs, a type of transference would then give the result on $\mathcal{H}$. In modern proofs, the model is all but a guideline. One model operator is densely defined via

$$\mathcal{T}_\sigma : h_I \mapsto \sigma_I h_I,$$

where the multiplier is a sign toss $\sigma_I = \pm 1 \forall I \in \mathcal{D}$. The model operator $\mathcal{T}_\sigma$ is called the martingale transform for the obvious reasons (but for the sake of this survey, we suppress the valuable connection of these thematics to probability theory). Much like the Hilbert transform, $\mathcal{T}_\sigma$ has L^p bounds, uniform in σ , for $1 < p < \infty$ and a uniform weak estimate at the endpoint $p = 1$. The sharp uniform norm estimate for $\mathcal{T}_\sigma$ is also known but differs from the one for $\mathcal{H}$.

In the view of the author, the two papers by Nazarov–Treil–Volberg on the two-weight problem for $\mathcal{T}_\sigma$ [25] and Treil–Volberg on the boundedness of the Hilbert transform in a matrix weighted space [34] significantly helped jump-start the first sharp weighted estimates in the second part of the 1990s.

2.1. The two-weight problem for $\mathcal{T}_\sigma$

Nazarov–Treil–Volberg [25] characterize the boundedness of $\mathcal{T}_\sigma : L^2_w(\mathbb{R}) \rightarrow L^2_v(\mathbb{R})$ and develop the Bellman function technique. This means they get several necessary conditions on w, v that are also sufficient for boundedness of $\mathcal{T}_\sigma$ for all σ . To get this characterization, the authors expand test functions f and g in disbalanced Haar systems, adapted to the weights w and v . The disbalanced Haar system can easily be derived for $u = w$ or $u = v$ as $h_I^u = \alpha_I^u \chi_I + \beta_I^u h_I$ with some constants α_I^u, β_I^u only depending upon I and the weight u so as to ensure orthogonality and normalization in $L^2_u(\mathbb{R})$. The dualized form

$$|\langle \mathcal{T}_\sigma f, g \rangle_{L^2(\mathbb{R})}| \leq \sum_{I \in \mathcal{D}} |\langle f, h_I \rangle_{L^2(\mathbb{R})} \langle g, h_I \rangle_{L^2(\mathbb{R})}| \quad (2.1)$$

thus turns into an estimate of four terms when switching to the disbalanced Haar systems. The term stemming from h_I^w and h_I^v is governed by a dyadic condition similar to the A_2 condition for two different weights, while mixed terms with h_I^w and χ_I^v or h_I^v and χ_I^w require square function-type conditions with embedding theorems. The term from χ_I^w and χ_I^v is often referred to as the bilinear term; it required in this case a Sawyer-type condition on positive objects.

2.2. The angle between past and future

Treil–Volberg [34] characterize the boundedness of the Hilbert transform on vector functions in $L^2_W(\mathbb{R}^d)$ through the use of the dyadic model $\mathcal{T}_\sigma$. That is, they show that the Muckenhoupt condition for matrix weights

$$[W]_{A_2} = \sup_I \|\langle W \rangle_I^{1/2} \langle W^{-1} \rangle_I^{1/2}\|^2 < \infty$$

is necessary and sufficient for the boundedness of $\mathcal{H} : L^2_W(\mathbb{R}^d) \rightarrow L^2_W(\mathbb{R}^d)$. Part of their argument uses a beautiful matrix version of the so-called Fefferman–Kenig–Pipher A_∞ inequality [12] and contains quantitative estimates that have not even been improved today. This paper also derives many of the basic properties of matrix A_2 weights. The constants in the proof were not tracked by the authors as the paper pre-dated the spiked interest in quantitative estimates, but by shortening the proof via tricks in Petermichl–Pott [28] that gave a $3/2$ bound in the scalar case, Bickel–Petermichl–Wick [3] showed that one could get an estimate of the form $\Phi(t) = t^{3/2} \log^2(t)$ for $\mathcal{T}_\sigma$. When adding the so-called Haar shift $\mathcal{S}_{\text{cl}}$ (see later) to the toolbox, one could also obtain this estimate for $\mathcal{H}$.

3. First sharp estimates

These advances in the use of the Haar system and development of novel techniques were followed by the first sharp estimates for dyadic operators, for example, by Hukovic–Treil–Volberg for the square function [13] and by Wittwer for the martingale transform [35]. Later, the first continuous classical singular integral operator was estimated by Petermichl–Volberg with its sharp dependence, the Beurling–Ahlfors operator on the complex plane [30]. The latter estimate solved a famous regularity problem on the Beltrami equation. Several years later, the author proved the A_2 conjecture, namely the linear bound for the Hilbert transform.

3.1. The square function

The dyadic square function is

$$\mathcal{S}f(x) = \left(\sum_{I \in \mathcal{D}} \langle f, h_I \rangle_{L^2(\mathbb{R})}^2 \frac{\chi_I(x)}{|I|} \right)^{1/2}.$$

It was shown by Hukovic–Treil–Volberg [13] that

$$\|\mathcal{S}\|_{L^2_w(\mathbb{R}) \rightarrow L^2_w(\mathbb{R})} \lesssim [w]_{A_2^{\text{dy}}},$$

where the dependence is sharp and the supremum when determining the A_2 characteristic runs over dyadic intervals only (indicated by the superscript ‘dy’). The authors use the reduction to disbalanced systems. The core estimate is proved via a Carleson embedding and the Bellman proof of the embedding condition.

Theorem 3.1 (Hukovic–Treil–Volberg). *There exists $C > 0$ such that*

$$\|\mathcal{S}\|_{L^2_w(\mathbb{R}) \rightarrow L^2_w(\mathbb{R})} \leq C[w]_{A_2^{\text{dy}}},$$

and this estimate is sharp.

3.2. The linear estimate of $\mathcal{T}_\sigma$

The following celebrated theorem of Wittwer [35] proves through a mix of techniques in [25, 34] the estimate

$$\sup_\sigma \|\mathcal{T}_\sigma\|_{L^2_w(\mathbb{R}) \rightarrow L^2_w(\mathbb{R})} \lesssim [w]_{A_2^{\text{dy}}}.$$

Since square functions are expectations of martingale multipliers, this estimate is stronger than that in [13] and also more involved as it has to address the bilinear term in the four-sum proof coming from the indicators in the disbalanced systems. One example of a core estimate is the following:

$$\sum_{I \in \mathcal{D}} |\langle w, h_I \rangle_{L^2(\mathbb{R})} \langle w^{-1}, h_I \rangle_{L^2(\mathbb{R})}| \lesssim [w]_{A_2}. \quad (3.1)$$

The estimate can be accomplished by a number of Bellman functions. One example is the function

$$B_Q(x, y) = \sqrt{Q} \sqrt{xy} - \frac{1}{4}xy$$

on the domain $\Omega = \{x, y \geq 0, 1 \leq xy \leq Q\}$. There holds on Ω : $0 \leq B_Q(x, y) \leq Q$, and for its Hessian, we have

$$-d^2 B_Q = \begin{pmatrix} \sqrt{Q} \frac{\sqrt{y}}{4\sqrt{x^3}} & \frac{1}{4} - \sqrt{Q} \frac{1}{4\sqrt{xy}} \\ \frac{1}{4} - \sqrt{Q} \frac{1}{4\sqrt{xy}} & \sqrt{Q} \frac{\sqrt{x}}{4\sqrt{y^3}} \end{pmatrix} \geq \begin{pmatrix} 0 & \pm \frac{1}{4} \\ \pm \frac{1}{4} & 0 \end{pmatrix}.$$

To see the inequality for $-d^2 B_Q$, one replaces the $1/4$ by c and estimates the determinant

$$Q \frac{1}{16xy} - \left(c^2 + Q \frac{1}{16xy} - c \sqrt{Q} \frac{1}{2\sqrt{xy}} \right) \geq -c^2 + \frac{c}{2},$$

which is non-negative when $0 \leq c \leq \frac{1}{2}$. The choice $1/4 \pm 1/4$ locates c in this range. Via a standard Bellman function argument, we obtain the desired estimate. The terms $\pm 1/4$ on the $dx dy$ slots will arrange to collect terms of the form $\sim |\langle w, h_I \rangle_{L^2(\mathbb{R})} \langle w^{-1}, h_I \rangle_{L^2(\mathbb{R})}|$ when plugging averages of w and v into the Bellman function.

Theorem 3.2 (Wittwer). *There exists $C > 0$ such that for all σ ,*

$$\|\mathcal{T}_\sigma\|_{L_w^2(\mathbb{R}) \rightarrow L_w^2(\mathbb{R})} \leq C[w]_{A_2^{\text{dy}}},$$

and this estimate is sharp.

3.3. The Beurling–Ahlfors operator

Petermichl–Volberg [30] solve a famous regularity problem via a sharp weighted $L_w^2(\mathbb{C})$ estimate for the Beurling–Ahlfors operator, defined through convolution in the complex plane with the kernel $1/\pi z^2$:

$$\mathcal{B}f(z) = \frac{1}{\pi} \int_{\mathbb{C}} \frac{f(\zeta)}{(z - \zeta)^2} dA(\zeta),$$

again understood in the principal value sense. Strictly speaking, an exponent different from 2 was required, but for this type of operator, the exponent $p = 2$ is critical, and via the extrapolation theorem by Rubio de Francia, one obtains the required sharp $L_w^p(\mathbb{C})$ bound with dependence on the A_p characteristic of the weight. The operator $\mathcal{B}$ exchanges ∂ and $\bar{\partial}$ and thus plays a central role in the Beltrami equation, which is given by $\bar{\partial}f - \mu\partial f = 0$ for the Beltrami coefficient $\mu \in L^\infty$ with $\|\mu\|_{L^\infty} = k < 1$. Astala–Iwaniec–Saksman [1] proved that $\mathcal{I} - \mu\mathcal{B}$ and $\mathcal{I} - \mathcal{B}\mu$ are invertible in L^p for $p \in (1 + k, 1 + 1/k)$ using any weighted estimate for $\mathcal{B}$. The sharp estimate for $\mathcal{B}$ proved the critical exponent case, namely that these maps have dense ranges in $L^{1+1/k}$ and are injective in L^{1+k} . Injectivity at the critical exponent shows that weakly quasiregular maps are quasiregular.

The proof of the sharp weighted estimate for $\mathcal{B}$ builds on the estimate by Wittwer by a novel type of Bellman argument using the heat extension instead of dyadic extensions via dyadic averages. Indeed, Wittwer proved the dualized estimate

$$\sum_{I \in \mathcal{D}} |\langle f, h_I \rangle_{L^2(\mathbb{R})} \langle g, h_I \rangle_{L^2(\mathbb{R})}| \lesssim [w]_{A_2^{\text{dy}}} \|f\|_{L_w^2(\mathbb{R})} \|g\|_{L_{w^{-1}}^2(\mathbb{R})}. \quad (3.2)$$

Via the Fourier transform, one can see that if $\mathcal{R}_1$ is the first Riesz transform in $\mathbb{R}^2$, then for test functions $f, g \in C_0^\infty$, there holds

$$\int_{\mathbb{R}^2} \mathcal{R}_1^2 f g dx_1 dx_2 = -2 \int_0^\infty \int_{\mathbb{R}^2} \frac{\partial f}{\partial x_1} \frac{\partial g}{\partial x_1} dx_1 dx_2 dt,$$

where the letters f, g stand for the heat extensions of the functions f, g originally defined on $\mathbb{R}^2$. The absolute value of the right-hand side is estimated by a constant multiple of

$$\int_0^\infty \int_{\mathbb{R}^2} \left| \frac{\partial f}{\partial x_1} \frac{\partial g}{\partial x_1} \right| dx_1 dx_2 dt,$$

which bears resemblance to the right-hand side of inequality (2.1) or left-hand side of (3.2). In turn, we have for the Beurling–Ahlfors operator $\mathcal{B} = \mathcal{R}_1^2 - \mathcal{R}_2^2 - 2i\mathcal{R}_1\mathcal{R}_2$, where we see that the imaginary part is a rotation of the real part. Estimating $\mathcal{R}_i^2$, where without loss of generality $i = 1$, is sufficient.

If one assumes a Bellman function solving (3.2), we plug heat averages into this Bellman function to obtain an estimate linear in an A_2 heat characteristic. Thanks to the rapid decay of the kernel, this is equivalent to the classical characteristic.

Theorem 3.3 (Petermichl–Volberg). *There exists $C > 0$ such that*

$$\|\mathcal{B}\|_{L_w^2(\mathbb{C}) \rightarrow L_w^2(\mathbb{C})} \leq C[w]_{A_2},$$

and this estimate is sharp.

3.4. The sharp estimate for the Hilbert transform

Through

$$\langle \mathcal{H}f, g \rangle_{L^2(\mathbb{R})} \leq \frac{1}{2\pi} \int_{\mathbb{D}} |\nabla f| |\nabla g| \log \frac{1}{|z|} dA(z),$$

Petermichl–Wittwer [31] approached the estimate for the Hilbert transform. In this estimate, f, g on the right-hand side are the Poisson extensions of the functions f, g originally defined on $\mathbb{T}$. A similar method had previously been applied in [24] with a Bellman function that was less powerful, getting therefore a worse estimate. Plugging Poisson extensions into the Bellman function for (3.2) gives a nice estimate with respect to the Poisson A_2 characteristic. But due to the slow decay of the Poisson kernel, this only gives a quadratic bound in terms of the desired classical characteristic. Indeed, the Poisson A_2 can be much larger, and to show sharpness at the critical exponent $p = 2$ is quite involved and was accomplished by Kakaroumpas–Treil [18], a paper that had a great impact on finalizing the blowup of the norm of the Hilbert transform in the matrix weighted case.

To the situation’s rescue and in order to prove Theorem 1.3, another tool had become available: a representation formula for the Hilbert transform using the dyadic shift

$$\mathcal{S}_{\text{cl}} : h_I \mapsto h_{I_-} - h_{I_+}.$$

Petermichl [26] showed that when considering dyadic grids of the form

$$\mathcal{D}^{\alpha, r} = \{([\alpha, \alpha + r2^k) + nr2^k) : k, n \in \mathbb{Z}\}$$

for $\alpha \in \mathbb{R}$ and $r > 0$, one can consider the shift operators densely defined on the Haar functions built on intervals in $\mathcal{D}^{\alpha,r}$ as

$$\mathcal{S}_{\text{cl}}^{\alpha,r} : h_I^{\alpha,r} \mapsto h_{I-}^{\alpha,r} - h_{I+}^{\alpha,r}.$$

It can be shown that an average in translations α followed by a logarithmic average in r produces a non-zero multiple of the Hilbert transform. This reduces weighted bounds for $\mathcal{H}$ to estimates for $\mathcal{S}_{\text{cl}}$, a much simpler task if one uses the Haar base. Originally used only for upper estimates, we will see that this operator also plays a role in the blowup, or the lower estimate, in the matrix weighted case.

Staying with dyadic methods, the main tool at that time was Bellman functions. The core estimate of this work is motivated by the action of the dyadic shift:

$$\sum_{I \in \mathcal{D}} |\langle w, h_I \rangle_{L^2(\mathbb{R})} \langle w^{-1}, h_{I-} \rangle_{L^2(\mathbb{R})}| \lesssim [w]_{A_2}.$$

There is a subtle but substantial difference to inequality (3.1): the index ‘ $-$ ’ on the second Haar function makes the term hard to estimate with Bellman functions as it requires a pairing of differences at a different time of the dyadic martingale. The solution consisted of the observation that the Bellman function used for this problem has a Hessian comparable to a diagonal matrix. This allows us to collect $(dx)^2 + (dy)^2$ instead of the seemingly smaller $|dx dy|$. With the increments separated, one could manage to insert an increment from the ‘wrong’ time. The sharp weighted estimate could thus be completed for the Haar shift.

Theorem 3.4 (Petermichl). *There exists $C > 0$ such that*

$$\|\mathcal{S}_{\text{cl}}\|_{L_w^2(\mathbb{R}) \rightarrow L_w^2(\mathbb{R})} \leq C [w]_{A_2}^{\text{dy}},$$

and this estimate is sharp and independent of the choice of the dyadic grid.

Theorem 1.3 followed as a corollary. This idea of ‘separating increments’ was developed later by Dragičević–Treil–Volberg [11] in the form of the so-called ellipse lemma. Finally, it was shown by Treil [32] using an abstract argument that one can always do such operations, even if one does not know the Bellman function explicitly. Treil’s estimate also has advantages when the dyadic shift operators have higher complexity (for example, $h_I \mapsto h_{I-}$ has complexity 3). The original argument of Petermichl [27] can be extended to shift operators of higher complexity, but the estimate one obtains is exponential in complexity, while that of Treil is linear. The author thinks the method of Treil has important implications in the area of convexity, but its features also became relevant, for example, for the representation formula of Hytönen [16] for all Calderón–Zygmund operators via generalized (high-complexity) dyadic shifts and general A_2 conjecture, the extension of the linear estimates to all Calderón–Zygmund operators.

4. Sparse revolution

In parallel to the Bellman method, more classical approaches via a form of sparse operator existed, and they became very popular due to developments around 2015. A sparse collection $\mathcal{R}$ of dyadic intervals has the following property. Defining

$$E_I = I \setminus \bigcup_{J \in \mathcal{R}, J \subsetneq I} J, \quad (4.1)$$

then $\mathcal{R}$ is sparse if, say, $|E_I| \geq 1/2|I|$. A sparse operator is an operator of the form

$$\mathcal{L}f(x) = \sum_{I \in \mathcal{R}} \langle |f| \rangle_I \chi_I(x),$$

where the $\mathcal{L}$ reminds us of Lerner. It had been known that sparse operators allow a simple weighted estimate through the use of maximal functions.

In 2015, there were several groups in parallel achieving a pointwise domination formula in terms of sparse operators for the Hilbert transform and other Calderón–Zygmund operators and in particular also for dyadic operators. We highlight the beautiful top-down proof by Lacey [20] that was extremely simple for the martingale multiplier or the dyadic shift but extended to Calderón–Zygmund operators. We state it for $\mathcal{H}$.

Theorem 4.1 (Lacey). *For any compactly supported function $f \in L^1(\mathbb{R})$, we have*

$$|\mathcal{H}f| \lesssim \mathcal{L}(f),$$

where $\mathcal{L}$ depends upon f and is a sum of at most three operators, each of which is sparse relative to a choice of dyadic grid on $\mathbb{R}$.

Building on an unweighted weak-type bound, a stopping time argument gives a pointwise domination of the dyadic version below, where the reader is again invited to pay close attention to the quantifiers. For all $f \in L^1$, there exists a sparse collection $\mathcal{R}$ so that almost surely

$$|\mathcal{T}f(x)| \lesssim \sum_{I \in \mathcal{R}} \langle |f| \rangle_I \chi_I(x)$$

for dyadic operators $\mathcal{T}$ such as $\mathcal{S}_{\text{cl}}$ or $\mathcal{T}_\sigma$. The sparse intervals are the stopping intervals of hitting the double threshold of $|f|$ or $|\mathcal{T}f|$. A weak-type bound is required to observe that one stops rarely, thus getting property (4.1). There also holds such an estimate for the Hilbert transform (as stated above) and other Calderón–Zygmund operators. In the case of the Hilbert transform, only three different dyadic grids are needed; in general, the number of dyadic grids required depends upon the dimension.

Singular integral operators such as $\mathcal{H}$ rely heavily on cancellation of their kernel, so a domination by a positive object is strange – it is achieved through a choice of a collection of sparse intervals that in turn depends upon f .

With the availability of these ideas, many interesting sharp estimates followed that were out of reach with the methods that had been used prior. The first application was the extension beyond Calderón–Zygmund theory by Bernicot–Frey–Petermichl [2] that also set the notion of weak sparse forms. Shortly after, a paper on the bilinear Hilbert transform by Culiuc–Di Plinio–Ou [6] followed, along with many more. The interest in sharp matrix weighted bounds had resurfaced around 2013. But this sudden speed motivated us to revisit the matrix weighted questions, hoping that with all these new techniques, the problem would be ready to be solved.

5. Matrix weights: Positive results

We give a few recent notable positive results. Some of them are the equivalent in the scalar weighted theory, while some are not. These results with an identical formulation as in the scalar case fortified the belief that the matrix $\mathbf{A}_2$ conjecture was true. Results with worse estimates were believed to be not sharp.

5.1. Matrix weighted Carleson lemma

Culiuc–Treil [7] were able to show the weighted Carleson lemma without the use of the $\mathbf{A}_2$ condition, so precisely the same as in the scalar case. While in the scalar case, the weighted and unweighted versions are indistinguishable, there was a notable effort to get the matrix weighted version. Here is the statement of the matrix version, as proved in [7].

Theorem 5.1 (Culiuc–Treil). *Let $I_0 = [0, 1)$. Let (A_I) be a sequence of positive semidefinite matrices. There exists a constant C only depending on d such that there holds*

$$\begin{aligned} \frac{1}{|K|} \sum_{J \in \mathcal{D}(K)} \langle W \rangle_J A_J \langle W \rangle_J &\leq \langle W \rangle_K \quad \forall K \in \mathcal{D} \\ \Rightarrow \sum_{J \in \mathcal{D}(I_0)} \|A_I^{1/2} \langle Wf \rangle_I\|^2 &\leq C \|f\|_{L_W^2(\mathbb{R}^d)}^2 \quad \forall f \in L_W^2(\mathbb{R}^d). \end{aligned}$$

The condition on the left-hand side is trivially necessary for the embedding as it represents a special case. The proof is a ‘Bellman function with a parameter’ argument. In the scalar case, the function

$$B(F, x, w, M) = F - \frac{x^2}{w + M}$$

represents the martingale variables $F = \langle f^2 w \rangle_J$, $x = \langle wf \rangle_J$, $w = \langle w \rangle_J$ and the super martingale $M = 1/|J| \sum_{I \in \mathcal{D}(J)} \alpha_I \langle w \rangle_I^2$. It is concave in $\{F \geq f^2, 0 \leq M \leq w\}$, and on this set, there holds $0 \leq B(F, x, M) \leq F$. Another crucial estimate is

$$\frac{\partial B}{\partial M} = \frac{f^2}{(w + M)^2} \geq \frac{f^2}{4w^2}. \quad (5.1)$$

This function gave the (sharp) estimate for the scalar (weighted) Carleson embedding. For its matrix analog $\mathbf{B}(\mathbf{F}, \mathbf{f}, \mathbf{W}, \mathbf{M}) = \mathbf{F} - \langle (\mathbf{W} + \mathbf{M})^{-1} \mathbf{f}, \mathbf{f} \rangle$, similar estimates hold, except that the simple derivative estimate (5.1) fails. Indeed,

$$\frac{\partial \mathbf{B}}{\partial \mathbf{M}} \Delta \mathbf{M} = \langle (\mathbf{W} + \mathbf{M})^{-1} \Delta \mathbf{M} (\mathbf{W} + \mathbf{M})^{-1} \mathbf{f}, \mathbf{f} \rangle,$$

and there is no way to estimate $(\mathbf{W} + \mathbf{M})^{-1} \Delta \mathbf{M} (\mathbf{W} + \mathbf{M})^{-1}$ from below. The increment $\Delta \mathbf{M}$ corresponds to some $\langle W \rangle_J A_J \langle W \rangle_J$, and in the factors $(\mathbf{W} + \mathbf{M})^{-1}$, the $\mathbf{M}$ has to disappear to have the resulting $\langle W \rangle_J^{-1} \langle W \rangle_J A_J \langle W \rangle_J \langle W \rangle_J^{-1} = A_J$. The idea is thus to multiply $\mathbf{M}$ by a scalar t and let $t \rightarrow 0$, certainly making the $\mathbf{M}$ disappear. Though this argument produces an inner derivative and a factor t . Simply letting $t \rightarrow 0$ does not produce the estimate; however, a bit more work in complex analysis allows us to do this with a penalty of an expected dimensional constant, here d^2 , where d is the side length of the matrices. The scalar case has a numeric estimate 4, and it is known that this is optimal. It is interesting that Kraus–Moucha–Roth [19] found a complex analysis argument for the part as $t \rightarrow 0$ that gave the estimate $4d^2$. It remains an open question if the dimensional growth is optimal.

The result of Culiuc–Treil [7] readily implied with standard methods an estimate of the ‘matrix-weighted unweighted maximal function’, that is, Doob’s inequality with matrix measure $W(x)dx$. Namely, $M_D^W : L_W^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R})$,

$$M_D^W f(x) = \sup_{I \in \mathcal{D}, x \in I} \|W^{1/2}(x) \langle W \rangle_I^{-1} \langle Wf \rangle_I\|_{\mathbb{R}^d},$$

is bounded. This estimate was obtained through linearization by Petermichl–Pott–Reguera [29] as a consequence of [7].

Theorem 5.2 (Petermichl–Pott–Reguera). *Let W be any matrix weight. Then there exists C only depending on d such that*

$$\|M_D^W\|_{L_W^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R})} \leq C.$$

Compare with the scalar case, where the supremum is usually taken before the multiplication with $w(x)$ and $M_w : L_w^2(\mathbb{R}) \rightarrow L_w^2(\mathbb{R})$,

$$M_w f(x) = \sup_{I: x \in I \in \mathcal{D}} |\langle f \rangle_{I, w}| = \sup_{I: x \in I \in \mathcal{D}} w(I)^{-1} \left| \int_I f w \right| = \sup_{I: x \in I \in \mathcal{D}} \langle w \rangle_I^{-1} |\langle f w \rangle_I|.$$

But $\|M^w f\|_{L^2(\mathbb{R})} = \|M_w f\|_{L_w^2(\mathbb{R})}$ since

$$w^{1/2}(x)M_w f(x) = \sup_{I:x \in I \in \mathcal{D}} |w^{1/2}(x)\langle f \rangle_{I,w}| = \sup_{I:x \in I \in \mathcal{D}} |w^{1/2}(x)\langle w \rangle^{-1}\langle fw \rangle_I|.$$

Here, it does not matter if we take the supremum before or after we multiply pointwise by $w^{1/2}(x)$, but the matrix case leaves no choice.

5.2. Linear bound for the square function

Hytönen–Petermichl–Volberg [15] proved the linear bound for the matrix weighted square function, later extended by Treil [33] to the nonhomogeneous case. The square function in this setting again has the weight W incorporated in a natural way. It is

$$\mathcal{S}_W : L_W^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}), \quad \mathcal{S}_W f(x) = (\mathbb{E} \|W(x)^{1/2} \mathcal{T}_\sigma f(x)\|_{\mathbb{R}^d}^2)^{1/2}.$$

Theorem 5.3 (Hytönen–Petermichl–Volberg). *There exists C only depending on d such that*

$$\|\mathcal{S}_W\|_{L_W^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R})} \leq C[W]_{A_2}.$$

This is the same sharp bound as in the scalar case as $\mathcal{S} : L_w^2(\mathbb{R}) \rightarrow L_w^2(\mathbb{R})$ is equivalent to $\mathcal{S}_w : L_w^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$. From far, it looked as if this was the natural progression – in the scalar case, the square function was also quite a bit easier than, for example, the estimate for $\mathcal{H}$, which eventually followed. However, the authors noticed that a typical core difficulty does not arise if one has the integrated form of the square function at hand:

$$\|\mathcal{S}_W f\|_{L^2(\mathbb{R})}^2 = \sum_{I \in \mathcal{D}} \langle \langle W \rangle_I \langle f, h_I \rangle_{L^2(\mathbb{R}^d)}, \langle f, h_I \rangle_{L^2(\mathbb{R}^d)} \rangle_{\mathbb{R}^d},$$

where there is a slight abuse of notation as h_I are real valued. When trying to estimate $\|W(x)^{1/2} \mathcal{T}_\sigma f(x)\|_{\mathbb{R}^2}$ (without the expectation) in $L^2(\mathbb{R})$, there are various options on how to dominate by a version of a sparse operator that in turn will be estimated by some appropriate maximal functions. There are a number of ways to do this. A common theme is the pointwise multiplier $W(x)^{1/2}$ that persists throughout. The suprema of the maximal operators typically act only after this pointwise multiplication. This makes these maximal operators difficult to stop, and every attempt cashes in an extra half power of $[W]_{A_2}$, either by splitting off terms or employing awkward definitions of maximal operators. So there were good reasons why the square function could be estimated with the linear bound, while other operators could not.

5.3. The 3/2 bound

The authors Nazarov–Petermichl–Treil–Volberg [23] chose one of their numerous proofs for the publication of the estimate

$$\|\mathcal{H}\|_{L^2_W(\mathbb{R}^d) \rightarrow L^2_W(\mathbb{R}^d)} \lesssim [W]_{\mathbf{A}_2}^{3/2},$$

using the so-called convex body domination followed by an estimate involving certain sparse square functions. These sparse square function estimates were interesting in their own right and helped in turn to estimate the matrix weighted square function $\mathcal{S}_W$ linearly. The proof in [23] is by sparse domination and as such also works for all Calderón–Zygmund operators.

The convex body average of a vector function f over the interval I is defined as

$$\langle\langle f \rangle\rangle_I = \{\langle\varphi_I f\rangle_I : \varphi_I : I \rightarrow [-1, 1]\}.$$

When $f \in L^1_{\chi_I}(\mathbb{R}^d)$ (L^1 on I with values in $\mathbb{R}^d$) is non-trivial, then there exists a unique subspace $E \in \mathbb{R}^d$ such that $\langle\langle f \rangle\rangle_I \subset E$ such that $\langle\langle f \rangle\rangle_I$ has non-empty interior in E , hence the name convex body. The convex body sparse operators for a sparse collection $\mathcal{R}$ are then

$$\mathcal{L}f(x) = \sum_{I \in \mathcal{R}} \langle\langle f \rangle\rangle_I \chi_I(x).$$

It is thus shown in [23] that there holds the domination

$$\mathcal{H}f(x) \in C \sum_i \sum_{I \in \mathcal{R}_i} \langle\langle f \rangle\rangle_I \chi_I(x).$$

Using simple functions, it is easy to see that if $g(x) \in \langle\langle f \rangle\rangle_I$ almost surely on I , then there exists a measurable function $K : I \times I \rightarrow \mathbb{R}$ with $\|K\|_{L^\infty} \leq |I|^{-1}$ such that $g(x) = \int_I K(x, y)f(y)dy$. This approach was one of many – all giving $\Phi(t) = t^{3/2}$. The theorem was stated in the introduction, Theorem 1.4. The choice to write down the estimate in terms of square functions turned out natural in retrospect, as it generalized beautifully to the case beyond the Calderón–Zygmund theory in [17] by Kakaroumpas–Nguyen–Vardakis.

6. First negative results in matrix weights

It had been observed very early that matrix weights lacked certain basic features we were used to from the scalar setting, for example, the openness condition of the weight classes [5]. While we had further observed that many techniques failed completely or at least presented with difficulty that ruined constants, it remained a very difficult task to show that there is a definite failure of certain quantitative results in the matrix case. In this section, we present some of these negative results.

6.1. The convex body maximal operator is not bounded

The use of convex bodies helped to mimic what would be absolute values inside the average, in the scalar case, expressions like $\langle |f| \rangle_I$, which arise in the definition of the sparse operators. While the maximal operators below (D for Doob and LP for Littlewood–Paley),

$$M_D^w f(x) = \sup_{I: x \in I \in \mathcal{D}} |w(x)^{1/2} \langle w \rangle_I^{-1} \langle f w \rangle_I|,$$

$$M_{LP}^w f(x) = \sup_{I: x \in I \in \mathcal{D}} w(x)^{1/2} \langle w \rangle_I^{-1} \langle |f| w \rangle_I,$$

are essentially the same, both bounded by the constant 2, their matrix analogs have very different behavior. For instance,

$$M_D^W f(x) = \sup_{I \in \mathcal{D}, x \in I} \|W^{1/2}(x) \langle W \rangle_I^{-1} \langle W f \rangle_I\|_{\mathbb{R}^d}$$

is bounded from $L_W^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R})$. To mimic the absolute value inside the average, one uses convex bodies:

$$M_{LP}^W f(x) = M^W f(x) = \sup_{I \in \mathcal{D}, x \in I} \sup_{\varphi_I: I \rightarrow [-1, 1]} \|W^{1/2}(x) \langle W \rangle_I^{-1} \langle \varphi_I W f \rangle_I\|_{\mathbb{R}^d}.$$

As was shown by Nazarov–Petermichl–Škreb–Treil [22], this convex body maximal function is in fact unbounded. The unboundedness is derived from the failure of the convex body Carleson lemma with matrix weight, which is also part of the work in [22]. If (A_I) is again a sequence of positive definite 2×2 matrices and if there holds the Carleson condition, then it does not follow that we may insert any φ_I into the conclusion of the Culiuc–Treil matrix weighted Carleson lemma.

Theorem 6.1 (Nazarov–Petermichl–Škreb–Treil). *There exist A_I, W_n, f and admissible φ_I so that $\|f\|_{L_{W_n}^2(\mathbb{R}^2)}$ is uniformly bounded but*

$$\begin{aligned} \frac{1}{|K|} \sum_{J \in \mathcal{D}(K)} \langle W_n \rangle_J A_J \langle W_n \rangle_J &\leq \langle W_n \rangle_K \quad \forall K \in \mathcal{D}, \\ \sum_{J \in \mathcal{D}(I_0)} \|A_I^{1/2} \langle \varphi_I W_n f \rangle_I\|_{\mathbb{R}^2}^2 &\rightarrow \infty \quad \text{as } n \rightarrow \infty. \end{aligned}$$

The counterexample uses a weight that only varies via a rotation in its eigenbasis and increasing eccentricity by a consecutive and patient procedure that rotates in every step the matrix weight while maintaining the martingale property. A constant $f \equiv e$ (almost any vector) will work to show the blowup. Access to the function φ_I allowed us to choose $\varphi_I = \chi_{I_+}$ inducing a shift in time on the term

$\langle \varphi_I W_n \rangle_I \sim \langle W_n \rangle_{I_+}$ as compared to just $\langle W_n \rangle_I$. While W_n and (A_I) were chosen to be in harmony so that operators A_I never see the large eigenvalues of W_n , this was brought out of balance by the time shift induced by φ_I , collecting a constant term on each one of the weight's moves. The index n describes how many moves were done. Letting $n \rightarrow \infty$ gives the blowup.

We were able to pass the example to the convex body maximal function.

Theorem 6.2 (Nazarov–Petermichl–Škreb–Treil). *There exists a matrix weight W such that M_{LP}^W does not map $L_W^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R})$.*

6.2. The bilinear Carleson lemma fails

A much simpler example gives the failure of a crucial ingredient in the original proof of the linear bound for $\mathcal{H}$, a bilinear weighted Carleson lemma. In the scalar case, it is the following.

Theorem 6.3 (Petermichl–Wittwer). *If (α_I) is a non-negative sequence of scalars, then*

$$\begin{aligned} & \frac{1}{K} \sum_{I \in \mathcal{D}(K)} \alpha_I \leq 1 \quad \forall K \in \mathcal{D} \\ \Rightarrow & \sum_{I \in \mathcal{D}} \frac{\alpha_I}{\langle w \rangle_I \langle w^{-1} \rangle_I} \langle w | f | \rangle_I \langle w^{-1} | g | \rangle_I \lesssim \|f\|_{L_w^2(\mathbb{R})} \|g\|_{L_{w^{-1}}^2(\mathbb{R})} \end{aligned}$$

for all $f \in L_w^2(\mathbb{R})$ and $g \in L_{w^{-1}}^2(\mathbb{R})$.

The dyadic version is found in [27], but the original version is by Petermichl–Wittwer [31], written with Poisson averages instead of dyadic averages, and it was stated with two more conditions that were later observed to be implied by the one condition above. These were

$$\begin{aligned} & \frac{1}{|K|} \sum_{I \in \mathcal{D}(K)} \frac{\alpha_I}{\langle w \rangle_I} \leq \langle w^{-1} \rangle_K \quad \forall K \in \mathcal{D}, \\ & \frac{1}{|K|} \sum_{I \in \mathcal{D}(K)} \frac{\alpha_I}{\langle w^{-1} \rangle_I} \leq \langle w \rangle_K \quad \forall K \in \mathcal{D}. \end{aligned} \tag{6.1}$$

The statement of Theorem 6.3 was new, but the proof was, up to a small additional observation, implicit in [25]. Notice that other than in the linear case, the Carleson condition is not necessary for the conclusion.

A matrix version by Petermichl–Pott–Reguera [29] seemed to be very weak, and it did not help at all with the $\mathbf{A}_2$ conjecture.

Theorem 6.4 (Petermichl–Pott–Reguera). *If (α_I) is a non-negative scalar sequence, then*

$$\frac{1}{K} \sum_{I \in \mathcal{D}(K)} \alpha_I \leq 1 \quad \forall K \in \mathcal{D}$$

$$\Rightarrow \sum_{I \in \mathcal{D}} \alpha_I |\langle \langle W \rangle_I^{-1} \langle Wf \rangle_I, \langle W^{-1} \rangle_I^{-1} \langle W^{-1}g \rangle_I \rangle_{\mathbb{R}^d}| \lesssim \|f\|_{L^2_W(\mathbb{R}^d)} \|g\|_{L^2_{W^{-1}}(\mathbb{R}^d)}$$

for all $f \in L^2_W(\mathbb{R}^d)$, $g \in L^2_{W^{-1}}(\mathbb{R}^d)$.

The proof was no longer by Bellman function but a simpler argument using level sets.

It is however interesting that the redundancy of the extra conditions (6.1) works in the matrix case. Petermichl–Pott–Reguera [29] showed that for non-negative (α_I) , there holds

$$\frac{1}{K} \sum_{I \in \mathcal{D}(K)} \alpha_I \leq 1 \quad \forall K \in \mathcal{D} \Rightarrow \frac{1}{K} \sum_{I \in \mathcal{D}(K)} \alpha_I \langle W^{-1} \rangle^{-1} \leq 4 \langle W \rangle_K \quad \forall K \in \mathcal{D}$$

and the symmetric statement with W and W^{-1} exchanged.

Domelevo–Petermichl–Škreb [9] showed that one cannot in any way improve Theorem 6.4. Indeed, replacing (α_I) by a matrix sequence (A_I) fails, and replacing the inner product by a norm fails even in the presence of the scalar sequence. Not even assuming $W \in \mathbf{A}_2$ is helpful as one needs instead an actual barrier on the conditioning numbers of W . To use this theorem in a four-sum proof to estimate, say, $\mathcal{T}_\sigma$ in the matrix case, the inner product would have to be exchanged by a product of norms and the sequence has to be matrix valued.

7. The failure of the matrix $\mathbf{A}_2$ conjecture

It was eventually shown by Domelevo–Petermichl–Treil–Volberg [10] that in the matrix weighted case, the $3/2$ exponent obtained in [23] cannot be improved. In the scalar case, one can choose a simple weight with one singularity to see the optimality of the linear estimate. A power weight $w_\alpha(x) = |x|^\alpha$ belongs to A_2 if and only if $-1 < \alpha < 1$ with $[w_\alpha]_{A_2} = 1/(1 - \alpha^2)$. Let us choose α close to 1. Through the choice of an appropriate related test function $f_\alpha = x^{-\alpha} \chi_{(0,1)}$, one can compute directly that the weighted norm of $\mathcal{H}$ must grow at least linearly as seen by the following two

computations:

$$\begin{aligned}\|f_\alpha\|_{L^2_{w_\alpha}(\mathbb{R})}^2 &= \int_0^1 x^{-2\alpha} x^\alpha dx = \frac{1}{1-\alpha}, \\ \|\mathcal{H} f_\alpha\|_{L^2_{w_\alpha}}^2 &\geq \int_2^\infty \left(\int_0^1 \frac{1}{x-t} f_\alpha(t) dt \right)^2 w_\alpha(x) dx \\ &\gtrsim \int_2^\infty \frac{1}{x^2} \frac{1}{(1-\alpha)^2} w_\alpha(x) dx \gtrsim \frac{1}{(1-\alpha)^3}.\end{aligned}$$

It is curious that the proof of the matrix bound in [23] returns a linear bound for specific situations, such as when one tries to do anything similar to the scalar example, that is, to choose a weight with one singularity. To be more clear, simple sparse operators, where the collection $\mathcal{R}$ has the property that each interval in $\mathcal{R}$ has at most one direct descendant, enjoy linear matrix weighted bounds. Leaning on this obstacle, Domelevo–Petermichl–Treil–Volberg use a sparse collection that just fails to be simple and construct on top of it a dyadic sparse paraproduct that has the feared $3/2$ exponent in its estimate. Such a weight has arbitrarily many singularities.

We first show the desired lower bound for some dyadic operators in terms of the dyadic $\mathbf{A}_2^{\text{dy}}$ condition. The weight we construct is necessarily quite complicated, constructed by a top-down approach: we apply subtle rotations of the eigenvectors of the averages in one step (which is only possible in the vector case), and in the next step, we disbalance the weight, which is in essence the action also giving the sharp blowup in the scalar case if one looks at the power weights. The rotation, which is only possible in the vector case, adds some subtle extra blowup: for the constructed weight, we then prove the estimate from below for several dyadic operators relevant to our task. We start with a special sparse paraproduct, which will be the main building block for all the dyadic estimates. We deduce that the classical Haar shift $\mathcal{S}_{\text{cl}}$ has the desired blowup as well. We then prove the blowup for a special dyadic operator $\mathcal{H}^{\text{dy}}$ which in our situation serves as a dyadic model for the Hilbert transform. While doing so, a number of quite recent estimates come into play – to allow us to not worry about all the terms, we utilize many times estimates that we know are linear – among them some sparse square functions from [23] or the square function estimate itself [15].

As explained earlier, the average of the Haar shifts $\mathcal{S}_{\text{cl}}^{\alpha,r}$ gives the Hilbert transform – as such, getting upper estimates for $\mathcal{H}$ from the estimates of the Haar shift is more immediate, intuitive and obvious; this idea has been used for years. The case of a blowup is more involved: in our passage from the dyadic operators $\mathcal{H}^{\text{dy}}$ to $\mathcal{H}$, we take inspiration from Bourgain [4] in the use of separating frequencies (the periodizations). A modification of his construction by considering second-order martingale differences allowed us to obtain the blowup for $\mathcal{H}$ (with a modified weight) from the blowup for $\mathcal{H}^{\text{dy}}$. The choice of the dyadic operator $\mathcal{H}^{\text{dy}}$ is forced upon us by the construction; we do not have any freedom here.

The construction does not stop here. The weight we constructed for the dyadic operator and its periodization are in the dyadic class $\mathbf{A}_2^{\text{dy}}$; however, they both have large jumps and do not satisfy the classical matrix $\mathbf{A}_2$ condition.

To overcome this difficulty, we borrow from the technique of quasi-periodization, developed by Nazarov [21] in his counterexample to the Sarason conjecture. This technique allows us to get the classical $\mathbf{A}_2$ condition, but only in a two-weight situation. However, in the present paper, we are in the one-weight situation and thus have even less freedom. So, in the final step, we use a modification of the so-called iterated remodeling technique, developed by Kakaroumpas–Treil [18] for the scalar case.

While the passage from the dyadic blow-up estimate to $\mathcal{H}$ is complicated and quite rigid in that the Hilbert transform seems to make specific demands as to which dyadic operator it will accept, the construction of the matrix weight and the blowup of the sparse paraproduct are key. We try to explain qualitatively the construction of the weight and follow up on why the $3/2$ comes into play.

For the construction of W , we start on I_0 and set the initial values W_0, V_0 of our martingales W, V . By the end of the construction, at time ∞ , they will be our weight W and its inverse W^{-1} ,

$$W_0 = \alpha_0 a_0 a_0^* + \beta_0 b_0 b_0^*, \quad V_0 = \beta_0^\# a_0 a_0^* + \alpha_0^\# b_0 b_0^*,$$

so that they are extremal with respect to the $\mathbf{A}_2^{\text{dy}}$ characteristic, say $\mathbf{Q}$, so that $\alpha_0 \beta_0^\# = \alpha_0^\# \beta_0 = \mathbf{Q}$. We have the initial eccentricity $E_0 = \alpha_0 / \beta_0 = \alpha_0^\# / \beta_0^\#$ and set our starting angle for the first rotation $\delta_0 = 0.1 E_0^{-1/2}$. Set $\mathcal{R}_0 = \{I_0\}$. Assume $\mathcal{R}_n = \{I \in \mathcal{R}(I_0) : |I| = 2^{-2n}\}$ is constructed with

$$W_n = \alpha_n a_n a_n^* + \beta_n b_n b_n^*, \quad V_n = \beta_n^\# a_n a_n^* + \alpha_n^\# b_n b_n^*,$$

$\alpha_n \beta_n^\# = \alpha_n^\# \beta_n = \mathbf{Q}$, $E_n = \alpha_n / \beta_n = \alpha_n^\# / \beta_n^\#$, and we set $\delta_n = 0.1 E_n^{-1/2}$. We then rotate the eigenvectors by $\pm \delta_n$ while maintaining the martingale property, receiving two new sets of main axes $(a_n^\pm, b_n^\pm)$ with modified coefficients. In the next step, we stretch the large eigenvalue by the right amount $r = 2 - 1/\mathbf{Q}$ and put the result on the interval to the right with its coefficients α_{n+1} and β_{n+1} . We put on the left side what we have to in order to preserve the martingale property. These right intervals are then the set $\mathcal{R}_{n+1}$, and the construction is such that again $\alpha_{n+1} \beta_{n+1}^\# = \alpha_{n+1}^\# \beta_{n+1} = \mathbf{Q}$.

We will eventually terminate this construction with one more step, after enough steps have been made so that enough damage has been done to the estimate. This last step is simple but essential so as to end up with an actual matrix weight function and its inverse.

In the following, the function $f : I_0 \rightarrow \mathbb{R}^2$, and we denote by $h_I^\infty = \sqrt{|I|} h_I$ the L^∞ normalized Haar functions. From above, $\mathcal{R}$ is the sparse collection and

$\mathcal{R}_<(I_0) = \mathcal{R} \setminus \{I_0\}$. The sparse paraproduct is

$$\Pi f(x) = \sum_{I \in \mathcal{R}_<(I_0)} \langle f \rangle_I h_I^\infty(x).$$

Then f will take the form $\chi_{I_0} W^{-1}e$ for some appropriate e , and we will compute $\|\Pi f\|_{L_W^2(\mathbb{R}^2)}$ for a specially constructed $W : I_0 \rightarrow \mathbb{R}^{2 \times 2}$. Denote $\mathcal{R}_<(J) = \{I \subsetneq J : I \in \mathcal{R}\}$. Then

$$\begin{aligned} & \|\Pi(W^{-1}\chi_{I_0}e)\|_{L_W^2(\mathbb{R}^2)}^2 \\ &= \sum_{I \in \mathcal{R}_<(I_0)} \sum_{J \in \mathcal{R}_<(I_0)} \langle \langle W^{-1} \rangle_I h_I^\infty e, \langle W^{-1} \rangle_J h_J^\infty e \rangle_{L_W^2(\mathbb{R}^2)} \\ &= \sum_{I \in \mathcal{R}_<(I_0)} \|\langle W \rangle_I^{1/2} \langle W^{-1} \rangle_I e\|_{\mathbb{R}^2}^2 |I| \\ &\quad + 2 \sum_{I \in \mathcal{R}_<(I_0)} \sum_{J \in \mathcal{R}_<(I)} \int \langle \langle W^{-1} \rangle_I e, W(s) h_I^\infty(s) h_J^\infty(s) \langle W^{-1} \rangle_J e \rangle_{\mathbb{R}^2} ds. \end{aligned}$$

The blowup we need will be in the second sum 2Σ , and this term will be positive. Denote $\mathcal{R}_\pm(I) = \{J \in \mathcal{R}_<(I) : J \subset I_\pm\}$. When $J \in \mathcal{R}_\pm(I)$, then $\langle W h_I^\infty h_J^\infty \rangle_J = \pm \langle W h_J^\infty \rangle_J$, and therefore,

$$\begin{aligned} \Sigma &= \sum_{I \in \mathcal{R}_<(I_0)} \sum_{J \in \mathcal{R}_+(I)} \langle \langle W^{-1} \rangle_I e, \langle W h_J^\infty \rangle_J \langle W^{-1} \rangle_J e \rangle_{\mathbb{R}^2} |J| \\ &\quad - \sum_{I \in \mathcal{R}_<(I_0)} \sum_{J \in \mathcal{R}_-(I)} \langle \langle W^{-1} \rangle_I e, \langle W h_J^\infty \rangle_J \langle W^{-1} \rangle_J e \rangle_{\mathbb{R}^2} |J|. \end{aligned}$$

In the scalar case, the summands would be estimated by $\langle w^{-1} \rangle_I [w]_{A_2} |J|$. If $I \in \mathcal{R}_<(I_0)$ with $|I| = 2^{-2n}$ and we sum over $J \in \mathcal{R}_<(I)$ with $|J| = 2^{-2k}$ for $k > n$, then

$$\begin{aligned} \sum_{J \in \mathcal{R}_{k <}(I)} \langle w^{-1} \rangle_I \langle w \rangle_J \langle w^{-1} \rangle_J |J| &\leq [w]_{A_2} \langle w^{-1} \rangle_I \sum_{J \in \mathcal{R}_{k <}(I)} |J| \\ &= [w]_{A_2} \langle w^{-1} \rangle_I 2^{n-k} |I|. \end{aligned}$$

We would then sum over $k > n$:

$$\sum_{J \in \mathcal{R}_{k <}(I)} \langle w^{-1} \rangle_I \langle w \rangle_J \langle w^{-1} \rangle_J |J| \leq \sum_{k > n} [w]_{A_2} \langle w^{-1} \rangle_I 2^{n-k} |I| \leq [w]_{A_2} \langle w^{-1} \rangle_I |I|.$$

For scalar weights, $\sum_{I \in \mathcal{R}(I_0)} \langle w^{-1} \rangle_I |I| \lesssim [w]_{A_2} \langle w^{-1} \rangle_{I_0}$, giving the quadratic bound for Σ and thus linear bound for Π .

In the matrix case, there is no estimate for $\langle W^{-1} \rangle_J \langle W \rangle_J$. Our example is constructed so that most terms have an estimate from below of the form

$$\mathbf{Q} r^{k-n} \langle \langle W^{-1} \rangle e, e \rangle_{\mathbb{R}^2} |J|,$$

where $r = 2 - 1/\mathbf{Q}$. Thus, instead of summing the geometric series $\sum_{l=1}^{\infty} 2^{-l} = 1$, we sum $\sum_{l=1}^{\infty} (r/2)^k = 2\mathbf{Q} - 1$ which gives us an extra factor $\mathbf{Q}$ in the sum.

Thanks to the stretching in every second step, the weight is very disbalanced so that the paraproduct Π is comparable to a slice of the classical Haar shift. By a ‘slice’ we mean the dyadic sum of the Haar shift restricted to every second generation of dyadic intervals. As it turns out, this is not enough to get a lower estimate for $\mathcal{H}$. The dyadic operator $\mathcal{H}^{\text{dy}}$ involves the (sliced) odd operator $\mathcal{S}_{\text{cl}} - \mathcal{S}_{\text{cl}}^* + c\mathcal{S}_0$, where $\mathcal{S}_0 : h_{I_{\pm}} \mapsto \pm h_{I_{\mp}}$ is a new type of odd dyadic shift, recently used by Domelevo–Petermichl [8] in a question on Banach space geometry. For our situation, the $\mathcal{S}_0$ did not stress us as the slice we used has small norm, but the demand on $\mathcal{S}_{\text{cl}} - \mathcal{S}_{\text{cl}}^*$ was a cliffhanger: both these operators $\mathcal{S}_{\text{cl}}$ and $\mathcal{S}_{\text{cl}}^*$ have the same contributions, but luckily with opposing signs.

Funding. This work was partially supported by the Alexander von Humboldt Foundation.

References

- [1] K. Astala, T. Iwaniec, and E. Saksman, [Beltrami operators in the plane](#). *Duke Math. J.* **107** (2001), no. 1, 27–56 Zbl [1009.30015](#) MR [1815249](#)
- [2] F. Bernicot, D. Frey, and S. Petermichl, [Sharp weighted norm estimates beyond Calderón–Zygmund theory](#). *Anal. PDE* **9** (2016), no. 5, 1079–1113 Zbl [1344.42009](#) MR [3531367](#)
- [3] K. Bickel, S. Petermichl, and B. D. Wick, [Bounds for the Hilbert transform with matrix \$A_2\$ weights](#). *J. Funct. Anal.* **270** (2016), no. 5, 1719–1743 Zbl [1335.44001](#) MR [3452715](#)
- [4] J. Bourgain, [Some remarks on Banach spaces in which martingale difference sequences are unconditional](#). *Ark. Mat.* **21** (1983), no. 1–2, 163–168 Zbl [0533.46008](#) MR [0727340](#)
- [5] M. Bownik, [Inverse volume inequalities for matrix weights](#). *Indiana Univ. Math. J.* **50** (2001), no. 1, 383–410 Zbl [0992.42006](#) MR [1857041](#)
- [6] A. Culiuc, F. Di Plinio, and Y. Ou, [Domination of multilinear singular integrals by positive sparse forms](#). *J. Lond. Math. Soc. (2)* **98** (2018), no. 2, 369–392 Zbl [1402.42013](#) MR [3873113](#)

- [7] A. Culiuc and S. Treil, [The Carleson embedding theorem with matrix weights](#). *Int. Math. Res. Not. IMRN* **2019** (2019), no. 11, 3301–3312 Zbl [1428.42018](#) MR [3961702](#)
- [8] K. Domelevo and S. Petermichl, The dyadic and the continuous Hilbert transforms with values in Banach spaces. [v1] 2022, [v2] 2023, arXiv:[2212.00090v2](#)
- [9] K. Domelevo, S. Petermichl, and K. A. Škreb, [Failure of the matrix weighted bilinear Carleson embedding theorem](#). *Linear Algebra Appl.* **582** (2019), 452–466 Zbl [1426.42013](#) MR [3995793](#)
- [10] K. Domelevo, S. Petermichl, S. Treil, and A. Volberg, The matrix A_2 conjecture fails, i.e. $3/2 > 1$. 2024, arXiv:[2402.06961v1](#)
- [11] O. Dragičević, S. Treil, and A. Volberg, [A theorem about three quadratic forms](#). *Int. Math. Res. Not. IMRN* **2008** (2008), article no. rnn072, 9 pp. Zbl [1245.42012](#) MR [2439569](#)
- [12] R. A. Fefferman, C. E. Kenig, and J. Pipher, [The theory of weights and the Dirichlet problem for elliptic equations](#). *Ann. of Math. (2)* **134** (1991), no. 1, 65–124 Zbl [0770.35014](#) MR [1114608](#)
- [13] S. Hukovic, S. Treil, and A. Volberg, [The Bellman functions and sharp weighted inequalities for square functions](#). In V. P. Havin and N. K. Nikolski (Eds.), *Complex analysis, operators, and related topics*, pp. 97–113, Oper. Theory Adv. Appl. 113, Birkhäuser, Basel, 2000 Zbl [0972.42011](#) MR [1771755](#)
- [14] R. Hunt, B. Muckenhoupt, and R. Wheeden, [Weighted norm inequalities for the conjugate function and Hilbert transform](#). *Trans. Amer. Math. Soc.* **176** (1973), 227–251 Zbl [0262.44004](#) MR [0312139](#)
- [15] T. Hytönen, S. Petermichl, and A. Volberg, [The sharp square function estimate with matrix weight](#). *Discrete Anal.* (2019), article no. 2, 8 pp. Zbl [1431.42019](#) MR [3936542](#)
- [16] T. P. Hytönen, [The sharp weighted bound for general Calderón–Zygmund operators](#). *Ann. of Math. (2)* **175** (2012), no. 3, 1473–1506 Zbl [1250.42036](#) MR [2912709](#)
- [17] S. Kakaroumpas, T. H. Nguyen, and D. Vardakis, [Matrix-weighted estimates beyond Calderón–Zygmund theory](#). *Math. Ann.* **394** (2026), no. 3, article no. 57, 56 pp. Zbl [08165566](#) MR [5033744](#)
- [18] S. Kakaroumpas and S. Treil, [“Small step” remodeling and counterexamples for weighted estimates with arbitrarily “smooth” weights](#). *Adv. Math.* **376** (2021), article no. 107450, 52 pp. Zbl [1453.42011](#) MR [4178922](#)
- [19] D. Kraus, A. Moucha, and O. Roth, [A sharp Bernstein-type inequality and application to the Carleson embedding theorem with matrix weights](#). *Anal. Math. Phys.* **12** (2022), no. 1, article no. 40, 6 pp. Zbl [1482.42063](#) MR [4376241](#)
- [20] M. T. Lacey, [An elementary proof of the \$A_2\$ bound](#). *Israel J. Math.* **217** (2017), no. 1, 181–195 Zbl [1368.42016](#) MR [3625108](#)
- [21] F. Nazarov, A counterexample to Sarason’s conjecture. Unpublished manuscript, <https://users.math.msu.edu/users/fedja/prepr.html> visited on 20 March 2026
- [22] F. Nazarov, S. Petermichl, K. A. Škreb, and S. Treil, [The matrix-weighted dyadic convex body maximal operator is not bounded](#). *Adv. Math.* **410** (2022), no. part A, article no. 108711, 20 pp. Zbl [1504.42065](#) MR [4491249](#)

- [23] F. Nazarov, S. Petermichl, S. Treil, and A. Volberg, [Convex body domination and weighted estimates with matrix weights](#). *Adv. Math.* **318** (2017), 279–306 Zbl 1382.42010 MR 3689742
- [24] F. Nazarov and S. Treil, The weighted norm inequalities for Hilbert transform are now trivial. *C. R. Acad. Sci. Paris Sér. I Math.* **323** (1996), no. 7, 717–722 Zbl 0858.44004 MR 1416164
- [25] 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 Zbl 0951.42007 MR 1685781
- [26] S. Petermichl, [Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol](#). *C. R. Acad. Sci. Paris Sér. I Math.* **330** (2000), no. 6, 455–460 Zbl 0991.42003 MR 1756958
- [27] 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 Zbl 1139.44002 MR 2354322
- [28] S. Petermichl and S. Pott, [An estimate for weighted Hilbert transform via square functions](#). *Trans. Amer. Math. Soc.* **354** (2002), no. 4, 1699–1703 Zbl 1005.42002 MR 1873024
- [29] S. Petermichl, S. Pott, and M. C. Reguera, [A matrix weighted bilinear Carleson lemma and maximal function](#). *Anal. Math. Phys.* **9** (2019), no. 3, 1163–1180 Zbl 1428.42021 MR 4014862
- [30] 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 Zbl 1025.30018 MR 1894362
- [31] S. Petermichl and J. Wittwer, [A sharp estimate for the weighted Hilbert transform via Bellman functions](#). *Michigan Math. J.* **50** (2002), no. 1, 71–88 Zbl 1040.42008 MR 1897034
- [32] S. Treil, Sharp A_2 estimates of Haar shifts via Bellman function. In *Recent trends in analysis*, pp. 187–208, Theta Ser. Adv. Math. 16, Theta, Bucharest, 2013 Zbl 1299.42046 MR 3411052
- [33] S. Treil, [Mixed \$A_2\$ - \$A_\infty\$ estimates of the non-homogeneous vector square function with matrix weights](#). *Proc. Amer. Math. Soc.* **151** (2023), no. 8, 3381–3389 Zbl 1514.42019 MR 4591773
- [34] S. Treil and A. Volberg, [Wavelets and the angle between past and future](#). *J. Funct. Anal.* **143** (1997), no. 2, 269–308 Zbl 0876.42027 MR 1428818
- [35] J. Wittwer, [A sharp estimate on the norm of the martingale transform](#). *Math. Res. Lett.* **7** (2000), no. 1, 1–12 Zbl 0951.42008 MR 1748283

Stefanie Petermichl

Institut für Mathematik, Universität Würzburg, Emil-Fischer-Straße 41, 97074 Würzburg, Germany; stefanie.petermichl@uni-wuerzburg.de