Variational Carleson operators in UMD spaces

Alex Amenta\textsuperscript{1} | Gennady Uraltsev\textsuperscript{2}

\textsuperscript{1} Mathematisches Institut, Universität Bonn, Bonn, Germany
\textsuperscript{2} Department of Mathematics, University of Virginia, Charlottesville, Virginia, USA

Correspondence
Gennady Uraltsev, Department of Mathematics, University of Virginia, Charlottesville, VA, USA.
Email: gennady.uraltsev@gmail.com

Funding information
Alexander von Humboldt Foundation

Abstract
We prove $L^p$-boundedness of variational Carleson operators for functions valued in intermediate UMD spaces. This provides quantitative information on the rate of convergence of partial Fourier integrals of vector-valued functions. Our proof relies on bounds on wave packet embeddings into outer Lebesgue spaces on the time-frequency-scale space $\mathbb{R}^3_+$, which are the focus of this paper.

MSC 2020
42B20 (primary), 42B25, 46E40 (secondary)

Contents
1. INTRODUCTION .................................... 1364
2. PRELIMINARIES .................................... 1368
   2.1. Notions and lemmas in Banach-valued analysis .............. 1368
   2.2. Outer Lebesgue spaces .................................. 1377
   2.3. The time-frequency-scale space .......................... 1382
   2.4. (Truncated) wave packets and $\mathcal{V}_{\gamma'}$ .............. 1387
3. REDUCTION OF THE MAIN THEOREM ...................... 1391
   3.1. Reduction to wave packet embedding bounds ............... 1391
   3.2. Reduction to domination by a scalar embedding .......... 1394
4. SCALAR DOMINATION OF THE TRUNCATED WAVE PACKET EMBEDDING ... 1394
   4.1. The Lebesgue term .................................... 1395
   4.2. Splitting the randomised term ........................... 1397
   4.3. Controlling $H_\gamma$ .................................. 1400
   4.4. Bounding the variation term ........................... 1402
   4.5. Bounding the first error term $\text{Err}^{(1)}$ ................ 1403
   4.6. Bounding the second error term $\text{Err}^{(2)}_{t(t)}$ .......... 1405
   4.7. Bounding the third error term $\text{Err}^{(3)}$ ............... 1407

© 2022 The Authors. The publishing rights in this article are licensed to the London Mathematical Society under an exclusive licence.
1 | INTRODUCTION

Consider a Schwartz function $f \in \mathcal{S}(\mathbb{R}; X)$ taking values in a Banach space $X$. For all frequencies $\xi \in \mathbb{R}$, the partial Fourier integral of $f$ up to $\xi$ is

$$\mathcal{C}_\xi f(x) := \int_{-\infty}^{\xi} f(\eta)e^{2\pi i \eta x} \, d\eta.$$ 

By Fourier inversion, $f$ can be recovered as the pointwise limit of partial Fourier integrals

$$\lim_{\xi \to \infty} \mathcal{C}_\xi f(x) = f(x) \quad \forall x \in \mathbb{R}. \quad (1.1)$$

This limit holds for all Schwartz functions, but when $f$ is merely assumed to be in the Bochner space $L^p(\mathbb{R}; X)$ for some $p \in (1, \infty)$, the pointwise convergence of partial Fourier integrals is a difficult matter. When the functions under consideration are scalar-valued, that is, when $X = \mathbb{C}$, the Carleson–Hunt theorem says that the limit $(1.1)$ exists for almost all $x \in \mathbb{R}$. Various proofs of this theorem are known, all using some form of time-frequency analysis; see [10, 12, 18, 23, 28, 35]. More generally, when $X$ is a Banach space with the intermediate UMD property (defined in Section 2.1.1), the convergence $(1.1)$ still holds for almost all $x \in \mathbb{R}$, as shown by Hytönen and Lacey [24]. The UMD property is equivalent to the $L^p$-boundedness of the Hilbert transform on $X$-valued functions, so it is a necessary condition for the existence of partial Fourier integrals, which must be ensured before convergence can even be discussed. However, it is not known whether the formally stronger ‘intermediate’ UMD property is necessary for the $X$-valued Carleson–Hunt theorem.$^\dagger$

The almost-everywhere validity of the limit $(1.1)$ for functions in $L^p(\mathbb{R}; X)$ follows from the $L^p$-boundedness of the Carleson maximal operator $\mathcal{C}_\ast$, defined on $f \in \mathcal{S}(\mathbb{R}; X)$ by

$$\mathcal{C}_\ast f(x) := \sup_{\xi \in \mathbb{R}} \| \mathcal{C}_\xi f(x) \|_X = \| \xi \mapsto \mathcal{C}_\xi f(x) \|_{L^\infty(\mathbb{R}; X)} \quad \forall x \in \mathbb{R}.$$ 

However, this argument does not provide any information on the rate of convergence, and it relies on having already established the convergence for Schwartz functions. To get quantitative information without need for an $a$ priori result on a dense subclass, one can consider variational operators. For $r \in [1, \infty)$, the $r$-variational Carleson operator is defined by

$$\mathcal{V}_r^\ast f(x) := \| \xi \mapsto \mathcal{C}_\xi f(x) \|_{V^r(\mathbb{R}; X)}, \quad (1.2)$$

$^\dagger$ It is not necessary for the Banach-valued ‘little Carleson theorem’ proven in [32], which controls the growth rate of partial Fourier series rather than convergence. Here, only the UMD property is needed.
where the $r$-variation of a path $u : \mathbb{R} \to X$ in a Banach space $X$ is

$$\|u\|_{V^r(\mathbb{R};X)} := \sup_{\epsilon \in \Delta} \left( \sum_{j=0}^{\infty} \|u(\epsilon_{j+1}) - u(\epsilon_j)\|_X \right)^{1/r}$$

with supremum taken over all increasing sequences $(\epsilon_j)_{j \in \mathbb{N}}$ in $\mathbb{R}$; here

$$\Delta := \{ \epsilon \in \mathbb{R}^\mathbb{N} : \epsilon_j \leq \epsilon_{j+1}, j \in \mathbb{N} \}$$

is the infinite simplex.

In the scalar-valued setting, Oberlin, Seeger, Tao, Thiele, and Wright showed that the $r$-variational Carleson operator $V^r_*$ is $L^p$-bounded for all $r' < p < \infty$, provided that $r > 2$ [31]. In this article, we extend their result to intermediate UMD spaces (defined in Section 2.1.1).

**Theorem 1.1.** Let $X$ be an $r_0$-intermediate UMD Banach space for some $r_0 \in [2, \infty)$. Then

$$\|V^r_* f\|_{L^p(\mathbb{R})} \lesssim_{p,r,X} \|f\|_{L^p(\mathbb{R};X)}$$

for all $r_0 < r < \infty$ and $(r/(r_0 - 1))' < p < \infty$.

When $r_0 = 2$, that is, when $X$ is isomorphic to a Hilbert space, the condition on $p$ becomes $r' < p < \infty$, as in the scalar-valued setting. As $r \to \infty$, the condition on $p$ tends to $1 < p < \infty$, and since $\mathcal{C}_c f(x) \leq V^r_* f(x)$ for all $f \in \mathcal{S}(\mathbb{R};X)$ and $x \in \mathbb{R}$, Theorem 1.1 implies $L^p$-boundedness of the Carleson maximal operator $\mathcal{C}_c$ for all $p \in (1, \infty)$ when $X$ is intermediate UMD, which is the main result of Hytönen and Lacey [24].

The bounds in Theorem 1.1 are a consequence of uniform bounds on linearisations of $\mathcal{V}^r_*$, in which the pointwise $r$-variation is replaced by the pointwise $\ell^r$-norm with respect to a selection function $\epsilon : \mathbb{R} \to \Delta$ prescribing a sequence of frequencies for each $x \in \mathbb{R}$. More precisely, given a measurable function $\epsilon : \mathbb{R} \to \Delta$ (meaning that each coordinate function $\epsilon_j : \mathbb{R} \to \mathbb{R}$ is measurable), we define a sequence-valued function $\mathcal{V}^r_{\epsilon} f = (\mathcal{V}^r_{\epsilon,j} f)_{j \in \mathbb{N}}$ on $\mathbb{R}$ by

$$\mathcal{V}^r_{\epsilon,j} f(x) := \int_{\epsilon_j(x)}^{\epsilon_{j+1}(x)} \hat{f}(\xi) e^{2\pi i \xi x} d\xi = \mathcal{C}_{\epsilon_{j+1}(x)}(x) - \mathcal{C}_{\epsilon_j(x)}(x),$$

where we abuse notation and write $\epsilon_j(x) := \epsilon(x)_j$. Then we have the pointwise control

$$\mathcal{V}^r_* f(x) \leq \sup_{\epsilon : \mathbb{R} \to \Delta} \|\mathcal{V}^r_{\epsilon} f(x)\|_{\ell^r(\mathbb{N};X)},$$

and Theorem 1.1 readily follows from the following result, which we prove.

**Theorem 1.2.** Let $X$ be an $r_0$-intermediate UMD Banach space for some $r_0 \in [2, \infty)$. Then for all measurable functions $\epsilon : \mathbb{R} \to \Delta$, we have

$$\|\mathcal{V}^r_{\epsilon} f\|_{L^p(\mathbb{R};\ell^r(\mathbb{N};X))} \lesssim_{p,r,X} \|f\|_{L^p(\mathbb{R};X)} \quad \forall f \in \mathcal{S}(\mathbb{R};X)$$

for all $r_0 < r < \infty$ and all $(r/(r_0 - 1))' < p < \infty$, with implicit constant independent of $\epsilon$. 
Our proof of Theorem 1.2 uses the framework of embeddings into outer Lebesgue spaces on the time-frequency-scale space $\mathbb{R}^3_+$, as in the scalar-valued setting by the second author in [36]. First, for all $f \in \mathcal{S}(\mathbb{R}; X)$ and all finitely-supported sequences $g \in c_00(\mathbb{N}; \mathcal{S}(\mathbb{R}; X^*))$ of $X^*$-valued Schwartz functions, the dual form to $\mathcal{V}$ is represented by

$$
\int_{\mathbb{R}} \langle \mathcal{V} f(x); g(x) \rangle \, dx = \sum_{\square \in \{+,-\}} \int_{\mathbb{R}^3_+} \langle \text{Emb}[f](\eta, y, t); A^{\square}_{\mathcal{V}}[g](\eta, y, t) \rangle \, d\eta \, dy \, dt,
$$
as the sum of two terms, each of which is the integral pairing of a wave packet embedding $\text{Emb}[f]$ of the function $f$ with a truncated wave packet embedding $A^{\pm}_{\mathcal{V}}[g]$ of the sequence $g$ (see Definition 3.1). These integral pairings satisfy the Hölder-type bounds

$$
\int_{\mathbb{R}^3_+} \left| \langle \text{Emb}[f](\eta, y, t); A^{\pm}_{\mathcal{V}}[g](\eta, y, t) \rangle \right| \, d\eta \, dy \, dt \lesssim \| \text{Emb}[f] \|_{L^p_{\mathcal{L}_{\mathcal{F}}} L^q_{\mathcal{L}_{\mathcal{G}}}} \| A^{\pm}_{\mathcal{V}}[g] \|_{L^p'_{\mathcal{L}_{\mathcal{F}}' L^q'_{\mathcal{L}_{\mathcal{G}}'}}};
$$

the quantities on the right-hand side are outer Lebesgue quasi-norms of functions on $\mathbb{R}^3_+$ (defined in Sections 2.2 and 2.3). Under the given assumptions on $X$, the wave packet embedding satisfies the bounds

$$
\| \text{Emb}[f] \|_{L^p_{\mathcal{L}_{\mathcal{F}}} L^q_{\mathcal{L}_{\mathcal{G}}}} \lesssim \| f \|_{L^p(\mathbb{R}; X)} \quad \forall f \in \mathcal{S}(\mathbb{R}; X)
$$

for all $p \in (1, \infty]$ and all $q > \min(p, r_0)'(r_0 - 1)$, as proven by the authors in [3]. Here we show the truncated wave packet embedding bounds

$$
\| A^{\pm}_{\mathcal{V}}[g] \|_{L^p'_{\mathcal{L}_{\mathcal{F}}' L^q'_{\mathcal{L}_{\mathcal{G}}'}}} \lesssim \| g \|_{L^p'(\mathbb{R}; \ell^{r_0'}(\mathbb{N}; X^*))} \quad \forall g \in c_00(\mathbb{N}; \mathcal{S}(\mathbb{R}; X^*))
$$

for suitable exponents $p$, $q$, and $r$ (see Theorem 3.3). We prove these by reduction to an auxiliary scalar-valued embedding map, whose boundedness was already established by the second author in [36]. This reduction is our Theorem 3.6, which is the main technical result of the paper. Its proof relies on variational estimates for convolution operators, along with various other technical estimates, using the UMD and cotype assumptions on $X$ (but interestingly, not the intermediacy assumption: this is because the necessary arguments are essentially of Calderón–Zygmund type, rather than time-frequency).

We comment briefly on previous research on variational Carleson operators. In the scalar-valued setting, besides the initial results of Oberlin et al. [31], there are the weighted estimates of Do and Lacey [15, 16], the outer Lebesgue approach of the second author [36], and the sparse domination result of Di Plinio, Do, and the second author [13] (which can also be proven by the helicoidal method of Benea and Muscalu [4, §7.2]). For vector-valued functions, bounds have been obtained by Benea and Muscalu [4] (for iterated Lebesgue spaces) and by the first author, Lorist, and Veraar [1] (for Banach function spaces with UMD concavifications).

In the general setting of UMD Banach spaces, the only result concerning variational Carleson operators available until now was the Walsh model considered by Hytönen, Lacey, and Parissis [25]. In this model, the real line is replaced by the Walsh group, resulting in an idealised dyadic analysis which retains core features of the problem while eliminating pesky tail estimates. Their theorem has essentially the same hypotheses as Theorem 1.2, the only difference being that ‘$r_0$-intermediate UMD’ is replaced with ‘tile-type $r$ for all $r > r_0$’, which follows from our assumption.
In fact, our results follow under the assumption that $X$ is UMD and satisfies the $L^p$-$L^2$ tree orthogonality estimates of [3, Theorem 6.5] for all $p > r_0$. The paper [3] itself deals with the boundedness of a related operator, namely the Bilinear Hilbert Transform, that is also subject to time-frequency-scale techniques. An analysis of this operator parallel to that of [3] has been independently carried out by [14]. We believe that the approach therein can also be used to recover the results of the current paper.

It is possible that the tree orthogonality estimates are equivalent to tile-type; for now we leave this problem open and remark that there is no known way of proving either condition without assuming $r_0$-intermediate UMD.

Remark 1.3. Our results can be reformulated for functions on the one-dimensional torus, and deduced from the corresponding results on $\mathbb{R}$ by standard transference methods. In the scalar-valued case, this is done in [31, Appendix A]; extending this to the vector-valued case does not pose additional difficulties (for an argument in the case of Banach function spaces, see [1]). Thus our results imply variational estimates for Fourier series as well as Fourier integrals.

Outline of the paper. Section 2 contains preliminary discussions and results on technical tools needed through the paper: UMD and intermediate UMD spaces, type and cotype, $\gamma$-radonifying operators, $R$-bounds, outer Lebesgue spaces, and truncated wave packets. This section includes important technical results, in particular, Lemma 2.18 (an estimate for sequences of convolution-type operators between mixed $L^2$-$\gamma$ and $L^2$-$\ell^r$ spaces for UMD spaces with cotype $r$) and Proposition 2.36 (truncated wave packet representation of the Fourier projection onto an interval). Section 3 contains the reduction of Theorem 1.2 to the embedding domination result of Theorem 3.6, which mostly consists of applications of the preliminary results. The heart of the paper is Section 4, in which Theorem 3.6 is actually proven. In Section 5, we discuss $r$-variational Carleson operators for functions valued in Banach function spaces. Finally, in Section 6, we give a detailed sketch of the proof of the wave packet embedding bounds from [3].

Notation. For Banach spaces $X$ and $Y$, we let $\mathcal{L}(X, Y)$ denote the Banach space of bounded linear operators from $X$ to $Y$, and we let $\mathcal{L}(X) := \mathcal{L}(X, X)$. For $p \in [1, \infty]$, we let $L^p(\mathbb{R}; X)$ denote the Bochner space of strongly measurable functions $\mathbb{R} \to X$ such that the function $x \mapsto \|f(x)\|_X$ is in $L^p(\mathbb{R})$; for technical details, see [26, Chapter 1]. We use the notation $\langle \cdot ; \cdot \rangle$ to denote the duality pairing between a Banach space $X$ and its dual $X^*$. For $p \in [1, \infty]$, we let $p'$ denote the conjugate exponent $p' := p/(p-1)$. We use the Japanese bracket notation

$$\langle x \rangle := (1 + |x|^2)^{1/2} \quad \forall x \in \mathbb{R}.$$ 

For each point $(y, \eta, t) \in \mathbb{R}_+^3 = \mathbb{R}^2 \times \mathbb{R}_+$, we define the translation, modulation, and ($L^1$-normalised) dilation operators on measurable functions $f : \mathbb{R} \to X$ by

$$\text{Tr}_y f(z) := f(z - y), \quad \text{Mod}_\eta f(z) := e^{2\pi i \eta z} f(z), \quad \text{Dil}_t f(z) := t^{-1} f\left(\frac{z}{t}\right),$$ 

and we define the composition

$$\Lambda(y, \eta, t) := \text{Tr}_y \text{Mod}_\eta \text{Dil}_t.$$
For $x \in \mathbb{R}$ and $r > 0$, we let $B_r(x)$ denote the ball centred at $x$ with radius $r$, and we write $B_r := B_r(0)$.

## 2 | PRELIMINARIES

### 2.1 | Notions and lemmas in Banach-valued analysis

#### 2.1.1 | UMD and intermediate UMD Banach spaces

**Definition 2.1.** A Banach space $X$ has the **UMD property** if the $X$-valued Hilbert transform, defined by

$$
Hf(x) := \text{PV} \int_{\mathbb{R}} f(x - y) \frac{dy}{y} \quad \forall f \in \mathcal{S}(\mathbb{R}; X), \ x \in \mathbb{R},
$$

is bounded on $L^p(\mathbb{R}; X)$ for all $p \in (1, \infty)$.

UMD stands for Unconditionality of Martingale Differences, and naturally there is an equivalent definition in these terms [6, 7]. Various forms of Littlewood–Paley theory are accessible for UMD-valued functions (and their validity is equivalent to the UMD property); the particular Littlewood–Paley estimate we need appears below as Proposition 2.10. Examples of UMD spaces include Hilbert spaces, most reflexive function spaces (Lebesgue, Sobolev, Besov, and so on), as well as reflexive non-commutative $L^p$-spaces (which are not function spaces). Useful references on these spaces include [8, 26, 33].

Although UMD spaces are the most natural Banach spaces for harmonic analysis, we will need a seemingly stronger property called $r$-**intermediate UMD**. See [5] or [26, Appendix C] for interpolation theory jargon.

**Definition 2.2.** For $r \in [2, \infty)$, a Banach space $X$ is $r$-**intermediate UMD** if there exists a compatible couple $(Y, H)$, where $Y$ is UMD and $H$ is a Hilbert space, such that $X$ is isomorphic to the complex interpolation space $[Y, H]_{2/r}$. We say $X$ is **intermediate UMD** if it is $r$-intermediate UMD for some $r$.

For example, an infinite-dimensional Lebesgue space $L^p$ with $p \neq 2$ (either classical or non-commutative) is $r$-intermediate UMD provided that $r > \max(p, p')$. One should think of $r$-intermediate UMD spaces as being slightly nicer than $L^r$ (unless $r = 2$, as 2-intermediate UMD is the same as being isomorphic to a Hilbert space, which is already the best possible situation to be in). Rubio de Francia conjectured in [11] that every UMD space is actually intermediate UMD, but beyond the setting of Banach function spaces (discussed in Section 5), the conjecture is wide open.

#### 2.1.2 | Type, cotype, and variational estimates for convolutions

The concepts of type and cotype, and of martingale type and cotype, are important in the geometry of Banach spaces. We briefly introduce them here and discuss consequences for variational estimates for families of convolution operators. In the definition below, $(\varepsilon_n)_{n=1}^N$ denotes a finite
Rademacher sequence, that is, a sequence of independent random variables on a probability space taking the values ±1 with probability 1/2. Much more information is to be found in [27].

**Definition 2.3.** Let $X$ be a Banach space, $p \in [1, 2]$, and $q \in [2, \infty]$.

- The space $X$ has **type** $p$ if for all finite sequences $(x_j)$ in $X$ the estimate

  $$
  \mathbb{E} \left\| \sum_j \varepsilon_j x_j \right\|_X \lesssim \left( \sum_j \|x_j\|^p_X \right)^{1/p}
  $$

  holds, with implicit constant independent of the sequence. On the other hand, $X$ has **cotype** $q$ if the reverse estimate holds with $p$ replaced by $q$.

- The space $X$ has **martingale type** $p$ if for all finite $X$-valued $L^p$-martingales $(f_j)_{j=0}^N$ (on any measure space $S$, with respect to any $\sigma$-finite filtration) the estimate

  $$
  \|f_N\|_{L^p(S;X)} \lesssim \left( \|f_0\|_{L^p(S;X)}^p + \sum_{j=1}^N \|f_j - f_{j-1}\|_{L^p(S;X)}^p \right)^{1/p}
  $$

  holds with implicit constant independent of the martingale. On the other hand, $X$ has **martingale cotype** $q$ if the reverse estimate holds with $p$ replaced by $q$.

If $X$ has martingale type $p$, then it follows that $X$ has type $p$, but in general the converse is false. However, for UMD spaces, type $p$ is equivalent to martingale type $p$. The same statements are true for (martingale) cotype. See, for example, [33, §10.4].

Martingale cotype $q$ can be characterised by the vector-valued Lépingle inequality on the variation of martingale difference sequences, as shown by Pisier and Xu [34]. Thus, martingale cotype is tightly linked with variational estimates. We have in particular the following result for convolutions, which follows immediately from [20, Lemma 3.3].

**Theorem 2.4.** Let $X$ be a Banach space with martingale cotype $r_0 \in [2, \infty)$, and let $\varphi \in \mathcal{S}(\mathbb{R})$. Then for all $r > r_0$, we have

$$
\left\| t \mapsto (f \ast \text{Dil}_t \varphi)(\cdot) \right\|_{L^p(\mathbb{R};X)} \lesssim_{p, \varphi, X} \|f\|_{L^p(\mathbb{R};X)}
$$

for all $p \in (1, \infty)$. In particular, this estimate holds if $X$ is UMD with cotype $r_0$.

Every Banach space $X$ has type 1 and cotype $\infty$; the situation is more interesting if $X$ has non-trivial type (type $p$ for some $p > 1$) or finite cotype (cotype $q$ for some $q < \infty$). Here is what we need to know about type and cotype.

- UMD spaces have non-trivial type and finite cotype.
- $r$-intermediate UMD spaces have type $r'$ and cotype $r$.
- Type dualises: if $X$ has type $p$, then $X^*$ has cotype $p'$.

---

† This result follows from the vector-valued variational inequalities for averaging operators proved in [22], via an argument from [9]. Note that the cited lemma does not appear in the final published version [21] of the paper.
• Cotype dualises in presence of non-trivial type: if $X$ has cotype $q$ and non-trivial type, then $X^*$ has type $q'$.

• If $X$ has type $p$ and cotype $q$, then for each Hilbert space $H$ the space of $\gamma$-radonifying operators $\gamma(H, X)$ (defined very soon) also has type $p$ and cotype $q$.

2.1.3 $\gamma$-radonifying operators

Littlewood–Paley estimates are always phrased in terms of square functions. When dealing with UMD-valued functions, the appropriate notion of ‘square function’ is not an $L^2$-norm, but rather a $\gamma$-radonifying norm, as defined below. This concept is described in more depth in [3, §2.2] and [27, Chapter 9].

**Definition 2.5.** Let $H$ be a Hilbert space and $X$ a Banach space. A linear operator $T : H \to X$ is $\gamma$-summing if

$$
\|T\|_{\gamma_\infty(H, X)} := \left( \sup \mathbb{E} \left\| \sum_j \gamma_j T h_j \right\|_X^2 \right)^{1/2} < \infty
$$

with supremum taken over all finite orthonormal systems $(h_j)$ in $H$, with $(\gamma_j)$ a sequence of independent standard Gaussian random variables on some probability space. This norm defines the Banach space of $\gamma$-summing operators $\gamma_\infty(H, X)$. All finite rank operators $H \to X$ are $\gamma$-summing, and the Banach space of $\gamma$-radonifying operators $\gamma(H, X)$ is the closure of the finite rank operators in $\gamma_\infty(H, X)$.

**Remark 2.6.** When $X$ has finite cotype, and in particular if $X$ is UMD, we have $\gamma_\infty(H, X) = \gamma(H, X)$ for all Hilbert spaces $H$.

The space $\gamma(H, X)$ is a space of operators, but when $H$ is taken to be an $L^2$ space, it contains many elements that can be identified with $X$-valued functions.

**Definition 2.7.** For a Banach space $X$ and a measure space $(S, \mathcal{A}, \mu)$, we define

$$
\gamma(S; X) = \gamma_\mu(S; X) := \gamma(L^2_\mu(S), X).
$$

For a strongly measurable, weakly $L^2$ function $f : S \to X$ (meaning that the function $(f, x^*) : S \to \mathbb{C}$ is in $L^2(S)$ for all $x^* \in X^*$), we write $f \in \gamma(S; X)$ to mean that the Pettis integral operator $I_f : L^2(S) \to X$ defined by

$$
I_f g := \int_S f(s) g(s) \, d\mu(s)
$$

has cotype 2, but does not have non-trivial type. Its dual, $\epsilon^\infty$, has neither non-trivial type nor finite cotype.

This follows from the fact that $\gamma(H, X)$ is isomorphic to a closed subspace of $L^2(\Omega; X)$ for some probability space $\Omega$, which can be shown by identifying the $\gamma_\infty$-norm of a finite rank operator as the $L^2$-norm of a Gaussian sum [27, Proposition 9.1.3], and using that $\gamma(H, X)$ is the closure of the finite-rank operators in $\gamma_\infty(H, X)$. The Bochner space $L^2(\Omega; X)$ inherits type and cotype from $X$, and so do its closed subspaces.
is in $\gamma(S, X)$, and write

$$\|f\|_{\gamma(S; X)} := \|I_f\|_{\gamma(S; X^*)}.$$  

When $H$ is a Hilbert space and $S$ a measure space, we have $\gamma(S; H) = L^2(S; H)$ with equivalent norms [27, Proposition 9.2.9]. In general, one should think of $\gamma(S; X)$ as a function space analogous to $L^2(S; X)$, but better adapted to the geometry of $X$. One manifestation of this analogy is the $\gamma$-Hölder inequality [27, Theorem 9.2.14].

**Proposition 2.8.** Let $(S, A, \mu)$ be a measure space and $X$ a Banach space. Suppose $f : S \to X$ and $g : S \to X^*$ are in $\gamma(S; X)$ and $\gamma(S; X^*)$, respectively. Then $(f; g) : S \to \mathbb{C}$ is integrable, with

$$\int_S |\langle f; g \rangle| \, d\mu \leq \|f\|_{\gamma(S; X)} \|g\|_{\gamma(S; X^*)}.$$  

Conversely, if $X$ has non-trivial type and $f : S \to X$ is strongly measurable and weakly $L^2$, then $f \in \gamma(S; X)$ if and only if there exists a constant $C < \infty$ such that

$$\left| \int_S \langle f; g \rangle \, d\mu \right| \leq C \|g\|_{\gamma(S; X^*)}$$  

for all $g \in L^2(S) \otimes X^*$, in which case $\|f\|_{\gamma(S; X)} \lesssim_X C$.

The $\gamma$-spaces satisfy a Fubini-type relation with $L^p$-spaces, which shows one way in which $\gamma$-norms are not like classical $L^2$-norms. See [27, Theorem 9.4.8].

**Theorem 2.9.** Let $(S, A, \mu)$ be a $\sigma$-finite measure space, let $H$ be a Hilbert space, and $p \in [1, \infty)$. Then the mapping $U : L^p(S; \gamma(H, X)) \to \mathcal{L}(H, L^p(S; X))$ defined by

$$(Uf)h := f(\cdot)h \quad \forall h \in H$$

defines an isomorphism

$$L^p(S; \gamma(H; X)) \cong \gamma(H; L^p(S; X)).$$

When $S_1$ and $S_2$ are $\sigma$-finite measure spaces and $f : S_1 \times S_2 \to X$ is sufficiently nice, this says, in particular, that

$$\|f\|_{L^p(S_1; \gamma(S_2; X))} \approx \|f\|_{\gamma(S_2; L^p(S_1; X))}.$$  

UMD Banach spaces allow for various forms of Littlewood–Paley theory. We need the following estimate, written in terms of $\gamma$-norms, which we proved in [3, Theorem 2.21].

**Proposition 2.10.** Let $X$ be a UMD Banach space and $p \in (1, \infty)$. Fix a Schwartz function $\psi \in \mathcal{S}(\mathbb{R})$ such that $0 \notin \text{spt} \hat{\psi}$. Then for all $f \in L^p(\mathbb{R}; X)$,

$$\|f * \text{Dil}_t \psi\|_{L^p(\mathbb{R}; \gamma_+^1(\mathbb{R}_+; X))} \lesssim_{X, p, \psi} \|f\|_{L^p(\mathbb{R}; X)}.$$
The implicit constant depends only on the second-order Schwartz seminorms of \( \psi \) and the distance of \( \text{spt} \psi \) to the origin.

We will make use of the following lemma, which controls the Bochner integral of the product of a vector-valued function and a scalar-valued function by the product of corresponding \( L^2 \)- and \( \gamma \)-norms.

**Lemma 2.11.** Let \( X \) be a Banach space and \( (S, \mathcal{A}, \mu) \) a measure space. Then for all separably valued \( f : S \to X \) in \( \gamma(S;X) \) and all \( g \in L^2(S;\mathbb{C}) \), the function \( gf : S \to X \) is strongly integrable and

\[
\left\| \int_S g(s)f(s) \, d\mu(s) \right\|_X \leq \|g\|_{L^2(S)} \|f\|_{\gamma(S;X)}.
\]

**Proof.** Since \( f \) is weakly \( L^2 \) and \( g \in L^2(S) \), we find that \( gf \) is weakly integrable. By the Pettis measurability theorem, since \( f \) is separably valued, \( gf \) is strongly integrable. Duality and Proposition 2.8 then yield

\[
\left\| \int_S g(s)f(s) \, d\mu(s) \right\|_X = \sup_{x^* \in X, \|x^*\|_{X^*} = 1} \left\langle \int_S g(s)f(s) \, d\mu(s); x^* \right\rangle
\]

\[
\leq \sup_{x^*} \int_S \left| \langle f(s); \overline{g(s)x^*} \rangle \right| \, d\mu(s)
\]

\[
\leq \sup_{x^*} \|f\|_{\gamma(S;X)} \|\overline{g} \otimes x^*\|_{\gamma(S;X^*)}
\]

\[
= \|f\|_{\gamma(S;X)} \|g\|_{L^2(S)}
\]

as claimed. \( \square \)

### 2.1.4 R-bounds

No discussion of Banach-valued analysis is complete without a section on the \( R \)-bound of a family of operators. This important concept is explained much more thoroughly in [27, Chapter 8].

**Definition 2.12.** Let \( X \) and \( Y \) be Banach spaces and \( \mathcal{T} \subset \mathcal{L}(X,Y) \) a set of operators. The set \( \mathcal{T} \) is **\( R \)-bounded** if there exists a constant \( C < \infty \) such that for all finite sequences \( (T_j) \) in \( \mathcal{T} \) and \( (x_j) \) in \( X \), whenever \( (\varepsilon_j) \) is a finite Rademacher sequence,

\[
\mathbb{E} \left\| \sum_j \varepsilon_j T_j x_j \right\|_Y \leq C \mathbb{E} \left\| \sum_j \varepsilon_j x_j \right\|_X.
\]

The infimum of all possible \( C \) in this estimate is called the **\( R \)-bound** of \( \mathcal{T} \), and denoted by \( R(\mathcal{T}) \).

\( R \)-boundedness is generally stronger than uniform boundedness (unless \( Y \) has type 2 and \( X \) has cotype 2, in which case the two notions are equivalent). The \( R \)-bound is to the \( \gamma \)-norm as \( L^\infty \) is to
$L^2$; this is made precise in the following theorem (see [27, Theorem 9.5.1 and Remark 9.5.8]), and it is why we need to discuss $R$-bounds.

**Theorem 2.13 (γ-multiplier theorem).** Let $X$ and $Y$ be Banach spaces with finite cotype, and $(S, A, \mu)$ a measure space. Let $A : S \to \mathcal{L}(X, Y)$ be such that for all $x \in X$ the $Y$-valued function $s \mapsto A(s)(x)$ is strongly $\mu$-measurable, and that the range $A(S)$ is $R$-bounded. Then for every $f : S \to X$ in $\gamma(S; X)$, the function $Af : S \to Y$ is in $\gamma(S; Y)$, with

$$\|Af\|_{\gamma(S; Y)} \lesssim R(A(S))\|f\|_{\gamma(S; X)}.$$

In particular, for a Banach space $X$ with finite cotype, we have

$$\|a \cdot f\|_{\gamma(S; X)} \lesssim \|f\|_{\gamma(S; X)}$$

for all $a \in L^\infty(S)$, as each bounded set of scalar operators $\{cI : c \in \mathbb{C}, |c| \leq N\}$ is $R$-bounded as a consequence of Kahane’s contraction principle for Rademacher sums.

Controlling the $R$-bound of a family of operators is generally subtle, and depends strongly on the structure of the family. We will need to consider families of convolution operators in order to prove an important technical lemma (Lemma 2.18). But first we will need $R$-bounds for families of Fourier multipliers with symbols of bounded variation, which is a special case of [27, Theorem 8.3.4]. For a bounded function $m \in L^\infty(\mathbb{R}; \mathbb{C})$, let $T_m$ denote the Fourier multiplier with symbol $m$.

**Theorem 2.14.** Let $X$ be a UMD space and $p \in (1, \infty)$. Then for all $C > 0$,

$$R\left(\{T_m \in \mathcal{L}(L^p(\mathbb{R}; X)) : \|m\|_{V^1(\mathbb{R}; \mathbb{C})} \leq C\}\right) \lesssim p, X C.$$

As a corollary, we obtain $R$-bounds for families of convolutions with dilates of a fixed nice (but not necessarily Schwartz) function. Later this will be applied to the functions $\varphi(x) = \langle x \rangle^{-N}$ for $N$ sufficiently large.

**Corollary 2.15.** Let $X$ be a UMD space and $p \in (1, \infty)$. Fix an integrable function $\varphi \in L^1(\mathbb{R})$ such that $\hat{\varphi}$ has bounded variation. Then

$$R(\{[f \mapsto f \ast \text{Dil}_t \varphi] \in \mathcal{L}(L^p(\mathbb{R}; X)) : t > 0\} \lesssim (p, X) \|\hat{\varphi}\|_{V^1(\mathbb{R}; \mathbb{C})}.$$

**Proof.** For each $t > 0$, we have $\|\text{Dil}_t \varphi\|_{V^1(\mathbb{R}; \mathbb{C})} = \|\hat{\varphi}\|_{V^1(\mathbb{R}; \mathbb{C})}$ since $\text{Dil}_t \varphi$ is a rescaling of $\hat{\varphi}$. The result then follows from Theorem 2.14, since $f \mapsto f \ast \text{Dil}_t \varphi$ is the Fourier multiplier with symbol $\text{Dil}_t \varphi$. □

This result extends to families of generalised convolutions, in which the convolving function (‘$g$’ in ‘$f * g$’) is allowed to depend on the variable $x \in \mathbb{R}$, as in Lemma 2.17. In what follows, it will be convenient to write

$$\text{Conv}_g f := f * g \quad \text{and} \quad \text{Mult}_g f := fg.$$
Remark 2.16. In the following two lemmas we will speak of piecewise smooth functions \( P : \mathbb{R} \times \mathbb{R}_+ \to \mathcal{S}(\mathbb{R}) \). By this, we mean that there exists a sequence of points \( 0 = t_0 < t_1 < t_2 < \cdots \) with \( t_n \to \infty \) such that the restrictions \( P|_{\mathbb{R} \times (t_i, t_{i+1})} \) are smooth. We use this concept purely to avoid discussion of measurability of \( \mathcal{S}(\mathbb{R}) \)-valued functions.

**Lemma 2.17.** Let \( X \) be a UMD Banach space and \( p \in (1, \infty) \). Given a piecewise smooth function \( P : \mathbb{R} \times \mathbb{R}_+ \to \mathcal{S}(\mathbb{R}) \), for each \( t > 0 \) define the bounded operator \( C_{P,t} : L^p(\mathbb{R}; X) \to L^p(\mathbb{R}; X) \) by

\[
(C_{P,t}f)(x) := (f \ast \text{Dil}_t P_{x,t})(x).
\]

Then we have the R-bound

\[
\mathcal{R}\{C_{P,t} : t > 0\} \lesssim p, X \sup_{x, t} \|P_{x,t}\|_s,
\]

where \( \| \cdot \|_s \) is a Schwartz seminorm of sufficiently high order.

**Proof.** First assume \( f \in \mathcal{S}(\mathbb{R}; X) \). Fix \( N \) large, and write by Fourier inversion

\[
(f \ast \text{Dil}_t P_{x,t})(x) = \int_{\mathbb{R}} f(y) \text{Dil}_t P_{x,t}(x-y) \, dy
\]

\[
= \int_{\mathbb{R}} f(y) \frac{1}{t} \left( \frac{x - y}{t} \right)^{-N} \bar{P}_{x,t}(x-y) \, dy
\]

\[
= \int_{\mathbb{R}} f(y) \frac{1}{t} \left( \frac{x - y}{t} \right)^{-N} \int_{\mathbb{R}} \bar{P}_{x,t}(\xi)e^{i\xi(x-y)/t} \, d\xi \, dy
\]

\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-i\xi y/t} f(y) \frac{1}{t} \left( \frac{x - y}{t} \right)^{-N} \, dy \, e^{i\xi x/t} \bar{P}_{x,t}(\xi) \, d\xi
\]

\[
= \int_{\mathbb{R}} (\text{Mod}_{\xi/t} f \ast \text{Dil}_t J_N)(x) P_{\xi,t}^\vee(x) \langle \xi \rangle^{-N} \, d\xi,
\]

where

\[
\bar{P}_{x,t}(w) := \langle w \rangle^N P_{x,t}(w)
\]

\[
J_N(w) := \langle w \rangle^{-N}
\]

\[
P_{\xi,t}^\vee(x) := \langle \xi \rangle^N \text{Mod}_{-x/t} \bar{P}_{x,t}(\xi).
\]

Thus we have

\[
(f \ast \text{Dil}_t P_{x,t})(x) = \int_{\mathbb{R}} (T_{\xi,t}f)(x) \langle \xi \rangle^{-N} \, d\xi
\]

for all \( x \in \mathbb{R} \), where

\[
T_{\xi,t}^P := \text{Mult}_{P_{\xi,t}^\vee} \text{Conv}_{\text{Dil}_t J_N \text{Mod}_{\xi/t}}.
\]
By density of the Schwartz functions in $L^p(\mathbb{R}; X)$, we get the representation
\[ C_{p,t} = \int_{\mathbb{R}} T^p_{\hat{\xi},t} \langle \xi \rangle^{-N} \, d\xi \]
in the strong operator topology on $L^p(\mathbb{R}; X)$. Using [27, Theorem 8.5.2], it follows that
\[ R\{C_{p,t} : t > 0\} \lesssim R\{T^p_{\hat{\xi},t} : t > 0, \xi \in \mathbb{R}\}. \quad (2.2) \]
Since the functions $P^\varnothing_{\hat{\xi},t}$ satisfy
\[ \sup_{\xi \in \mathbb{R}, t > 0} \| P^\varnothing_{\hat{\xi},t} \| \lesssim \sup_{x \in \mathbb{R}, t > 0} \sup_{\xi \in \mathbb{R}} |\langle \xi \rangle|^N \hat{P}_x(t) \lesssim \sup_{x \in \mathbb{R}, t > 0} \| P_{x,t} \|_* \],
the operators $\{\text{Mult}_{P^\varnothing_{\hat{\xi},t}} : \xi \in \mathbb{R}\}$ have $R$-bound controlled by $\sup_{x,t} \| P_{x,t} \|_*$. Furthermore, by Corollary 2.15, the operators $\{\text{ConvDil}_{t,j} J_N : t > 0\}$ have $R$-bound controlled by a constant depending on $p$ and $X$ (for sufficiently large $N$). Finally, since modulations are just multiplication by unimodular functions, the operators $\{\text{Mod}_{\hat{\xi}/t} : \xi \in \mathbb{R}\}$ have $R$-bound controlled by 1. Using (2.2) and (2.1), the claim follows. □

2.1.5 | An important technical lemma

All of the concepts described above — UMD, cotype, $\gamma$-norms, $R$-bounds — come together in a key technical lemma that we will use in bounding various error terms.

**Lemma 2.18.** Let $Y$ be a UMD space with cotype $r \geq 2$. Let $N_\pm : \mathbb{N} \times \mathbb{R} \to \mathbb{R}_+$ be measurable functions with $N_-(j,z) < N_+(j,z)$ for all $(j,z) \in \mathbb{N} \times \mathbb{R}$. For each $z \in \mathbb{R}$, let $I_{j,z}$ be the interval $(N_-(j,z), N_+(j,z))$, and suppose that the intervals $(I_{j,z})_{j \in \mathbb{N}}$ have finite overlap. Fix a measurable function $q : \mathbb{N} \times \mathbb{R} \times \mathbb{R}_+ \to \mathbb{C}$ and a function $P : \mathbb{N} \times \mathbb{R} \times \mathbb{R}_+ \to \mathcal{S}(\mathbb{R})$ such that for each $j \in \mathbb{N}$, the function $P_j : \mathbb{R} \times \mathbb{R}_+ \to \mathcal{S}(\mathbb{R})$ is piecewise smooth (in the sense of Remark 2.16). Then for all $h \in L^2(\mathbb{R}; \gamma(\mathbb{R}; Y))$, we have
\[
\left\| \int_{I_{j,z}} h(\cdot, \sigma) * \text{Dil}_\sigma P_{j,z,\sigma}(z) q_{j,z}(\sigma) \frac{d\sigma}{\sigma} \right\|_{L^2_{\mathcal{S}}(\mathbb{R}; X_j(z)_Y))} \lesssim \| h \|_{L^2(\mathbb{R}; \gamma(\mathbb{R}; Y))} \sup_{j,z,\sigma} \| P_{j,z,\sigma} \|_* \sup_{j,z} \| q_{j,z}(\sigma) \|_{1^2_{\mathcal{S}}(I_{j,z})},
\]
where $\| \cdot \|_*$ denotes a Schwartz seminorm of sufficiently high order.

**Proof.** For $j \in \mathbb{N}$ and $z \in \mathbb{R}$, we have
\[
\left\| \int_{I_{j,z}} h(\cdot, \sigma) * \text{Dil}_\sigma P_{j,z,\sigma}(z) q_{j,z}(\sigma) \frac{d\sigma}{\sigma} \right\|_Y \lesssim \| q_{j,z} \|_{1^2_{\mathcal{S}}(I_{j,z})} \| L^1_{\mathcal{S}}(\mathbb{R}; \gamma(\mathbb{R}; \mathbb{R}_+; Y)) \|
\]
by Lemma 2.11. Since $\gamma_{d/\sigma}(\mathbb{R}_+; Y)$ has cotype $r$, we then get

$$
\left\| \int_{I_{j,z}} h(\cdot, \sigma) * \text{Dil}_\sigma P_{j,z,\sigma}(z) q_{j,z}(\sigma) \frac{d\sigma}{\sigma} \right\|_{\ell_r^r(\mathbb{N}; Y)} \lesssim \sup_{j \in \mathbb{N}} \| q_{j,z} \|_{L^2_{d/\sigma}(I_{j,z})} \left( \sum_{j \in \mathbb{N}} \left\| 1_{I_{j,z}}(\sigma) h(\cdot, \sigma) * \text{Dil}_\sigma P_{j,z,\sigma}(z) \right\|_{Y_{d/\sigma}(\mathbb{R}_+; Y)}^r \right)^{1/r}.
$$

where $(\varepsilon_j)_{j \in \mathbb{N}}$ is a Rademacher sequence on a probability space $\Omega$. Taking the $L^2$-norm in $z$ and applying the $\gamma$-Fubini theorem gives

$$
\left\| \int_{I_{j,z}} h(\cdot, \sigma) * \text{Dil}_\sigma P_{j,z,\sigma}(z) q_{j,z}(\sigma) \frac{d\sigma}{\sigma} \right\|_{L^2_{d/\sigma}(\mathbb{R}; Y)} \lesssim \sup_{j,z} \| q_{j,z} \|_{L^2_{d/\sigma}(I_{j,z})} \mathbb{E} \left\| \sum_{j \in \mathbb{N}} \varepsilon_j 1_{I_{j,z}}(\sigma) h(\cdot, \sigma) * \text{Dil}_\sigma P_{j,z,\sigma}(z) \right\|_{Y_{d/\sigma}(\mathbb{R}_+; Y)}.
$$

It remains to control the second factor. By the $\gamma$-multiplier theorem and another application of the $\gamma$-Fubini theorem, for all $\omega \in \Omega$, we have

$$
\left\| \sum_{j \in \mathbb{N}} \varepsilon_j(\omega) 1_{I_{j,z}}(\sigma) h(\cdot, \sigma) * \text{Dil}_\sigma P_{j,z,\sigma}(z) \right\|_{Y_{d/\sigma}(\mathbb{R}_+; L^2_{dz}(\mathbb{R}; Y))} \lesssim R\{C_{\tilde{P}_{z,\sigma}} : \sigma > 0\} \| h \|_{L^2_{d/\sigma}(\mathbb{R}; Y)}.
$$

where the operators $C_{\tilde{P}_{z,\sigma}} \in \mathcal{L}(L^2_{dz}(\mathbb{R}; Y))$ are defined as in Lemma 2.17 with

$$
\tilde{P}_{z,\sigma} := \sum_{j \in \mathbb{N}} \varepsilon_j(\omega) 1_{I_{j,z}}(\sigma) P_{j,z,\sigma}(z).
$$

Lemma 2.17 and finite overlap of the intervals $I_{j,z}$ then gives

$$
R\{C_{\tilde{P}_{z,\sigma}} : \sigma > 0\} \lesssim \sup_{z,\sigma} \| \tilde{P}_{z,\sigma} \|_* \lesssim \sup_{j,z,\sigma} \| P_{j,z,\sigma} \|_*,
$$

with implicit constants independent of $\omega \in \Omega$, which after an integration in $\Omega$ completes the proof. \qed
2.2 Outer Lebesgue spaces

We quickly describe abstract outer spaces and the associated outer Lebesgue quasi-norms. We use the formulation in [3, §2.5], to which the reader is referred for a detailed exposition. Here we review the main points.

For a topological space $\mathcal{X}$, we let $\mathcal{B}(\mathcal{X})$ denote the $\sigma$-algebra of Borel sets in $\mathcal{X}$, and for a Banach space $X$ we let $\mathcal{B}(\mathcal{X}; X)$ denote the set of strongly Borel measurable functions $\mathcal{X} \to X$.

Definition 2.19. Let $\mathcal{X}$ be a topological space.

- A $\sigma$-generating collection on $\mathcal{X}$ is a subset $\mathcal{B} \subset \mathcal{B}(\mathcal{X})$ such that $\mathcal{X}$ can be written as a union of countably many elements of $\mathcal{B}$. We denote by $\mathcal{B}^\cup$ the set of at-most-countable unions of elements of $\mathcal{B}$:

  $$\mathcal{B}^\cup := \left\{ E \in \mathcal{B}(\mathcal{X}) : E = \bigcup_{n \in \mathbb{N}} B_n \text{ with } B_n \in \mathcal{B} \cup \{\emptyset\} \text{ for all } n \in \mathbb{N} \right\}.$$

- A local measure (on $\mathcal{B}$) is a $\sigma$-subadditive function $\mu : \mathcal{B} \cup \{\emptyset\} \to [0, \infty]$ such that $\mu(\emptyset) = 0$.

- Given a Banach space $X$, an $X$-valued local size (on $\mathcal{B}$) is a collection of quasi-norms $\mathcal{S} = (\| \cdot \|_{\mathcal{S}(B)})_{B \in \mathcal{B}}$, with each $\| \cdot \|_{\mathcal{S}(B)} : \mathcal{B}(\mathcal{X}; X) \to [0, \infty]$, satisfying

  - **Positive homogeneity** $\| \lambda F \|_{\mathcal{S}(B)} = |\lambda| \| F \|_{\mathcal{S}(B)}$ for all $F \in \mathcal{B}(\mathcal{X}; X)$ and $\lambda \in \mathbb{C}$;
  
  - **Global positive-definiteness** $\| F \|_{\mathcal{S}(B)} = 0$ for all $B \in \mathcal{B}$ if and only if $F = 0$;
  
  - **Quasi-triangle inequality** there exists a constant $C \in [1, \infty)$ such that $\| F + G \|_{\mathcal{S}(B)} \leq C(\| F \|_{\mathcal{S}(B)} + \| G \|_{\mathcal{S}(B)})$ for all $B \in \mathcal{B}$ and $F, G \in \mathcal{B}(\mathcal{X}; X)$.

- An $(X$-valued) outer space is a tuple $(\mathcal{X}, \mathcal{B}, \mu, \mathcal{S})$ consisting of a topological space $\mathcal{X}$, a $\sigma$-generating collection $\mathcal{B}$ on $\mathcal{X}$, a local measure $\mu$, and an $X$-valued local size $\mathcal{S}$, all as above.

We often do not make reference to the Banach space $X$, which is encoded in the domain of $\mathcal{S}$.

Given an outer space $(\mathcal{X}, \mathcal{B}, \mu, \mathcal{S})$, we extend $\mu$ to an outer measure on $\mathcal{X}$ via countable covers: for all $E \subset \mathcal{X}$,

$$\mu(E) := \inf \left\{ \sum_{n \in \mathbb{N}} \mu(B_n) : B_n \in \mathcal{B} \cup \{\emptyset\}, \bigcup_{n \in \mathbb{N}} B_n \supset E \right\}.$$

We abuse notation and write $\mu$ for both the local measure and the corresponding outer measure. The generated outer measure extends the local measure $\mu$ under the reasonable assumption on the local measure that for any $B' \in \mathcal{B}$, we have

$$B' \subset \bigcup_{n \in \mathbb{N}} B_n \text{ with } B_n \in \mathcal{B} \cup \{\emptyset\} \Rightarrow \mu(B') \leq \sum_{n \in \mathbb{N}} \mu(B_n).$$

We define the outer size (or outer supremum) of $F \in \mathcal{B}(\mathcal{X}; X)$ by

$$\| F \|_{\mathcal{S}} := \sup_{B \in \mathcal{B}} \sup_{V \in \mathcal{B}^\cup} \| 1_{\mathcal{X} \setminus V} F \|_{\mathcal{S}(B)};$$
this definition guarantees in particular that
\[
\|1_{\chi \setminus E}F\|_S \lesssim \|F\|_S \quad \text{and} \quad \|1_E F\|_S \lesssim \|F\|_S \quad \forall E \in \mathcal{B}^U.
\] (2.3)

We say that two local sizes \(S_1, S_2\) on \(B\) are equivalent if
\[
\|F\|_{S_1} \simeq \|F\|_{S_2}
\]
for all \(F \in \mathcal{B}(\chi; X)\). Similarly, we say outer measures \(\mu_1\) and \(\mu_2\) are equivalent if
\[
\mu_1(E) \simeq \mu_2(E) \quad \forall E \in \mathcal{B}(\chi)
\]
with an implicit constant independent of \(E\). The collection of all (local) outer measures and local/outer sizes are cones, being closed with respect to positive linear combinations, that is, the expressions
\[
(a_1 \mu_1 + a_2 \mu_2)(E) := a_1 \mu_1(E) + a_2 \mu_2(E),
\]
\[
\|F\|_{a_1 S_1 + a_2 S_2} := a_1 \|F\|_{S_1} + a_2 \|F\|_{S_2}
\]
define outer measures and outer sizes for \(a_{1,2} \geq 0, (a_1, a_2) \neq (0, 0)\). Given local sizes indexed by \(B \in \mathcal{B}\), we define
\[
\|F\|_{(a_1 S_1 + a_2 S_2)(B)} := a_1 \|F\|_{S_1(B)} + a_2 \|F\|_{S_1(B)} \quad \forall B \in \mathcal{B}
\]
so that the corresponding outer size is equivalent to the combination \(a_1 S_1 + a_2 S_2\) of respective outer sizes.

At the heart of this integration theory is a way of relating the outer measures and outer sizes: we define the outer super-level measure of \(F\) by
\[
\mu(\|F\|_S > \lambda) := \inf \{\mu(V) : V \in \mathcal{B}^U, \|1_{\chi \setminus V} F\|_S \leq \lambda\} \quad \forall \lambda \geq 0.
\]
Notice that \(\mu(\|F\|_S > \lambda)\) need not be the outer measure of any particular set; it should be thought of as a generalised distribution function. We also define the ‘measure of the support’ by
\[
\mu(\text{spt}(F)) := \mu(\|F\|_S > 0).
\]

**Definition 2.20.** Let \((\chi, \mathcal{B}, \mu, S)\) be an \(X\)-valued outer space. We define the outer Lebesgue quasi-norms of a function \(F \in \mathcal{B}(\chi; X)\), and their weak variants, by setting
\[
\|F\|_{L_p^S} := \left( \int_0^\infty \lambda^p \mu(\|F\|_S > \lambda) \frac{d\lambda}{\lambda} \right)^{1/p} \quad \forall p \in (0, \infty),
\]
\[
\|F\|_{L_p^\infty} := \sup_{\lambda > 0} \lambda \mu(\|F\|_S > \lambda)^{1/p} \quad \forall p \in (0, \infty),
\]
\[
\|F\|_{L_\infty^S} := \|F\|_S.
\]
It is straightforward to check that these are quasi-norms (modulo functions \( F \) with \( \mu(\text{spt}(F)) = 0 \)), and that equivalent local sizes and local measures give rise to equivalent outer Lebesgue quasi-norms.

We construct *iterated outer spaces* by using an outer Lebesgue quasi-norms themselves as local sizes.

**Definition 2.21.** Let \((\mathcal{X}, \mathcal{B}, \mu, \mathcal{S})\) be an outer space. Let \(\mathcal{B}'\) be a \(\sigma\)-generating collection on \(\mathcal{X}\), and let \(\nu\) be a local measure on \(\mathcal{B}'\). Then for all \(q \in (0, \infty)\) we define the *iterated local size* \(L^q_{\mu, \mathcal{S}}\) on \(\mathcal{B}'\) (which depends on \(\nu\)) by

\[
\|F\|_{L^q_{\mu, \mathcal{S}}(\mathcal{B}')} := \frac{1}{\nu(B')}^{1/q} \|1_{B'} F\|_{L^q_{\mu, \mathcal{S}}} \quad (B' \in \mathcal{B}').
\] (2.4)

The expression above makes sense if \(\nu(B') \neq 0\) for all \(B' \in \mathcal{B}'\), and it defines a local size if the quasi-triangle inequality is satisfied with a common constant for all \(B' \in \mathcal{B}'\). In our applications, this is always the case and is straightforward to check.

We call \((\mathcal{X}, \mathcal{B}', \nu, L^q_{\mu, \mathcal{S}})\) an *iterated* outer space. On this space, one can then construct outer Lebesgue norms \(\|\cdot\|_{L^p \nu, L^q_{\mu, \mathcal{S}}}\) following Definition 2.20.

We will use the following properties of outer Lebesgue quasi-norms. The first is a Radon–Nikodym-type domination result.

**Proposition 2.22.** Let \((\mathcal{X}, \mathcal{B}, \mu, \mathcal{S})\) be an outer space such that the outer measure generated by \(\mu\) is \(\sigma\)-finite. Let \(\mathcal{m}\) be a positive Borel measure on \(\mathcal{X}\) such that

\[
\int_B \|F(x)\|_X \, d\mathcal{m}(x) \lesssim \|F\|_{\mathcal{S}(B)} \mu(B) \quad \forall B \in \mathcal{B}, \forall F \in \mathcal{B}(\mathcal{X}; X)
\]

and

\[
\mu(A) = 0 \Rightarrow \mathcal{m}(A) = 0 \quad \forall A \in \mathcal{B}(\mathcal{X}).
\]

Then we have

\[
\int_{\mathcal{X}} \|F(x)\|_X \, d\mathcal{m}(x) \lesssim \|F\|_{L^1_{\mu, \mathcal{S}}} \quad \forall F \in L^1(\mathcal{X}, \mathcal{m}; X).
\]

The proof is a straightforward modification of [36, Lemma 2.2] and [17, Proposition 3.6].

**Proof.** First we show that for any Borel subset \(E \subset \mathcal{X}\) the bound

\[
\int_E \|F(x)\|_X \, d\mathcal{m}(x) \lesssim \|1_E F\|_{\mathcal{S}} \mu(E)
\] (2.5)

holds. We start by selecting an at-most-countable collection \(B \subset \mathcal{B}\) such that

\[
E \subset \bigcup_{B \in B} B \quad \text{and} \quad \sum_{B \in B} \mu(B) \leq \mu(E),
\]
and after having enumerated $B$ we set $\Delta B_n := B_n \setminus \bigcup_{n' < n} B_{n'}$, so that $E \subset \bigcup_{n \in \mathbb{N}} B_n$ and

$$
\int_E \|F(x)\|_X \, dm(x) \leq \sum_{n \in \mathbb{N}} \int_{B_n} \|1_{\mathbb{X} \setminus \bigcup_{n' < n} B_{n'}} F(x)\|_X \, dm(x)
$$

$$
\leq \sum_{n \in \mathbb{N}} \mu(B_n) \|1_{\mathbb{X} \setminus \bigcup_{n' < n} B_{n'}} F\|_S \leq \mu(E) \|1_E F\|_S,
$$

having used (2.3). This shows the bound (2.5).

Now we are ready to show the main claim. Given $F \in L^1_\mu S$, we proceed by choosing sets $E_n \in \mathcal{B} \cup \mathcal{F}$ for each $n \in \mathbb{Z}$ such that

$$
\mu(E_n) \approx \mu(\|F\|_S > 2^n) \quad \text{and} \quad \|1_{\mathbb{X} \setminus E_n} F\|_S \leq 2^n,
$$

and thus

$$
\sum_{n \in \mathbb{Z}} \mu(E_n) 2^n \approx \|F\|_{L^1_\mu S},
$$

We may assume that $E_n \supset E_{n+1}$ for any $n \in \mathbb{Z}$ by using the sets $\bigcup_{n' < n} E_{n'}$ in place of $E_n$. It follows from (2.3) that

$$
\|1_{\mathbb{X} \setminus \bigcup_{n' > n} E_{n'}} F\|_S \lessapprox 2^n
$$

by noticing that $\|1_{\mathbb{X} \setminus \bigcup_{n' > n} E_{n'}} F\|_S = \|1_{\mathbb{X} \setminus \bigcup_{n' > n} E_{n'} \setminus E_n} F\|_S$. Furthermore,

$$
\sum_{n \in \mathbb{Z}} \mu \left( \bigcup_{n' \geq n} E_{n'} \right) 2^n \approx \sum_{n \in \mathbb{Z}} \mu(E_n) 2^n \sum_{n \leq n'} 2^{n-n'} \lesssim \|F\|_{L^1_\mu S}.
$$

Now we set

$$
F_n := 1_{E_n \setminus E_{n+1}} F,
$$

$$
F_{+\infty} := 1_{\bigcap_{n \in \mathbb{Z}} E_n} F,
$$

$$
F_{-\infty} := 1_{\mathbb{X} \setminus \bigcup_{n \in \mathbb{Z}} E_n} F,
$$

and note that $\|F_{-\infty}\|_S = 0$, and thus $F_{-\infty} = 0$, so that

$$
\int_X \|F(x)\|_X \, dm(x) = \int_{\bigcap_{n \in \mathbb{Z}} E_n} \|F_{+\infty}(x)\|_X \, dm(x) + \sum_{n \in \mathbb{Z}} \int_{E_n} \|F_n(x)\|_X \, dm(x)
$$

$$
\lesssim \sum_{n \in \mathbb{Z}} \mu(E_n) \|F_n\|_S \lesssim \sum_{n \in \mathbb{Z}} \mu(E_n) 2^n \lesssim \|F\|_{L^1_\mu S},
$$

as required. The term with $F_{+\infty}$ vanishes since the integral is taken over a set of vanishing measure $m$; as a matter of fact, the set $\bigcap_{n \in \mathbb{Z}} E_n$ satisfies $\mu(\bigcap_{n \in \mathbb{Z}} E_n) \leq \inf_{n \in \mathbb{Z}} \mu(E_n) = 0$ and thus $m(\bigcap_{n \in \mathbb{Z}} E_n) = 0$. \qed
Next is an ‘outer Hölder inequality’.

**Proposition 2.23.** Let \( \mathbb{X} \) be a topological space, \( \mathcal{B} \) a \( \sigma \)-generating collection on \( \mathbb{X} \), and \( \mu \) a local measure on \( \mathcal{B} \). Let \( X \) be a Banach space, \( S \) an \( X \)-valued local size, \( S^* \) an \( X^* \)-valued local size, and \( S_C \) a \( C \)-valued local size (all on \( \mathcal{B} \)). Suppose these local sizes satisfy the Hölder relation

\[
\|\langle F; G \rangle\|_{S_C} \lesssim \|F\|_S \|G\|_{S^*},
\]  

(2.6)

for all \( F \in \mathcal{B}(\mathbb{X}; X) \) and \( G \in \mathcal{B}(\mathbb{X}; X^*) \), where \( \langle F; G \rangle(x) := \langle F(x); G(x) \rangle \) for \( x \in \mathbb{X} \). Then for all \( p \in (0, \infty] \), we have

\[
\|\langle F; G \rangle\|_{L_p^\mu S_C} \lesssim_p \|F\|_{L_p^\mu S} \|G\|_{L_p'\mu S^*}.
\]  

(2.7)

A proof in the trilinear setting already appears in [2, Proposition 4.3]. We reproduce it here in the bilinear setting.

**Proof.** Assume that the factors on the right-hand side of (2.7) are finite and non-zero, for otherwise there is nothing to prove. By homogeneity, we may assume that \( \|F\|_{L_p^\mu S} = \|G\|_{L_p'\mu S^*} = 1 \). For each \( n \in \mathbb{Z} \) and some \( C > 1 \) to be determined later, let \( E_n, E_n^* \in \mathcal{B} \cup \) be such that

\[
\mu(E_n) \approx \mu(\|F\|_S > C^{-1}2^n), \quad \|1_{\mathbb{X}\setminus E_n} F\|_S \leq C^{-1}2^n, \quad \|1_{\mathbb{X}\setminus E_n} F\|_S \leq 2^n,
\]

\[
\mu(E_n^*) \approx \mu(\|G\|_{S^*} > C^{-1}2^n), \quad \|1_{\mathbb{X}\setminus E_n^*} G\|_S \leq C^{-1}2^n, \quad \|1_{\mathbb{X}\setminus E_n^*} G\|_S < C^{-1}2^n
\]

so that

\[
\sum_{n \in \mathbb{Z}} 2^n \mu(E_n) \leq 1 \quad \text{and} \quad \sum_{n \in \mathbb{Z}} 2^n \mu(E_n^*) \leq 1.
\]

According to the assumption (2.6) on the sizes \( S_C, S, S^* \) as well as the bound (2.3), we have that

\[
\|1_{\mathbb{X}\setminus (E_n \cup E_n^*)} FG\|_{S_C} \lesssim \|1_{\mathbb{X}\setminus (E_n \cup E_n^*)} F\|_S \|1_{\mathbb{X}\setminus (E_n \cup E_n^*)} G\|_{S^*} \lesssim C^{-2}2^n.
\]

If \( C \) is chosen large enough, this implies that

\[
\mu(\|FG\|_{S_C} > 2^n) \lesssim \mu(E_n) + \mu(E_n^*)
\]

and thus

\[
\sum_{n \in \mathbb{Z}} 2^n \mu(\|FG\|_{S_C} > 2^n) \leq \sum_{n \in \mathbb{N}} 2^n (\mu(E_n) + \mu(E_n^*)) \leq 1.
\]

The desired claim then follows by noticing that

\[
\|FG\|_{L_p^\mu S_C} = \int_{0}^{+\infty} \lambda \mu(\|FG\|_{S_C} > \lambda) \frac{d\lambda}{\lambda} \lesssim \sum_{n \in \mathbb{Z}} 2^n \mu(\|FG\|_{S_C} > 2^n).
\]

\[\square\]
Outer Lebesgue quasi-norms are typically estimated by interpolation. For a proof of the following result, see [17, Proposition 3.5].

**Proposition 2.24.** Let $(X, B, \mu, \mathcal{S})$ be an $X$-valued outer space. Let $\Omega$ be a $\sigma$-finite measure space, and let $T$ be a quasi-sublinear operator mapping $L^{p_0}(\Omega; X) + L^{p_1}(\Omega; X)$ into $\mathcal{B}(X; X)$ for some $1 \leq p_0 < p_1 \leq \infty$. Suppose that

$$
\|Tf\|_{L^{p_0, \infty} \mu, \mathcal{S}} \lesssim \|f\|_{L^{p_0}(\Omega; X)},
$$

$$
\|Tf\|_{L^{p_1, \infty} \mu, \mathcal{S}} \lesssim \|f\|_{L^{p_1}(\Omega; X)}
$$

for all $f \in L^{p_0}(\Omega; X) + L^{p_1}(\Omega; X)$. Then for all $p \in (p_0, p_1)$,

$$
\|Tf\|_{L^{p} \mu, \mathcal{S}} \lesssim \|f\|_{L^{p}(\Omega; X)} \quad \forall f \in L^{p}(\Omega; X).
$$

We omit the proof of this statement as it coincides precisely with the well-known proof of the Marcinkiewicz interpolation theorem, as can be found for example in [19]. This classical proof only uses properties of the distribution function $\lambda \mapsto \nu(\{x : |F(x)| > \lambda\})$ associated with a (classical) measure $\nu$, and these properties also hold for the outer distribution function $\lambda \mapsto \mu(\|F\|_\mathcal{S} > \lambda)$.

### 2.3 The time-frequency-scale space

We recall a number of notions from [3]. The time-frequency-scale space is $\mathbb{R}^3_+$. For $(\xi, x, s) \in \mathbb{R}^3_+$, we have mutually inverse local coordinate maps $\pi_{(\xi, x, s)}, \pi^{-1}_{(\xi, x, s)}$ mapping $\mathbb{R}^3_+$ to itself, defined by

$$
\pi_{(\xi, x, s)}(\theta, \zeta, \sigma) := (\xi + \theta(s\sigma)^{-1}, x + s\xi, s\sigma)
$$

$$
\pi^{-1}_{(\xi, x, s)}(\eta, y, t) := \left( t(\eta - \xi), \frac{y - x}{s}, t \right).
$$

**Definition 2.25.** Fix bounded intervals $\Theta$ and $\Theta_{in}$, with $0 \in \Theta_{in} \subset \Theta$. Define the model tree (over $\Theta$) by

$$
\mathcal{T}_\Theta := \left\{ (\theta, \zeta, \sigma) \in \mathbb{R}^3_+ : \theta \in \Theta, |\zeta| < 1 - \sigma \right\}
$$

and the tree with top $(\xi, x, s) \in \mathbb{R}^3_+$

$$
T_{(\xi, x, s)} := \pi_{(\xi, x, s)}(\mathcal{T}_\Theta)
$$

(we suppress $\Theta$ in the notation here, otherwise it gets in the way). For $T = T_{(\xi, x, s)}$, we write $\pi_T := \pi_{(\xi, x, s)}$ and $(\xi_T, x_T, s_T) = (\xi, x, s)$. The inner and outer parts of $T$ are

$$
T^{\text{in}} = \pi_{(\xi_T, x_T, s_T)}(\mathcal{T}_\Theta \cap \{\theta \in \Theta_{in}\}),
$$

$$
T^{\text{out}} = \pi_{(\xi_T, x_T, s_T)}(\mathcal{T}_\Theta \cap \{\theta \notin \Theta_{in}\}).
$$
and we denote the family of all trees by \( \mathbb{T}_\Theta \). We define a local measure \( \mu \) on the \( \sigma \)-generating collection \( \mathbb{T}_\Theta \) by

\[
\mu(T) := s_T.
\]

See Figure 1 for a sketch of the model tree \( \mathfrak{T}_\Theta \), and Figure 2 for how two trees look in local coordinates with respect to one of them.
Associated to each tree, we have a pullback map acting on functions (actually associated with the top of the tree) defined below.

**Definition 2.26.** Let $X$ be a Banach space and $F \in \mathcal{B}(\mathbb{R}^3_+; X)$. For each $T \in \mathcal{T}_\Theta$, define the function $\pi^*_T F \in \mathcal{B}(\mathbb{R}^3_+; X)$ by

$$(\pi^*_T F)(\theta, \zeta, \sigma) := 1_\mathcal{T}(\theta, \zeta, \sigma) e^{-2\pi i \xi_T (x_T + y_T \zeta)} F \circ \pi_T(\theta, \zeta, \sigma).$$

We also have strips, which contain no frequency information. These correspond to tents in $\mathbb{R}^2_+$.

**Definition 2.27.** Define the **model strip** by

$$\mathcal{D} := \{ (\zeta, \sigma) \in \mathbb{R}^2_+ : |\zeta| < 1 - \sigma \},$$

and define the **strip with top** $(x, s) \in \mathbb{R}^2_+$ to be the set

$$D_{(x,s)} := \pi_{(0,x,s)}(\mathbb{R} \times \mathcal{D}).$$

We let $(x_D, s_D) := (x, s)$, and we denote the family of all strips by $\mathcal{D}$. We define a local measure $\nu$ on the $\sigma$-generating collection $\mathcal{D}$ by

$$\nu(D_{(x,s)}) := s.$$

Note that we have the expression

$$D_{(x,s)} = \{ (\eta, y, t) \in \mathbb{R}^3_+ : |y - x| < s - t \} = \bigcup_{\xi \in \mathbb{Q}} \pi_{(\xi,x,s)}(\mathcal{T}_\Theta)$$

for any interval $\Theta$ containing the origin, so each strip can be written as a countable union of trees in $\mathcal{T}_\Theta$.

Now we define various local sizes, allowing us to measure Banach-valued functions on $\mathbb{R}^3_+$. In the context of Theorem 1.1, this Banach space will be either $X$ or $X^*$. First we consider the following Banach-valued ‘Lebesgue’ local sizes on $\mathcal{T}_\Theta$.

**Definition 2.28.** Let $\Theta$ and $\Theta_{in}$ be intervals with $0 \in \Theta_{in} \subset \Theta$, and let $Y$ be a Banach space. For $p \in [1, \infty]$, we define the local sizes $\| F \|_{\Theta, \Theta_{in}, \Theta_{out}}^p$ as follows: for $F \in \mathcal{B}(\mathbb{R}^3_+; Y)$ and $T \in \mathcal{T}_\Theta$,

$$\| F \|_{\Theta_{in}}^p(T) := \| \pi^*_T F \|_{L^p_{\zeta \theta d\zeta d\sigma}}(\mathcal{T}_\Theta, Y)$$

$$= \left( \int_{\mathcal{T}_\Theta} \| \pi^*_T F(\theta, \zeta, \sigma) \|_Y^p d\theta d\zeta \frac{d\sigma}{\sigma} \right)^{\frac{1}{p}}$$

with the usual modification when $p = \infty$. These local sizes have ‘inner’ and ‘outer’ variants given by localisation to the inner and outer parts of trees:

$$\| F \|_{\Theta_{in}}^p(T) := \| 1_{\Theta_{in}} F \|_{\Theta_{in}}^p(T) \quad \| F \|_{\Theta_{out}}^p(T) := \| 1_{\Theta_{out}} F \|_{\Theta_{out}}^p(T).$$
We do not mention the target Banach space in the notation, as it is unambiguous from context. We also omit mention of $\Theta_{in}$, otherwise the notation gets ridiculous. Generally multiple instances of $\mathcal{B}$ can occur on the same line with respect to different Banach spaces, as in the case in the following local size-Hölder inequality.

**Proposition 2.29.** Let $\Theta$ and $\Theta_{in}$ be intervals with $0 \in \Theta_{in} \subset \Theta$, and suppose $p, q \in [1, \infty]$. Let $Y$ be a Banach space, $F \in \mathcal{B}(\mathbb{R}^3_+; Y)$, and $G \in \mathcal{B}(\mathbb{R}^3_+; Y^*)$. Then for any $T \in \mathbb{T}_\Theta$,

$$\|\langle F; G \rangle \|_{\mathcal{B}_\Theta(T)} \lesssim \|F\|_{\mathcal{B}(\mathbb{R}^3_+; Y)}^{\frac{p}{q'}} \|G\|_{\mathcal{B}(\mathbb{R}^3_+; Y^*)}^{\frac{q}{q'}}.$$

Next we define the ‘randomised’ local sizes, using $\gamma$-norms (defined in Section 2.1.3) to represent square functions.

**Definition 2.30.** Let $\Theta$ and $\Theta_{in}$ be intervals with $0 \in \Theta_{in} \subset \Theta$, and let $Y$ be a Banach space. We define the local size $\mathcal{S}_\Theta$ for $F \in \mathcal{B}(\mathbb{R}^3_+; Y)$ and $T \in \mathbb{T}_\Theta$ by

$$\|F\|_{\mathcal{S}_\Theta(T)} := \left( \int_{\mathbb{R}^2} \|\pi_T^\Theta(1_{T_{out}}F)(\vartheta, \zeta, \sigma)\|_{\mathcal{B}(\mathbb{R}^3_+; Y)}^2 \, d\zeta \, d\vartheta \right)^{\frac{1}{2}}.$$

Note that $\mathcal{S}_\Theta$ depends on $\Theta$ and $\Theta_{in}$ through the definition of $T_{out}$. We combine the randomised and Lebesgue local sizes into two ‘full local sizes’ as follows.

**Definition 2.31.** Let $\Theta$ and $\Theta_{in}$ be intervals with $0 \in \Theta_{in} \subset \Theta$, and let $Y$ be a Banach space. For $F \in \mathcal{B}(\mathbb{R}^3_+; Y)$ and $T \in \mathbb{T}_\Theta$, we define

$$\|F\|_{\mathcal{B}_\Theta(T)} := \|F\|_{\mathcal{S}_\Theta(T)} + \|F\|_{\mathcal{B}_\Theta(T)}^{\frac{1}{2}}$$

and

$$\|F\|_{\mathcal{B}_\Theta(T)} := \|F\|_{\mathcal{S}_\Theta(T)} + \|F\|_{\mathcal{B}_\Theta(T)}^{\frac{1}{2}}.$$

When applied to the wave packet embedding $F = \text{Emb}_{\varphi}[f]$ of a Banach-valued function $f : \mathbb{R} \to Y$ (see Definition 3.1), the quantity $\|F\|_{\mathcal{S}_\Theta(T)}$ represents a Littlewood–Paley square function of a modulation of $f$, as in Proposition 2.10. On the other hand, $\|F\|_{\mathcal{B}_\Theta(T)}$ represents a Hardy–Littlewood-type maximal function of $f$, and $\|F\|_{\mathcal{B}_\Theta(T)}$ arises by duality (as seen in the proof of Proposition 2.32).

The following local size-Hölder inequality follows simply from the definition of the local sizes, the classical Hölder inequality (for the Lebesgue parts), and the $\gamma$-Hölder inequality (for the randomised parts).

**Proposition 2.32.** Let $\Theta$ and $\Theta_{in}$ be intervals with $0 \in \Theta_{in} \subset \Theta$, and let $Y$ be a Banach space. Let $F \in \mathcal{B}(\mathbb{R}^3_+; Y)$ and $G \in \mathcal{B}(\mathbb{R}^3_+; Y^*)$. Then for all $T \in \mathbb{T}_\Theta$ and $A \in \mathbb{T}_{\Theta_{in}}$,

$$\|1_{\mathbb{R}^3_+ \setminus A} \langle F; G \rangle \|_{\mathcal{B}_\Theta(T)} \lesssim \|F\|_{\mathcal{B}_\Theta(T)} \|G\|_{\mathcal{B}_\Theta(T)}.$$ (2.8)
Proof. For $T \in \mathbb{T}_\Theta$, expand the definition of the local size $\Omega^1_{\Theta}(T)$ to obtain
\[
\|1_{\mathbb{R}^3_+ \setminus A} \langle F; G \rangle \|_{\Omega^1_{\Theta}(T)} = \int_{\mathbb{T}_\Theta} |\pi_T^*(1_{\mathbb{R}^3_+ \setminus A} F; 1_{\mathbb{R}^3_+ \setminus A} G)(\theta, \zeta, \sigma)| \, d\theta \, d\zeta \, \frac{d\sigma}{\sigma}
\]
\[
= \left( \int_{\mathbb{T}_\Theta \cap \{\theta \in \Theta \}} + \int_{\mathbb{T}_\Theta \cap \{\theta \notin \Theta \}} \right) |\pi_T^*(1_{\mathbb{R}^3_+ \setminus A} F; 1_{\mathbb{R}^3_+ \setminus A} G)(\theta, \zeta, \sigma)| \, d\theta \, d\zeta \, \frac{d\sigma}{\sigma}.
\]
The first summand is controlled by
\[
\|\pi_T^*(1_{\mathbb{R}^3_+ \setminus A} F)\|_{L^\infty(\mathbb{T}_\Theta; Y)} \|\pi_T^*(1_{\mathbb{R}^3_+ \setminus A} G)\|_{L^1(\mathbb{T}_\Theta \times B_1 \times (0,1); Y^*)}
\]
using Hölder’s inequality, and the second summand by
\[
\|\pi_T^*(1_{\mathbb{R}^3_+ \setminus A} F)\|_{L^2(\Theta \setminus \Theta_{in} \times B_1 \times (\gamma, \infty))} \|\pi_T^*(1_{\mathbb{R}^3_+ \setminus A} G)\|_{L^2(\Theta \setminus \Theta_{in} \times B_1 \times (\gamma, \infty))}
\]
by Cauchy–Schwarz and first part of Proposition 2.8. This shows the bound (2.8).

By combining Proposition 2.32 with Proposition 2.22, we obtain a Hölder-type inequality involving the classical integral and iterated outer Lebesgue quasi-norms. See [2, Corollary 4.13] for full details of this argument in the Walsh setting.

Corollary 2.33. Let $\Theta$ and $\Theta_{in}$ be intervals with $0 \in \Theta_{in} \subset \Theta$, and let $Y$ be a UMD Banach space. Let $p, q \in (0, \infty]$. Then for all $F \in \mathcal{B}(\mathbb{R}^3_+; Y)$ and $G \in \mathcal{B}(\mathbb{R}^3_+; Y^*)$,
\[
\int_{\mathbb{R}_+^3} |\langle F(\eta, y, t); G(\eta, y, t) \rangle| \, d\eta \, dy \, dt \lesssim \|F\|_{L_p^p L_q^q \mathcal{G}_\Theta} \|G\|_{L_p^q L_q^p \mathcal{G}_\Theta^*}.
\]

Proof. We apply Proposition 2.22 to the integral to obtain
\[
\int_{\mathbb{R}_+^3} |\langle F(\eta, y, t); G(\eta, y, t) \rangle| \, d\eta \, dy \, dt \lesssim \|\langle F(\eta, y, t); G(\eta, y, t) \rangle\|_{L_p^p L_q^q \mathcal{G}_\Theta} \|G\|_{L_p^q L_q^p \mathcal{G}_\Theta^*}.
\]
By the outer Hölder inequality (Proposition 2.23), the term on the right-hand side satisfies the required bounds
\[
\|\langle F(\eta, y, t); G(\eta, y, t) \rangle\|_{L_p^p L_q^q \mathcal{G}_\Theta} \lesssim \|F\|_{L_p^p L_q^q \mathcal{G}_\Theta} \|G\|_{L_p^q L_q^p \mathcal{G}_\Theta^*},
\]
as long as the sizes, which are themselves localised outer Lebesgue quasi-norms, satisfy the bounds
\[
\|\langle F; G \rangle\|_{L_p^p \mathcal{G}_\Theta} \lesssim \|F\|_{L_p^p \mathcal{G}_\Theta} \|G\|_{L_p^q \mathcal{G}_\Theta^*}.
\]
Recalling the definitions of the iterated local sizes (2.4), the condition to check becomes
\[ \\|1_D \langle F; G \rangle \|_{L^1_\mu \mathcal{O}} \lesssim \|1_D F\|_{L^q_\mu \mathcal{O}} \|1_D G\|_{L^{q'}_\mu \mathcal{O}} \]
for any \( D \in \mathcal{D} \). To show this inequality, we appeal once again to the outer Hölder inequality (Proposition 2.23), which yields the above bound as long as the inequality
\[ \|\langle F; G \rangle\|_{\mathcal{O} \Theta} \lesssim \|F\|_{\mathcal{O} \Theta} \|G\|_{\mathcal{O} \Theta} \]
for the sizes holds. But this is precisely the statement of Proposition 2.32, the ‘size-Hölder’ inequality, which represents the starting level of the iterated outer measure construction, and justifies the choice of specific form of the size \( \|\cdot\|_{\mathcal{O}^* \Theta} \) that is ‘dual’ in an appropriate sense with respect to the size \( \|\cdot\|_{\mathcal{O} \Theta} \). The choice of the size \( \|\cdot\|_{\mathcal{O} \Theta} \) is, on the other hand, dictated by the necessity to satisfy the conditions of Proposition 2.22 at the ‘last’ level of outer measure iteration.

2.4 (Truncated) wave packets and \( \mathcal{V}_\epsilon \)

For each \((\eta, y, t) \in \mathbb{R}^3_+\), recall the operator
\[ \Lambda(\eta, y, t) := \text{Tr}_y \text{Mod}_\eta \text{Dil}_t. \]
Given \( \varphi \in \mathcal{S}(\mathbb{R}) \), the function \( \Lambda(\eta, y, t) \varphi \) is called a wave packet at \((y, \eta, t)\) modelled on \( \varphi \).

In [36], the second author introduced the concept of a family of left (or right) truncated wave packets, defined below. These are crucial to the representation of the operator \( \mathcal{V}_\epsilon \) on the time-frequency-scale space \( \mathbb{R}^3_+ \).

Families of left- or right-truncated wave packets \( \Psi^\pm \) are constructed depending on parameters \((b, \epsilon)\). The point of the defining properties below is to guarantee, for \( \epsilon > 0 \) sufficiently small with respect to \( b \), that the Fourier support of \( \text{Mod}_\eta \text{Dil}_t \Psi_{\eta,t}(\cdot, \cdot) \) is contained in \([c_-, c_+]\), and that this Fourier support is a Whitney interval with respect to the set \((c_-, c_+)\). This is encoded more precisely in Proposition 2.36 and the following corollary. The Whitney condition is not required for these two results but is of fundamental importance in proving Theorem 3.3.

The definition we use is a slight modification of that in [36], making the concept a bit clearer and easier to work with. For \( b > 0 \), we let \( \Phi_b \subset \mathcal{S}(\mathbb{R}) \) denote the subspace of Schwartz functions with Fourier transform supported in the ball \( B_b \).

**Definition 2.34.** Let \( b, \epsilon > 0 \). A family of left-truncated wave packets with parameters \((b, \epsilon)\) is a smooth map
\[ \Psi^+ : \mathbb{R}^2_+ \times \{(c_-, c_+) \in (\mathbb{R} \cup \{\infty\})^2 : c_- < c_+\} \to \Phi_b, \]
which we write as \( \Psi^+_{\eta,t}(c_-, c_+) \) or \( \Psi^+((\eta, t), (c_-, c_+)) \), satisfying the following properties:
Smoothness For any fixed $N \in \mathbb{N}$ and $\alpha_+, \alpha_-, \alpha_\eta, \alpha_t \in \{0, \ldots, N\}$, the functions

$$(t^{-1} \partial_{c_-})^{\alpha_+} (t^{-1} \partial_{c_+})^{\alpha_-} (t^{-1} \partial_{\eta})^{\alpha_\eta} (t \partial_t)^{\alpha_t} \Psi_{(\eta, t)}^{(c_-, c_+), +} \in \Phi_0$$

are uniformly bounded in $\Phi_0$.

Frequency dependence The function $\Psi_{(\eta, t)}^{(c_-, c_+), +}$ is non-vanishing only if

$$\eta \in \left( c_- + (1 - \varepsilon)t^{-1}, \min(c_+ - (1 - \varepsilon)t^{-1}, c_- + (1 + \varepsilon)t^{-1}) \right);$$

Weak dependence on right endpoint For all $(\eta, t) \in \mathbb{R}_2^+$, the map

$$c_+ \mapsto \Psi_{(\eta, t)}^{(c_-, c_+), +}$$

is constant for $c_+ > \eta + 3t^{-1}$;

Frequency-scale relation for $(0, \infty)$ For all $(\eta, t) \in \mathbb{R}_2^+$,

$$\Psi_{(\eta, t)}^{(0, \infty), +} = \Psi_{(\eta, 1)}^{(0, \infty), +}.$$ 

A family of right-truncated wave packets with parameters $(\mathfrak{b}, \varepsilon)$ is a map

$$\Psi^+: \mathbb{R}_2^+ \times \{(c_-, c_+) \in \mathbb{R} \cup \{-\infty\}^2 : c_- < c_+ \} \rightarrow \Phi_0(\mathbb{R})$$

such that the function $(\eta, t, c_-, c_+) \mapsto \Psi_{(\eta, t)}^{(-c_-, c_-), -}$ is a family of left-truncated wave packets with parameters $(\mathfrak{b}, \varepsilon)$.

Remark 2.35. Consider a smooth map

$$\Psi_{(0, 1), +} : \mathbb{R}_2^+ \rightarrow \Phi_0$$

satisfying the conditions above with $(c_-, c_+) = (0, 1)$ fixed. Then families of left- and right-truncated wave packets $\Psi^{\pm}$ can be defined by rescaling:

$$\Psi_{(\eta, t)}^{(c_-, c_+), \pm} = \Psi_{(\eta, t)}^{(0, 1), \pm} \left( \frac{c_-}{c_+ - c_-}, \frac{t(c_+ - c_-)}{c_+ - c_-} \right).$$

When $c_+ = \infty$, this has to be interpreted as a limit in $c_+ \rightarrow \infty$.

The representation of $\mathcal{V}_\mathfrak{c}$ on $\mathbb{R}_3^+$ follows from representing the Fourier projection onto an interval as a superposition of (modulated and dilated) truncated wave packets. A version of the following proposition was already shown in [36, Lemma 3.1]; here we give a somewhat clearer proof.

Proposition 2.36. For any sufficiently small $\mathfrak{b} > 0$ and any $\varepsilon > 0$ sufficiently small with respect to $\mathfrak{b}$, there exist families $\Psi^{\pm}$ of left- and right-truncated wave packets with parameters $(\mathfrak{b}, \varepsilon)$ such that for any $c_- < c_+$ it holds that
\[
1_{(c_-, c_+)}(\xi) = \sum_{\square \in \{+,-\}} \int_{\mathbb{R}^2} \left( \Lambda_{(\eta,0,t)}^{(c_-,c_+)}(\square)^{\wedge} \right)(\xi) \, d\eta \, dt.
\]

**Proof.** By Remark 2.35, and since
\[
1_{(c_-, c_+)}(\xi) = 1_{(0,1)} \left( \frac{\xi - c_-}{c_+ - c_-} \right),
\]
it is sufficient to prove the result for \((c_-, c_+) = (0, 1)\). Let \(\chi \in C^\infty_c(B_x)\) be non-negative, even, and positive on \(B_{\varepsilon/2}\), set
\[
\chi^+(z) := \|\chi\|^{-1}_{L^1} \int_{-\infty}^{z} \chi(\zeta) d\zeta, \quad \chi^-(z) := \chi^+(z),
\]
and let \(\hat{\varphi} \in C^\infty_c(B_{b/2})\) be non-negative, even, and positive on \(B_{b/4}\).

Consider the expression
\[
m^+(\xi) := \int_{\mathbb{R}} \int_{0}^{\infty} \hat{\varphi}(t(\xi - \eta)) \chi(t\eta - 1) \chi^-(t(\eta - 1) + 1) \, d\eta \, dt. \tag{2.9}
\]

We claim that \(m^+(\xi)\) is non-negative, smooth on \((0, 1)\), supported in \([0, 3/4]\), non-vanishing in an open neighborhood of \((0, 1/2)\), and that \(\partial_\xi m^+(\xi)\) is supported in \([1/4, 3/4]\).

The first claim follows by noticing that \(m^+(\xi)\) is an integral of non-negative functions. The integrand in (2.9) vanishes unless \(t\eta \in B_{\varepsilon}(1), t > 2 - 2\varepsilon,\) and \(t\xi \in B_{b/2}(t\eta)\); thus the integral vanishes unless
\[
\xi \in \bigcup_{t=2-2\varepsilon} B_{t^{-1}(\frac{b}{2}+\varepsilon)}(t^{-1}) = \bigcup_{t=2-2\varepsilon} \left( \frac{1 - \varepsilon - b/2}{t}, \frac{1 + \varepsilon + b/2}{t} \right) \subset (0, 3/4).
\]

On the other hand, fix \(t\eta = 1\) and note that for \(t > 2 + \varepsilon\) the integrand is non-vanishing for \(t\xi \in B_{b/4}(1)\). Since the integrand depends continuously on the parameters \(\eta, t\), we can see that
\[
m^+(\xi) > 0 \text{ on } \bigcup_{t=2+\varepsilon} B_{b/4t}(1/t) \supset (0, 1/2]
\]
as long as \(\varepsilon > 0\) is sufficiently small with respect to \(b\). Furthermore, for any \(\xi > 0\), we have that
\[
\xi \partial_\xi m^+(\xi) = \int_{\mathbb{R}} \int_{0}^{\infty} \xi \partial_\xi \hat{\varphi}(t\xi - t\eta) \chi(t\eta - 1) \chi^-(t(\eta - 1) + 1) \, d\eta \, dt
\]
\[
= \int_{\mathbb{R}} \int_{0}^{\infty} (t\partial_t - \eta \partial_\eta) \hat{\varphi}(t\xi - t\eta) \chi(t\eta - 1) \chi^-(t(\eta - 1) + 1) \, d\eta \, dt
\]
\[
= -t \|\chi\|_{L^1} \int_{0}^{\infty} \hat{\varphi}(t\xi - t\eta) \chi(t\eta - 1) \chi(t(\eta - 1) + 1) \, d\eta \, dt.
\]
The integrand is non-vanishing only if \( t \eta \in B_\varepsilon(1), t \in B_{2\varepsilon}(2) \), and \( t \xi \in B_{\varepsilon+b/2}(1) \) and in particular only if \( \xi \in B_{b/2}(1/2) \) (again, provided that \( \varepsilon \) is small enough with respect to \( b \)). We also deduce that \( m^+(\xi) \) is smooth on \((0, 1)\) since its derivative is given, by the discussion above, as an integral of smooth functions over a compact set of parameters: \( t \eta \in B_\varepsilon(1), t \in B_{2\varepsilon}(2) \). Now set

\[
m^-(\xi) = m^+(1 - \xi),
\]

\[
m(\xi) = \begin{cases} 
m^-(\xi) + m^+(\xi) & \xi \in (0, 1), \\
\lim_{\xi \to 0^+} m^-(\xi) & \xi < 0, \\
\lim_{\xi \to 1^-} m^-(\xi) & \xi > 1.
\end{cases}
\]

From the discussion above, it follows that \( m(\xi) \) is positive and smooth, \( \partial_\xi m(\xi) \) is supported in \( B_{b/2}(1/2) \), the identity \( m(1 - \xi) = m(\xi) \) holds, and

\[
1_{(0,1)}(\xi) = \int_{\mathbb{R}^2_+} \hat{\phi}(t(\xi - \eta)) m^{-1}(\xi) \left( \chi(t\eta - 1)\chi^-(t(\eta - 1) + 1) + \chi^+(t\eta - 1)\chi(t(1 - \eta) + 1) \right) \mathrm{d}\eta \mathrm{d}t.
\]

Note that the integrand is non-vanishing only when \( \eta \in B_1 \) and \( t > 2 - 2\varepsilon \), and for this range of parameters one has that \( \hat{\phi}(\xi) m^{-1}(\xi/t + \eta) \) is contained in \( \in C_\infty^c(B_b) \) and uniformly bounded in any Schwartz seminorm. Now define

\[
\psi^{(0,1),+}_{(\eta, t)}(x) := \chi(t\eta - 1)\chi^-(t(\eta - 1) + 1) \left( \hat{\phi}(\cdot) m^{-1}(\cdot/t + \eta) \right)^\vee(x)
\]

\[
\psi^{(0,1),-}_{(\eta, t)}(x) := \chi^+(t\eta - 1)\chi(t(\eta - 1) + 1) \left( \hat{\phi}(\cdot) m^{-1}(\cdot/t + \eta) \right)^\vee(x).
\]

The function \( m^{-1} \) is constant on the support of \( \xi \to \hat{\phi}(t(\xi - \eta)) \) when \( t \eta < \frac{1-b}{2} t - \frac{b}{2} \) and in particular when

\[
\frac{1-b}{2} t(\eta - 1) < -\frac{1-b}{2} t\eta - \frac{b}{2}
\]

or even when

\[
t(\eta - 1) < -(1 + 4b) < -\frac{1+2b}{1-b}(1 + \varepsilon)
\]

as long as \( \varepsilon > 0 \) is small enough. Thus

\[
\psi^{(0,1),+}_{(\eta, t)}(x) = \chi(t\eta - 1)\chi^-(t(\eta - 1) + 1) m^{-1}(0)\varphi(x)
\]

\[
\psi^{(0,1),-}_{(\eta, t)}(x) = \chi^+(t\eta - 1)\chi(t(\eta - 1) + 1) m^{-1}(0)\varphi(x)
\]

for \( t(\eta - 1) < -(1 + 4b) \). It follows that \( \psi^{(0,1),\pm} \) satisfies the left- and right-truncated wave packet conditions.
Now we can prove the claimed representation of $\mathcal{V}_c$.

**Corollary 2.37.** For $b > 0$ sufficiently small, and $\varepsilon > 0$ sufficiently small with respect to $b$, there exists $\varphi \in \Phi_b$ and families of left- and right-truncated wave packets $\Psi^\pm$ with parameters $(b/2, \varepsilon)$ such that the wave packet representation

$$
\mathcal{V}_{c,j} f(x) = \sum_{\square \in \{+, -\}} \int_{\mathbb{R}^3} \langle f; \Lambda(\eta, y, t) \varphi \rangle \Lambda(\eta, y, t) \Psi(\gamma_j(x), \gamma_{j+1}(x)), \square (x) \, d\eta \, dy \, dt
$$

holds for any measurable $c : \mathbb{R} \rightarrow \Delta$ and $f \in \mathcal{S}(\mathbb{R})$. We denote the summands on the right-hand side by $\mathcal{V}^\pm_{c,j} f(x)$.

**Proof.** Let $\varphi \in \Phi_b$ be even and real-valued with $\hat{\varphi} = 1$ on $B_b/2$, and let $\Psi^\pm$ be families of truncated wave packets with parameters $(b/2, \varepsilon)$ as in Proposition 2.36. By Fourier support considerations, we have

$$
\Lambda(\eta, 0, t) \Psi(c_-, c_+, \pm, \eta, t) = \Lambda(\eta, 0, t) \varphi * \Lambda(\eta, 0, t) \Psi(c_-, c_+, \pm, \eta, t)
$$

for all frequencies $c_- < c_+$. Thus for all $x \in \mathbb{R}$

$$
\mathcal{V}_{c,j} f(x) = \int_{\mathbb{R}} f(x) e^{2\pi i \xi x} \, d\xi
$$

as claimed. □

## 3 REDUCTION OF THE MAIN THEOREM

### 3.1 Reduction to wave packet embedding bounds

Now we can begin to approach Theorem 1.2: for an $r_0$-intermediate UMD space $X$, we want to show

$$
\|\mathcal{V}_c f\|_{L^p(\mathbb{R}; L^p(\mathbb{N}; X))} \lesssim_{p, r, X} \|f\|_{L^p(\mathbb{R}; X)} \quad \forall f \in \mathcal{S}(\mathbb{R}; X)
$$
for all $r_0 < r < \infty$ and $(r/(r_0 - 1))' < p < \infty$, uniformly in $\epsilon : \mathbb{R} \to \Delta$. By duality, this is equivalent to showing
\[ \left| \int_{\mathbb{R}} \langle \mathcal{Y}_\epsilon f(x); g(x) \rangle \, dx \right| \lesssim \| f \|_{L^p(\mathbb{R}; X)} \| g \|_{L^{p'}(\mathbb{R}; (c_{00}(\mathbb{N}); \mathcal{S}(\mathbb{R}; X^*))} \quad (3.1) \]
for all $f \in \mathcal{S}(\mathbb{R}; X)$ and $g \in c_{00}(\mathbb{N}; \mathcal{S}(\mathbb{R}; X^*))$ (the set of finitely supported sequences of $X^*$-valued Schwartz functions). The duality pairing in (3.1) is given in terms of that between $X$ and $X^*$ by
\[ \langle \mathcal{Y}_\epsilon f(x); g(x) \rangle = \sum_{j \in \mathbb{N}} \langle \mathcal{Y}_\epsilon, j f(x); g_j(x) \rangle. \]

Due to the truncated wave packet representation of $\mathcal{Y}_\epsilon$ in Corollary 2.37, the left-hand side of (3.1) can be rephrased in terms of the following wave packet embeddings.

**Definition 3.1.** Let $Y$ be a Banach space and $\varphi \in \mathcal{S}(\mathbb{R})$. For all $f \in \mathcal{S}(\mathbb{R}; Y)$, we define the wave packet embedding $\text{Emb}_\varphi[f] : \mathbb{R}^3_+ \to Y$ of $f$ with respect to $\varphi$ by
\[ \text{Emb}_\varphi[f](\eta, y, t) := \langle f; \Lambda(\eta, y, t) \varphi \rangle = \int_{\mathbb{R}} f(x) t^{-1} e^{-2\pi i \eta (x-y)} \varphi(\frac{x-y}{t}) \, dx. \]

Given families $\Psi^\pm$ of left- and right-truncated wave packets and a measurable function $\epsilon : \mathbb{R} \to \Delta$, for all sequences $g = (g_j)_{j \in \mathbb{N}} \in c_{00}(\mathbb{N}; \mathcal{S}(\mathbb{R}; Y))$, we define the truncated wave packet embeddings with respect to $\epsilon$, $A_\epsilon^\pm[g] : \mathbb{R}_+^3 \to Y$, by
\[ A_\epsilon^\pm[g](\eta, y, t) := \int_{\mathbb{R}} \sum_{j \in \mathbb{N}} g_j(x) \Lambda(\eta, y, t) \Psi^\pm_{(\eta, t)}(\tau_j(x), \tau_{j+1}(x)) \pm(x) \, dx. \]

In [3], we proved the following bounds for the wave packet embedding; in Section 6, we provide a detailed sketch of the proof. This is the only place in which the $r_0$-intermediacy assumption is used, and it is not known if this condition is necessary.

**Theorem 3.2.** Let $\Theta$ and $\Theta_{in}$ be bounded intervals with $0 \in \Theta_{in} \subset \Theta$, and suppose $\varphi \in \mathcal{S}(\mathbb{R})$ has Fourier support in $\Theta_{in}$. Let $Y$ be an $r_0$-intermediate UMD space for some $r_0 \in [2, \infty)$. Then for all $p \in (1, \infty)$ and $q \in \min(p, r_0)'(r_0 - 1, \infty)$,
\[ \| \text{Emb}_\varphi[f] \|_{L^p(\mathbb{R}; \mathcal{S}(\mathbb{R}; Y))} \| f \|_{L^q(\mathbb{R}; X)} \lesssim \| f \|_{L^p(\mathbb{R}; X)}. \quad (3.2) \]

The main technical goal of this article is to prove bounds for the truncated wave packet embeddings.

**Theorem 3.3.** Let $Y$ be a UMD Banach space with type $r_0'$ for some $r_0 \in [2, \infty)$. For sufficiently small $b, \epsilon > 0$ and for each choice of sign $\pm$, the following holds. Consider the embedding $A_\epsilon^\pm$ defined with respect to a choice of left/right truncated wave packets $\Psi^\pm$ with parameters $(b, \epsilon)$. Let $\Theta$ be a bounded interval containing $\pm[0, 1+\epsilon)$, and let $\Theta_{in} = \Theta \cap \pm(-\infty, 1-\epsilon)$. Then for all $r' \in (1, r_0'), p' \in (1, \infty]$, and $q' \in (r', \infty]$, and all measurable $\epsilon : \mathbb{R} \to \Delta,$
\[\|A_\xi \pm [g]\|_{L^q_x L^p_t \ell^q_y \otimes_\Theta} \lesssim \|g\|_{L^p_x (\mathbb{R}; \ell^r_t (\mathbb{N}; Y))}.\]

Note that the conditions on the intervals \((\Theta, \Theta_{in})\) are different depending on the choice of sign \(\pm\). These technical conditions are not necessary (it is also possible to prove the result for both embeddings \(A^\pm\) with arbitrary \(\Theta_{in} \subset \Theta\) such that \(B_{2b} \subset \Theta_{in} \subset \Theta\), and these conditions are somewhat more natural) but they simplify the proofs considerably.

We now prove Theorem 1.2 as a consequence of Theorem 3.3. Recall that Theorem 1.2 is an equivalent linearised version of our goal, Theorem 1.1.

**Proof of Theorem 1.2, assuming Theorem 3.3.** We prove the dual bound (3.1) for all \(f \in \mathcal{S}(\mathbb{R}; X)\) and all \(g \in c_0 (\mathbb{N}; \mathcal{S}(\mathbb{R}; X^*))\). By the truncated wave packet representation, Corollary 2.37, for all sufficiently small \(b, \varepsilon > 0\) there exist \(\varphi \in \Phi_b\) and families of left- and right-truncated wave packets \(\Psi^\pm\) with parameters \((b/2, \varepsilon)\) such that

\[
\int_{\mathbb{R}} \langle \mathcal{V}_\xi f(x); g(x) \rangle \, dx = \sum_{\square \in \{+, -\}} \int_{\mathbb{R}^3_+} \langle \text{Emb}_\varphi [f](\eta, y, t); A^\square_\xi [g](\eta, y, t) \rangle \, d\eta \, dy \, dt,
\]

and thus to prove (3.1), it suffices to prove

\[
\left| \int_{\mathbb{R}^3_+} \langle \text{Emb}_\varphi [f](\eta, y, t); A^\pm_\xi [g](\eta, y, t) \rangle \, d\eta \, dy \, dt \right| \lesssim \|f\|_{L^p_x (\mathbb{R}; X)} \|g\|_{L^q_x (\mathbb{R}; \ell^r_y (\mathbb{N}; Y^*))}. \tag{3.3}
\]

Fix a sign \(\pm\) and define \((\Theta, \Theta_{in})\) as in Theorem 3.3. By Corollary 2.33, for all \(q \in [1, \infty]\), we have

\[
\left| \int_{\mathbb{R}^3_+} \langle \text{Emb}_\varphi [f](\eta, y, t); A^\pm_\xi [g](\eta, y, t) \rangle \, d\eta \, dy \, dt \right| \lesssim \|\text{Emb}_\varphi [f]\|_{L^p_x L^q_y \ell^q_\Theta} \|A^\pm_\xi [g]\|_{L^p_x L^q_y \ell^q_\Theta}.
\]

If \(q\) satisfies the conditions

\[
q > \min(p, r_0)'(r_0 - 1) \quad \text{and} \quad q' > r', \tag{3.4}
\]

then provided \(b\) and \(\varepsilon\) were chosen sufficiently small, Theorems 3.2 and 3.3 imply that

\[
\|\text{Emb}_\varphi [f]\|_{L^p_x L^q_y \ell^q_\Theta} \|A^\pm_\xi [g]\|_{L^p_x L^q_y \ell^q_\Theta} \lesssim \|f\|_{L^p_x (\mathbb{R}; X)} \|g\|_{L^q_x (\mathbb{R}; \ell^r_y (\mathbb{N}; Y^*))},
\]

proving (3.3) and hence the theorem, so it suffices to show that there exists a \(q\) satisfying (3.4). This reduces to the inequality

\[
r > \min(p, r_0)'(r_0 - 1),
\]

which is equivalent to the assumed condition \(p > (r/(r_0 - 1))'\).

Thus it suffices to prove Theorem 3.3. In the next section, we reduce this to the domination of the truncated wave packet embeddings \(A^\pm_\xi\) by an auxiliary ‘scalar’ embedding.
3.2 Reduction to domination by a scalar embedding

**Definition 3.4.** Let $Y$ be a Banach space, $\Theta$ a bounded interval containing the origin, and $N \in \mathbb{N}$. For all $g \in c_00(\mathbb{N}; \mathcal{S}(\mathbb{R}; Y))$, all $r \in (1, \infty)$, and all measurable $c: \mathbb{R} \to \Delta$, define the auxiliary embedding $M_{r, \Theta}^{r, N}[g]: \mathbb{R}_+^3 \to [0, \infty)$ by

$$M_{r, \Theta}^{r, N}[g](\eta, y, t) := \int_{\mathbb{R}} \left( \sum_{j \in \mathbb{N}} \|g_j(x)\|_Y^{r'} (\Theta - t \epsilon_j(x)) \right)^{1/r'} \frac{1}{t} \left\langle \frac{x - y}{t} \right\rangle^{-N} dx.$$ 

Note that this embedding basically ignores the geometry of $Y$, as it only depends on $g$ through the sequence $\|g\|_Y := (\|g_j(x)\|_Y)_{j \in \mathbb{N}}$. The following bounds follow by applying [36, Proposition 4.1] to the scalar-valued sequence $\|g\|_Y$.

**Proposition 3.5.** Let $Y$ be a Banach space and $\Theta$ a bounded interval containing the origin. If $N \in \mathbb{N}$ is sufficiently large, then for any $r' \in [1, \infty)$, $p' \in (1, \infty)$, and $q' \in (r', \infty)$, and for any function $g \in c_00(\mathbb{N}; \mathcal{S}(\mathbb{R}; Y))$,

$$\|M_{r, \Theta}^{r, N}[g]\|_{L_{p'}(\mathbb{R}; \ell_{q'}^{r'}(\mathbb{N}; Y)))} \lesssim \|g\|_{L_{p'}(\mathbb{R}; \ell_{q'}^{r'}(\mathbb{N}; Y)))}$$

with implicit constant independent of $\Theta$ and $N$.

Thus to prove Theorem 3.3, it suffices to show

$$\|A_\epsilon^+[g]\|_{L_{p'}^q \ell_{q'}^{r'}(\mathbb{N}; \mathfrak{S}_\Theta)} \lesssim \|M_{r, \Theta}^{r, N}[g]\|_{L_{p'}^q \ell_{q'}^{r'}(\mathbb{N}; \mathfrak{S}_\Theta)} \quad \forall g \in c_00(\mathbb{N}; \mathcal{S}(\mathbb{R}; Y))$$

for sufficiently large $N \in \mathbb{N}$, under the hypotheses of the theorem. By definition of the outer Lebesgue quasi-norms, this will follow from the following theorem, which extends [36, Proposition 5.1] to the vector-valued setting.

**Theorem 3.6.** Let $Y$ be a UMD Banach space with type $r_0'$ for some $r_0 \in [2, \infty)$. For sufficiently small $b, \varepsilon > 0$ and for each choice of sign $\pm$, the following holds. Consider the embedding $A_\epsilon^\pm$ defined with respect to a choice of left/right truncated wave packets $\Psi^\pm$ with parameters $(b, \varepsilon)$. Let $\Theta$ be a bounded interval containing $\pm [0, 1 + \varepsilon)$, and let $\Theta_{in} = \Theta \cap \pm (-\infty, 1 - \varepsilon)$. Let $E \in \mathcal{T}_\Theta$ be a union of trees. Then for all measurable $c: \mathbb{R} \to \Delta$, all $r \in (1, r_0')$, and all sufficiently large $N \in \mathbb{N}$,

$$\|1_{\mathbb{R}_+^3 \setminus E} A_\epsilon^\pm[g]\|_{\mathcal{S}_\Theta} \lesssim \|1_{\mathbb{R}_+^3 \setminus E} M_{r, \Theta}^{r, N}[g]\|_{\mathcal{S}_\Theta} \quad \forall g \in c_00(\mathbb{N}; \mathcal{S}(\mathbb{R}; Y))$$

(3.5)

with implicit constant independent of $c$.

Thus the only task left — the hardest task — is to prove Theorem 3.6.

4 SCALAR DOMINATION OF THE TRUNCATED WAVE PACKET EMBEDDING

We devote the entirety of this section to the proof of Theorem 3.6. From now on, we only consider the embedding $A_\epsilon^+$, defined in terms of a fixed family of left-truncated wave packets $\Psi^+$ with
sufficiently small parameters \((b, \varepsilon)\); the proof for \(A^{-}_{\tau}\) is symmetric. We fix intervals \(\Theta\) and \(\Theta_{in}\) as in the statement of the theorem: that is, \(\Theta\) is a bounded interval containing \([0, 1 + \varepsilon)\), and \(\Theta_{in} = \Theta \cap (-\infty, 1 - \varepsilon)\). To ease notation, we fix \(r, \varepsilon, \) and \(N\), and abbreviate \(A^{+}_{\tau}, M_{\tau, \Theta}^{r, N}, \) and \(\Psi^{+}\) by \(A, M, \) and \(\Psi\).

We need the following useful lemma, which follows from [36, Lemma 5.5].

**Lemma 4.1.** Let \(\tau : \mathbb{R} \to \mathbb{R}_{+}\) be a 1-Lipschitz function. Then for any non-negative function \(h : \mathbb{R} \to \mathbb{R}_{+}\) and any \(L > 1\), one has

\[
\int_{\mathbb{R}} h(z) \, dz \approx_{L} \int_{\mathbb{R}} \int_{B(\tau(x), L(x))} h(z) \, dz \, dx
\]

with implicit constant depending only on \(L\).

We will be considering sets \(T \setminus E\), where \(T\) is a tree and \(E\) is a union of trees. The boundary of \(E\), viewed in local coordinates with respect to \(T\), is encoded as a function \(\tau_{E}\), and the following lemma contains the requisite properties of this function.

**Lemma 4.2.** Let \(E \in \mathcal{T}^U\) and fix \(T \in \mathcal{T}\). The function

\[
\tau_{E}(\theta, \zeta) := \sup (\{0\} \cup \{\sigma : (\theta, \zeta, \sigma) \in \pi^{-1}_{T}(E)\})
\]

defines the boundary of \(\pi^{-1}_{T}(E)\) as a graph over \((\theta, \zeta) \in \mathbb{R}^{2}\). For every \(\theta \in \mathbb{R}\), the map \(\zeta \to \tau_{E}(\theta, \zeta)\) is non-negative and 1-Lipschitz. If \(\{(\theta) \times \mathbb{R}^{2}\} \cap \pi^{-1}_{T}(T \setminus E)\) is non-empty, then \(\tau_{E}(\theta, \zeta) \leq 3 + |\zeta|\).

Furthermore, there exists \(E' \in \mathcal{T}^U\) with \(E' \supseteq E\) such that \(T \setminus E = T \setminus E'\) and \(\tau_{E'}(\theta, \zeta) > -1 + |\zeta|\).

**Proof.** All properties other than the existence of \(E'\) are clear from the geometry of trees. The set \(E'\) can be defined by

\[
E' = E \cup \left( \bigcup_{k \in \mathbb{Z}, |k| \geq 2} D(x_{T} + ks_{T}, (|k| - 1)s_{T}) \right);
\]

and \(\tau_{E'}(\theta, \zeta) > -1 + |\zeta|\) follows. \(\square\)

Let us get to the proof of (3.5). There are two terms in the local size \(\mathcal{S}_{\Theta}^*\) (see Definition 2.31) that we need to control: the Lebesgue term, and the randomised term.

### 4.1 The Lebesgue term

First we prove

\[
\|1_{\mathbb{R}^3_{+} \setminus E}[g]\|_{\mathcal{O}_{in}^T} \lesssim \|1_{\mathbb{R}^3_{+} \setminus E}M[g]\|_{\mathcal{O}_{in}^\infty}
\]

(4.1)

for all trees \(T \in \mathcal{T}_{\Theta}\). By translation, dilation, and modulation symmetry, we may assume that \(T = T(0, 0, 1)\), and by homogeneity, we may assume that the right-hand side of (4.1) is equal to
1. Having made these reductions, it suffices to show

\[ \int_{B_1} \int_0^1 \left\| \pi^*_I(Y \setminus E_1) A[g](\theta, \zeta, \sigma) \right\|_Y \frac{d\sigma}{\sigma} d\zeta \lesssim c_\theta \]  
(4.2)

for every frequency \( \theta \in \Theta \cap \{ \theta < 1 - \varepsilon \} \) such that there exists a point \((\theta, \zeta, \sigma) \in \pi^{-1}_I(T^{in} \setminus E)\), with constant \(c_\theta\) which is integrable over all such \( \theta \). From now on, we fix such a frequency \( \theta \), and we omit it from our notation wherever possible.

The left-hand side of (4.2) is controlled by

\[ \int_{B_1} \int_0^1 \int_{\mathbb{R}} \sum_{j \in \mathbb{N}} \left\| g_j(z) \right\|_Y |\Lambda_{(\theta \sigma^{-1}, 0, \sigma)} \psi_{\sigma}(\epsilon_j(z), \epsilon_{j+1}(z)) \rangle (z - \zeta) | dz \frac{d\sigma}{\sigma} d\zeta. \]  
(4.3)

By the conditions defining truncated wave packets, \( \psi_{\sigma}(\epsilon_j(z), \epsilon_{j+1}(z)) \) is non-vanishing only if \( \theta - \sigma \epsilon_j \in B_\varepsilon(1) \) and \( \theta - \sigma \epsilon_{j+1} < -1 + \varepsilon \), thus, since \( \theta < 1 - \varepsilon \), the integrand vanishes unless \( \epsilon_j < 0 < \epsilon_{j+1} \), and

\[ |\Lambda_{(\theta \sigma^{-1}, 0, \sigma)} \psi_{\sigma}(\epsilon_j(z), \epsilon_{j+1}(z)) \rangle (z - \zeta) | \lesssim \sigma^{-1} \left\langle \frac{z - \zeta}{\sigma} \right\rangle^{-N}. \]

Thus, (4.3) is bounded by

\[ \int_{B_1} \left( \int_{\mathbb{R}} \int_{\mathbb{R}} \right) \int g^*(z) \left| Y^{1(B_\varepsilon(1))(\theta - \sigma \epsilon^*(z))} \sigma^{-1} \left\langle \frac{z - \zeta}{\sigma} \right\rangle^{-N} dz \frac{d\sigma}{\sigma} d\zeta \]  
(4.4)

where \( \tau = \tau_E(\theta, \cdot) : \mathbb{R} \to \mathbb{R}_+ \), \( j^*(z) \in \mathbb{N} \) is the index (unique if it exists) that satisfies \( \epsilon_j(z) < 0 < \epsilon_{j+1}(z) \), \( g^*(z) := g_j^*(z) \), and \( \epsilon^*(z) := \epsilon_{j^*(z)}(z) \); if no such index \( j^* \) exists, set \( g^*(z) := 0 \), making the other quantities irrelevant.

Since \( \theta - \sigma \epsilon^*(z) \in B_\varepsilon(1) \) and \( \sigma > \tau(\zeta) \) implies \( \theta - \tau(\zeta) \epsilon^*(z) \in \Theta \),\(^\dagger\) the first summand in (4.4) is controlled by

\[ \int_{B_1} \int_{\mathbb{R}} \left| g^*(z) \right| \left| Y^{1(B_\varepsilon(1))(\theta - \tau(\zeta) \epsilon^*(z))} \left\langle \frac{z - \zeta}{\tau(\zeta)} \right\rangle^{-N} dz \right. \]  
\[ \leq \int_{B_1} M[g](\theta \tau(\zeta)^{-1}, \zeta, \tau(\zeta)) d\zeta \lesssim 1, \]

recalling the normalisation of the right-hand side of (4.1). Using Lemma 4.1, we find that the second summand in (4.4) is controlled by

\[ \int_{B_1} \int_{4\tau(\zeta)}^1 \int_{\mathbb{R}} \int_{B_{1/2}(y)} \left| g^*(z) \right| \left| Y^{1(B_\varepsilon(1))(\theta - \sigma \epsilon^*(z))} \sigma^{-1} \left\langle \frac{z - \zeta}{\sigma} \right\rangle^{-N} dz dy \frac{d\sigma}{\sigma} d\zeta. \]  
(4.5)

\(^\dagger\) This argument implicitly uses that \( B_\varepsilon(1) \subset \Theta \), that is, that \( \Theta \) is larger than \( \Theta_{in} \). We will use this argument repeatedly.
We split this integral into two parts, according to the conditions $\sigma > \tau(y)$ and $\sigma < \tau(y)$. In the first case, reasoning as before we have $\theta - \tau(y)e^*(z) \in \Theta$, so the integral is bounded by

$$\int_{B_1} \int_{1/4\tau(\zeta)} \int_{\mathbb{R}} \int_{B_{\tau(y)/2}(y)} \| g^*(z) \|_Y 1_{\Theta}(\theta - \tau(y)e^*(z)) \, dz$$

$$\times 1_{B_{(1)}(\theta - \sigma e^*(z))} \sigma^{-1} \left( \frac{y - \zeta}{\sigma} \right)^{-N} d\sigma \frac{d\sigma}{d\zeta}$$

$$\lesssim \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \int_{1/4\tau(\zeta)} \int_{B_{\tau(y)/2}(y)} \| g^*(z) \|_Y 1_{\Theta}(\theta - \tau(y)e^*(z)) \, dz \right) d\gamma \langle y \rangle^{-2} dy$$

$$\lesssim \log \left( \frac{1 + \varepsilon - \theta}{1 - \varepsilon - \theta} \right) \int_{\mathbb{R}} \int_{\mathbb{R}} \| g^*(z) \|_Y 1_{\Theta}(\theta - \tau(y)e^*(z)) \, dz \langle y \rangle^{-2} dy$$

with $c_\theta$ integrable over $\{\theta \in \Theta : \theta < 1 - \varepsilon\}$. The normalisation of the right-hand side of (4.1) is used in the last step.

In the second case, in which $\sigma < \tau(y)$, the integrand of (4.5) is non-vanishing only if $\tau(y)/4 > \tau(\zeta)$. This implies $|\tau(y) - \tau(\zeta)| > 3\tau(y)/4$, and since $\tau$ is 1-Lipschitz, we have that $|y - \zeta| > 3\tau(y)/4$. Since $|z - y| < \tau(y)/2$, we get $4^{-1} |z - \zeta| \leq |y - \zeta| \leq 4 |z - \zeta|$, and thus $|z - \zeta| > \tau(y)/4$. Using this, and the trivial bound $\langle \frac{y - \zeta}{\sigma} \rangle^{-1} \lesssim \langle y \rangle^{-1}$, the integral (4.5) is controlled by

$$\int_{B_1} \int_{\mathbb{R}} \int_{1/4\tau(\zeta)} \int_{B_{\tau(y)/2}(y)} \| g^*(z) \|_Y 1_{B_{(1)}(\theta - \sigma e^*(z))} \sigma^{-1} \left( \frac{z - \zeta}{\sigma} \right)^{-N/2} d\zeta$$

$$\times \frac{\sigma}{\tau(y)} \sigma^{-1} \left( \frac{y - \zeta}{\sigma} \right)^{-N/2+2} 1_{\sigma < \tau(y)} d\sigma \frac{d\sigma}{d\zeta}$$

$$\lesssim \int_{\mathbb{R}} \int_{0} \left( \int_{B_1} M[g](\theta \sigma^{-1}, \zeta, \sigma) \sigma^{-1} \left( \frac{y - \zeta}{\sigma} \right)^{-N/2+2} 1_{4\tau(\zeta) < \sigma} d\zeta \right)$$

$$\times \frac{\sigma}{\tau(y)} 1_{\sigma < \tau(y)} d\sigma \langle y \rangle^{-2} dy \lesssim 1,$$
We need to show that

\[ \| 1_{\mathbb{R}^3 \setminus E} A_g \|_{\mathbb{R}(T)} \lesssim \| 1_{\mathbb{R}^3 \setminus E} M_T' \|_{\mathbb{R}^\infty} \]

for all trees \( T \). As with the Lebesgue term, we may assume that \( T = T(0,0,1) \), and that the right-hand side of (4.6) is equal to 1. Having made these reductions, it suffices to show

\[
\left( \int_{\mathbb{R}} \left| \int_{\mathbb{R}} \sum_{j \in \mathbb{N}} g_j(z) \overline{\Psi_{\sigma}^{(c_j(z),c_{j+1}(z)))}}(z - \zeta) \, dz \right|^2 \frac{d\zeta}{\gamma_{\tau(\zeta),1;Y}} \right)^{1/2} \lesssim C_\vartheta \tag{4.6}
\]

for every frequency \( \vartheta \in \Theta \cap \{ \vartheta > 1 - \varepsilon \} \) such that there exists a point \((\vartheta, \zeta, \sigma) \in \pi^{-1}(T_{\text{out}} \setminus E)\), with constant \( C_\vartheta \) which is square-integrable over all such \( \vartheta \). From now on, we fix such a frequency \( \vartheta \), and we omit it from our notation wherever possible. As in the previous section, we let \( \tau(\cdot) = \tau_E(\vartheta, \cdot) \). According to Lemma 4.2, we may assume without loss of generality (by replacing \( E \) with a new set \( E' \) such that \( T \setminus E = T \setminus E' \)) that \(-1 + |\zeta| < \tau_E(\zeta) < 3 + |\zeta| \).

The left-hand side of (4.6) is controlled by

\[
\left( \int_{B_1} \left\| \int_{\mathbb{R}} \sum_{j \in \mathbb{N}} g_j(z) \overline{\Psi_{\sigma}^{(c_j(z),c_{j+1}(z)))}}(z - \zeta) \, dz \right|^2 \frac{d\zeta}{\gamma_{\tau(\zeta),1;Y}} \right)^{1/2}, \tag{4.7}
\]

with

\[
\overline{\Psi_{\sigma}^{(c_j(z),c_{j+1}(z)))}}(z) := \Lambda_{(\vartheta_0,0,\sigma)} \Psi_{(\vartheta_1,-1,\sigma)}^{(c_j(z),c_{j+1}(z)))}.
\]

We estimate (4.7) by duality. For all \( h \in C_0^\infty(\mathbb{R};Y^*) \) with \( h(\zeta,\sigma) \) vanishing when \( \sigma < \tau(\zeta) \) (the smoothness is not important here), we will show that

\[
\left| \int_{B_1} \int_0^1 \int_{\mathbb{R}} \sum_{j \in \mathbb{N}} \left( g_j(z) \overline{\Psi_{\sigma}^{(c_j(z),c_{j+1}(z)))}}(z - \zeta) ; h(\zeta,\sigma) \right) \, dz \, d\sigma \, d\zeta \right| \lesssim C_\vartheta \| h \|_{L^2(B_1;\gamma_{\tau(\zeta),1;Y}^\infty)} \tag{4.8}
\]

Using Lemma 4.1, we can control the left-hand side of (4.8) by

\[
\int_{\mathbb{R}} \int_{B_1} \sum_{j \in \mathbb{N}} \left| g_j(z); \left( \int_0^\tau \int_0^{\tau(x)} (h(\cdot,\sigma) * \overline{\Psi_{\sigma}^{(c_j(z),c_{j+1}(z)))}})(z) \, d\sigma \right) \right| \, dz \, dx.
\]

The second bracketed term — the one including the integral from 0 to \( \tau(x) \) — is our first error term, which we call \( \text{Err}^{(1)} \). The estimate

\[
\text{Err}^{(1)} \lesssim C_\vartheta \| h \|_{L^2(B_1;\gamma_{\tau(\zeta),1;Y}^\infty)}
\]

is proven in Section 4.5. As for the first bracketed summand, first note that the integrand vanishes unless \( \tau(x) \in (0,1) \), and since \( \tau(x) > -1 + |x| \), we can restrict to \( x \in B_2 \). Furthermore, since
\[ \tilde{\Psi}_\sigma^{(\epsilon_j(z), \epsilon_{j+1}(z))} \] vanishes unless \( \theta - \sigma \epsilon_j(z) \in B_\varepsilon(1) \), and since \( \theta - \sigma \epsilon_j \in B_\varepsilon(1) \) and \( \sigma > \tau(x) \) implies \( \theta - \tau(x) \epsilon_j \in \Theta \), we have

\[
\int \int_{B_\tau(x)/2(x)} \left| \sum_{j \in \mathbb{N}} \left( g_j(z); \int_{\tau(x)}^{1} (h(\cdot, \sigma) \ast \tilde{\Psi}_\sigma^{(\epsilon_j(z), \epsilon_{j+1}(z))})(z) \frac{d\sigma}{\sigma} \right) \right| dz \, dx
\]

\[
\leq \int \int_{B_2} \left( \sum_{j \in \mathbb{N}} \| g_j(z) \|_Y \right) \left( \int_{\tau(x)}^{1} (h(\cdot, \sigma) \ast \tilde{\Psi}_\sigma^{(\epsilon_j(z), \epsilon_{j+1}(z))})(z) \frac{d\sigma}{\sigma} \right)_{Y^*} \, dz \, dx
\]

\[
\leq \int \int_{B_2} \left( \sum_{j \in \mathbb{N}} \| g_j(z) \|_Y^r \right)^{1/r} \mathcal{H}_\tau(x)(z) \, dz \, dx,
\]

where for \( t > 0 \) we write

\[
\mathcal{H}_t(z) := \sup_{r \in \Delta} \left( \sum_{j \in \mathbb{N}} \left( \int_{t}^{1} (h(\cdot, \sigma) \ast \tilde{\Psi}_\sigma^{(\epsilon_j(z), \epsilon_{j+1}(z))})(z) \frac{d\sigma}{\sigma} \right)^r \right)^{1/r}.
\]

The last line of (4.9) is bounded by

\[
\int_{B_2} M[g](\theta \tau(x)^{-1}, x, \tau(x)) \sup_{z \in B_\tau(x)/2(x)} \mathcal{H}_\tau(x)(z) \, dx
\]

\[
\leq \int_{B_2} \sup_{z \in B_\tau(x)/2(x)} \mathcal{H}_\tau(x)(z) \, dx.
\]

We claim that

\[
\sup_{z \in B_\tau(x)/2(x)} \mathcal{H}_\tau(x)(z) \lesssim (\mathcal{M}H_\tau)(x),
\]

where \( \mathcal{H}_\tau(z) := \sup_{t \in (0,1)} \mathcal{H}_t(z) \) and \( \mathcal{M} \) is the Hardy–Littlewood maximal operator. Assuming (4.12), we have that the last expression in (4.11) is controlled by

\[
\int_{B_2} (\mathcal{M}H_\tau)(x) \, dx \lesssim \| \mathcal{M}H_\tau \|_{L^2(\mathbb{R})} \lesssim \| H_\tau \|_{L^2(\mathbb{R})}.
\]

Thus, assuming (4.12), it suffices to show

\[
\| H_\tau \|_{L^2(\mathbb{R})} \lesssim C_\theta \| h \|_{L^2(B_1; Y^*, 0, 1; Y^*)}.
\]

To show (4.12), note that the Fourier support of \( \tilde{\Psi}_\sigma^{(\epsilon_j(z), \epsilon_{j+1}(z))} \) is contained in \((0, |\Theta| + b)\). Let \( Y \in \mathcal{S}(\mathbb{R}) \) be such that \( \hat{Y} = 1 \) on \([0, |\Theta| + b]\) and set \( Y_t = \text{Dil}_t Y \). It then holds that
\[ \mathcal{H}_t(z) = \sup_{\sigma \in \Delta} \left( \sum_{j \in \mathbb{N}} \left\| \int_{B_{1/2}} \int_0^1 h(\zeta, \sigma) \tilde{\Psi}_{\sigma}^{(\epsilon_j, \epsilon_{j+1})} * Y_t(z - \zeta) \, d\zeta \right\|_{L^r}^r \right)^{1/r} \]

\[ \leq \int_{\mathbb{R}} \sup_{\sigma \in \Delta} \left( \sum_{j \in \mathbb{N}} \left\| \int_{B_{1/2}} \int_0^1 h(\zeta, \sigma) \tilde{\Psi}_{\sigma}^{(\epsilon_j, \epsilon_{j+1})}(z' - \zeta) \, d\zeta \right\|_{L^r}^r \right)^{1/r} Y_t(z - z') \, dz' \]

\[ = \int_{\mathbb{R}} \mathcal{H}_t(z') Y_t(z - z') \, dz' \lesssim \inf_{z' \in B_t(z)} (\mathcal{M} H_*)(z'), \]

which implies (4.12).

### 4.3 Controlling $\mathcal{H}_*$

To estimate $\mathcal{H}_* = \sup_{t \in (0,1)} \mathcal{H}_t$, we compare it with the variation of a family of convolution operators applied to a function constructed from $h$.

The truncated wave packet $\tilde{\Psi}_{\sigma}^{(c_-, c_+)}$ is non-vanishing only if

\[ \theta \in (\sigma c_+ + (1 - \epsilon), \sigma c_- + (1 - \epsilon)) \cap (\sigma c_- + (1 + \epsilon), \sigma c_+ + (1 + \epsilon)). \]

Since $\theta > 1 - \epsilon$, $\tilde{\Psi}_{\sigma}^{(c_-, c_+)}$ is non-vanishing only if $c_+ > 0$. This means that in (4.10), we need only consider $\epsilon \in \Delta$ such that $\epsilon_{j+1} > 0$ for all $j \in \mathbb{N}$. Henceforth we fix $t \in (0,1)$ and $z \in \mathbb{R}$, and we bound $\mathcal{H}_t(z)$ pointwise.

For any $\epsilon \in \Delta$, we define

\[ \sigma^-_{\epsilon, j} := \inf\{ \sigma \in (t, 1) : \tilde{\Psi}_{\sigma}^{(\epsilon_j, \epsilon_{j+1})} \neq 0 \}, \]

\[ \sigma^+_{\epsilon, j} := \sup\{ \sigma \in (t, 1) : \tilde{\Psi}_{\sigma}^{(\epsilon_j, \epsilon_{j+1})} \neq 0 \}, \]

with $\sigma^\pm_{\epsilon, j}$ implicitly depending on $t \in (0,1)$; let us drop the indices $j \in \mathbb{N}$ for which the interval $(\sigma^-_{\epsilon, j}, \sigma^+_{\epsilon, j})$ is empty. We have

\[ \sigma^-_{\epsilon, j} \geq \frac{\theta + (1 - \epsilon)}{\epsilon_{j+1}} \quad \forall j \in \mathbb{N}, \quad \sigma^+_{\epsilon, j} \leq \frac{\theta - (1 - \epsilon)}{\epsilon_{j+1}} \quad \forall j \in \mathbb{N} \]

and thus

\[ \frac{\sigma^+_{\epsilon, j+1}}{\sigma^-_{\epsilon, j}} \leq \frac{\theta - (1 - \epsilon)}{\theta + (1 - \epsilon)} < 1 - 2 \frac{(1 - \epsilon)}{\theta + (1 - \epsilon)}. \]

It follows that for any $\lambda > 1$, the intervals $(\lambda^{-1} \sigma^-_{\epsilon, j}, \lambda \sigma^+_{\epsilon, j})$ have finite overlap, so that

\[ \left\| \sum_{j \in \mathbb{N}} 1_{(\lambda^{-1} \sigma^-_{\epsilon, j}, \lambda \sigma^+_{\epsilon, j})} \right\|_{L^\infty} \lesssim \lambda |\theta| \epsilon 1 \]

with implicit constant independent of $\theta$ and $\epsilon$. Fix a large $\lambda > 1$ to be determined later.
By construction, we have

\[ \mathcal{H}_t(z) = \sup_{\sigma \in \Delta} \left( \sum_{j \in \mathbb{N}} \left\| \int_{\sigma_{-j}}^{\sigma_{+j}} h(\cdot, \sigma) \ast \tilde{\Psi}_{\sigma(j, j+1)}(z) \frac{d\sigma}{\sigma} \right\|_{Y^*}^r \right)^{1/r} ; \]

do not forget that the endpoints \( \sigma_{\pm j} \) depend implicitly on \( t \). We bound \( \mathcal{H}_t \) pointwise by

\[ \mathcal{H}_t(z) \lesssim \mathcal{G}(z) + \text{Err}_t^{(2)}(z), \]

where

\[ \text{Err}_t^{(2)}(z) := \sup_{\sigma \in \Delta} \left( \sum_{j \in \mathbb{N}} \left\| \int_{\sigma_{-j}}^{\sigma_{+j}} h(\cdot, \sigma) \ast (\tilde{\Psi}_{\sigma(j, j+1)} - \tilde{\Psi}_\sigma)(z) \frac{d\sigma}{\sigma} \right\|_{Y^*}^r \right)^{1/r} \]

with

\[ \tilde{\Psi}_\sigma(z) := \text{Dil}_\sigma \text{Mod}_\theta \Psi^{(0, +\infty)}(\theta \sigma - 1, \sigma)(z), \]

and

\[ \mathcal{G}(z) := \sup_{\sigma} \left( \sum_{j \in \mathbb{N}} \left\| \int_{\sigma_{-j}}^{\sigma_{+j}} h(\cdot, \sigma) \ast \tilde{\Psi}_\sigma(z) \frac{d\sigma}{\sigma} \right\|_{Y^*}^r \right)^{1/r} \]

with supremum taken over all sequences \( \sigma = (\sigma_{\pm j})_{j \in \mathbb{N}} \) satisfying

\[ 0 < \ldots < \sigma_{j+1}^+ < \sigma_j^- < \sigma_j^+ < \sigma_{j-1}^- < \ldots < 1 \quad \text{and} \quad \sum_{j \in \mathbb{N}} 1(\lambda^{-1} \sigma_j^- \lambda \sigma_j^+) \leq C \quad (4.16) \]

for \( C \) larger than the implicit constant in \((4.15)\). The term \( \mathcal{G} \) is further decomposed by introducing the family of functions

\[ H_\rho(z) := \int_0^\rho h(\cdot, \sigma) \ast \tilde{\Psi}_\sigma(z) \frac{d\sigma}{\sigma}, \]

so that

\[ \mathcal{G}(z) = \sup_{\sigma} \left( \sum_{j \in \mathbb{N}} \left\| H_{\sigma_{+j}}(z) - H_{\sigma_{-j}}(z) \right\|_{Y^*}^r \right)^{1/r} . \]

We will compare \( \mathcal{G} \) with the variation of a family of convolution operators applied to \( H_1 \). In particular for a fixed \( Y \in \mathcal{S}(\mathbb{R}) \) (to be chosen later), we obtain the pointwise bound

\[ \mathcal{G}(z) \leq \left\| H_1 \ast Y_\sigma(z) \right\|_{V'((0, 1); Y^*)} + \text{Err}_t^{(3)}(z) \]
with $Y_\sigma := \text{Dil}_\sigma Y$ and

$$\text{Err}^{(3)}(z) := \sup_{\sigma} \left( \sum_{j \in \mathbb{N}} \| (H_{\sigma_j^+}(z) - H_1 \ast Y_{\sigma_j^+}(z)) - (H_{\sigma_j^-}(z) - H_1 \ast Y_{\sigma_j^-}(z)) \|_{Y^*} \right)^{1/r}.$$

The Fourier support of $\tilde{\Psi}_\sigma$ is contained in $B_{b/\sigma}(\theta/\sigma)$, so we choose $Y$ so that $\text{spt} \hat{Y} \subset B_{\theta + 2b}$ and $\hat{Y} = 1$ on $B_{\theta + b}$.

Now we have the pointwise estimate

$$H_t(z) \lesssim \| \sigma \mapsto H_1 \ast Y_\sigma(z) \|_{V^r(0,1;Y^*)} + \text{Err}^{(2)}_t(z) + \text{Err}^{(3)}(z)$$

and it remains to show

$$\| \sigma \mapsto H_1 \ast Y_\sigma \|_{V^r(0,1;Y^*)} \lesssim C_0 \| h \|_{L^2(B_1; \gamma d\sigma/(0,1;Y^*))}, \quad (4.17)$$

$$\| \text{Err}^{(2)}_t(z) \|_{L^2(\mathbb{R})} \lesssim C_0 \| h \|_{L^2(B_1; \gamma d\sigma/(0,1;Y^*))}, \quad (4.18)$$

$$\| \text{Err}^{(3)} \|_{L^2(\mathbb{R})} \lesssim C_0 \| h \|_{L^2(B_1; \gamma d\sigma/(0,1;Y^*))}, \quad (4.19)$$

for any measurable function $t : \mathbb{R} \to (0,1)$.

### 4.4 Bounding the variation term

Due to Theorem 2.4 (valid since $Y^*$ has martingale cotype $r_0 < r$), we have

$$\| \| \sigma \mapsto H_1 \ast Y_\sigma \|_{V^r(0,1;Y^*)} \|_{L^2(\mathbb{R})} \lesssim \| H_1 \|_{L^2(\mathbb{R};Y^*)},$$

so to prove (4.17) it remains to show the bound

$$\| H_1 \|_{L^2(\mathbb{R};Y^*)} \lesssim C_0 \| h \|_{L^2(B_1; \gamma d\sigma/(0,1;Y^*))}.$$  

We estimate by duality by fixing $\mathcal{H} \in L^2(\mathbb{R}; Y)$ and writing

$$\left| \int_{\mathbb{R}} \langle H_1(z); \mathcal{H}(z) \rangle \, dz \right| = \left| \int_{\mathbb{R}} \int_{B_1} \langle h(\zeta, \sigma); \int_{\mathbb{R}} \mathcal{H}(z) \tilde{\Psi}_\sigma(z - \zeta) \, dz \rangle \, d\sigma \, d\zeta \right| \lesssim \| h \|_{L^2(B_1; \gamma d\sigma/(0,1;Y^*))} \left\| \int_{\mathbb{R}} \mathcal{H}(z) \tilde{\Psi}_\sigma(z - \zeta) \, dz \right\|_{L^2(\mathbb{R}; \gamma d\sigma/(0,1;Y))}.$$  

By the frequency-scale compatibility condition for $\Psi^{(0,+\infty)}$, we have

$$\tilde{\Psi}_\sigma = \text{Dil}_\sigma \text{Mod}_\theta \Psi^{(0, +\infty)}_{(\theta^{-1}, \sigma)} = \text{Dil}_\sigma \text{Mod}_\theta \Psi^{(0, +\infty)}_{(\theta, 1)}.$$
Since $\text{Mod}_\theta \Psi^{(0,+\infty)}(\theta,1)$ is uniformly bounded in arbitrary Schwartz seminorms and has Fourier support in $B_0(\theta)$ (and in particular has vanishing mean), by the continuous Littlewood–Paley estimate for UMD spaces (Proposition 2.10), we have
\[
\left\| \int_{\mathbb{R}} \mathcal{G}(z) \overline{\Psi}_\sigma(z - \xi) \, dz \right\|_{L^2_\gamma(B_1;\gamma_{dc/j}(0,1;Y))} \lesssim \| \mathcal{H} \|_{L^2(B_1;Y)}
\]
as required (here there is no dependence on $\theta$ in the constant, since $b$ is assumed to be sufficiently small).

### 4.5 Bounding the first error term $\text{Err}^{(1)}$

We show that
\[
\text{Err}^{(1)} = \int \int \sum_{j \in \mathbb{N}} \langle g_j(z); \int_{B_1} h(\xi, \sigma) \overline{\Psi}_\sigma(\xi, \tau(z))(z - \xi \frac{d\sigma}{\sigma}) \rangle \, dz \, dx \\
\lesssim C \| h \|_{L^2(B_1;\gamma_{dc/j}(0,1;Y^*))}.
\]

Decompose $\text{Err}^{(1)}$ into the sum of two terms, $\text{Err}^{(1,1)}$ and $\text{Err}^{(1,2)}$, by splitting the integral in $\sigma$ according to the conditions $\sigma < 4\tau(\xi)$ (yielding $\text{Err}^{(1,1)}$) and $\sigma > 4\tau(\xi)$ (yielding $\text{Err}^{(1,2)}$). By Lemma 2.11 and some rearranging, $\text{Err}^{(1,1)}$ is controlled by
\[
\int_{B_1} \int \sum_{j \in \mathbb{N}} \| g_j(z) \|_Y \left( \int \frac{\tau(z)}{\tau(\xi)} \right)^{-\frac{1}{2}} \| h(\xi, \sigma) \|_{\gamma_{dc/j}(0,1;Y^*)} \, dz \, d\xi \lesssim \| h(\xi, \sigma) \|_{L^2(B_1;\gamma_{dc/j}(0,1;Y^*))}.
\]

For fixed $\xi \in B_1$ and $z \in \mathbb{R}$, the integrand vanishes unless $\epsilon_j(z) > 0$ for all $j \in \mathbb{N}$. Furthermore, for any fixed $\xi, z \in B_1 \times \mathbb{R}$, there are at most finitely many indices $j \in \mathbb{N}$ (with quantity independent of $\theta > 1 - \epsilon$) such that the integrand above does not vanish; this follows from the fact that if $1_{(-\infty,-1+\epsilon)}(\theta - \sigma \epsilon(j+1)(z)) \neq 0$ for some $\sigma \in (0,1)$, then for any $j' \geq j + 1$ one has $1_{B_1}(\theta - \sigma \epsilon(j+1)(z)) = 0$ unless $\sigma' < \sigma \frac{1}{1+\epsilon} < \sigma(1 - \frac{1}{2|\Theta|})$ (assuming $\epsilon > 0$ is small enough). Since $\theta - \sigma \epsilon(j)(z) \in B_1(1) \subset \Theta$ and $\sigma > \tau(\xi)$ implies $\theta - \tau(\xi) \epsilon_j(z) \in \Theta$, we may control $\text{Err}^{(1,1)}$ by
\[
\int_{B_1} \int \left( \sum_{j \in \mathbb{N}} \| g_j(z) \|_Y^r \right) \tau(z)^{-\frac{1}{2}} \left( \frac{\tau(z)}{\tau(\xi)} \right)^{-\frac{1}{2}} \| h(\xi, \sigma) \|_{\gamma_{dc/j}(0,1;Y^*)} \, dz \, d\xi \lesssim \| h(\xi, \sigma) \|_{L^2(B_1;\gamma_{dc/j}(0,1;Y^*))}
\]
as required.
By Lemma 2.11, we can control $\text{Err}^{(1,2)}$ by

$$
\int_R \int_{B_1(x)/2(x)} \int_{B_1} \|h(\zeta,\sigma)\|_{\gamma_{\delta/\sigma}(0,1;\gamma)} \left( \int_{4\tau(\zeta)}^\tau (\sum_{j\in\mathbb{N}} \|g_j(z)\|_{Y^*} |\Psi_{\sigma}^{(\epsilon_j(z),\epsilon_j+1(z))}(z - \zeta)| \frac{d\sigma}{\sigma} \right)^{1/2} \frac{d\zeta}{\zeta} \frac{d\sigma}{\sigma} \frac{dz}{z} dx.
$$

For any $z \in \mathbb{R}$ and $\sigma \in (0,1)$, there is at most one $j \in \mathbb{N}$ such that $\Psi_{\sigma}^{(\epsilon_j(z),\epsilon_j+1(z))}$ is non-vanishing. Furthermore, $\Psi_{\sigma}^{(\epsilon_j(z),\epsilon_j+1(z))}(z - \zeta)$ vanishes unless $\theta - \sigma \epsilon_j(z) \in B_\epsilon(1) \subset \Theta$, so we can control $\text{Err}^{(1,2)}$ by

$$
\int_{B_1} \|h(\zeta,\sigma)\|_{\gamma_{\delta/\sigma}(0,1;\gamma)} \int_R \int_{B_1(x)/2(x)} \left( \sum_{j\in\mathbb{N}} \|g_j(z)\|_{Y^*} \right)^2 1_{B_{\epsilon/1}}(\theta - \sigma \epsilon_j(z)) \left( \int_{4\tau(\zeta)}^\tau \frac{\sigma}{\sigma} \frac{d\sigma}{\sigma} \left( \sum_{j\in\mathbb{N}} \|g_j(z)\|_{Y^*} 1_{B_\epsilon(1)}(\theta - \sigma \epsilon_j(z)) \right)^{1/2} \right) \frac{d\zeta}{\zeta} \frac{d\sigma}{\sigma} \frac{dz}{z} dx.
$$

Here we are integrating over $(x,\zeta) \in \mathbb{R} \times B_1$ such that $\tau(x) > 4\tau(\zeta)$. This implies that $|\tau(x) - \tau(\zeta)| > 3\tau(x)/4$, and since $\tau$ is 1-Lipschitz we have that $|x - \zeta| > 3\tau(x)/4$. Since $|x - z| < \tau(x)/2$, we get $4^{-1}|z - \zeta| \leq |x - \zeta| \leq 4|z - \zeta|$, and hence $|z - \zeta| > \tau(x)/4$. This yields

$$
|\Psi_{\sigma}^{(\epsilon_j(z),\epsilon_j+1(z))}(z - \zeta)| \lesssim \frac{\sigma^2}{\tau(x)^2} \left( \frac{z - \zeta}{\sigma} \right)^{-N+2}.
$$

Let $A(\zeta)$ denote the integral in $(x, z)$ appearing in (4.20). Using (4.21) and the Minkowski inequality, for $\zeta \in B_1$, we can control $A(\zeta)$ by

$$
\int_R \int_{B_1(x)/2(x)} \int_{4\tau(\zeta)}^\tau \frac{\sigma}{\tau(x)^2} \left( \sum_{j\in\mathbb{N}} \|g_j(z)\|_{Y^*} 1_{B_\epsilon(1)}(\theta - \sigma \epsilon_j(z)) \right)^{1/2} \frac{d\zeta}{\zeta} \frac{d\sigma}{\sigma} \frac{dz}{z} dx.
$$

$$
\lesssim \int_R \left( \sum_{j\in\mathbb{N}} \|g_j(z)\|^2_{Y^*} 1_{B_\epsilon(1)}(\theta - \sigma \epsilon_j(z)) \right)^{1/2} \frac{d\zeta}{\zeta} \frac{d\sigma}{\sigma} \frac{dz}{z} dx.
$$
Thus we can control $\text{Err}^{(1,2)}$ by

$$
\int_{B_1} \|h(\zeta, \sigma)\|_{\gamma_{\Delta/\sigma}(0,1;Y^*)} \int_{\mathbb{R}} \langle x \rangle^{-2} \int_{4\tau(\zeta)}^{\tau(x)} \frac{\sigma^3}{\tau(x)^3} \\
\times \int_{\mathbb{R}} \left( \sum_{j \in \mathbb{N}} \|g_j(z)\|_{Y^*} \right) \left( \theta - \sigma \epsilon_j(z) \right) \frac{1}{2} \sigma^{-1} \left( \frac{z - \zeta}{\sigma} \right)^{-N/2} \frac{d\sigma}{\sigma} \frac{d\zeta}{\zeta} \\
\times \sigma^{-1} \left( \frac{x - \zeta}{\sigma} \right)^{-N/2+4} \frac{dx}{x} \frac{d\sigma}{\sigma} \\
\lesssim \int_{B_1} \|h(\zeta, \sigma)\|_{\gamma_{\Delta/\sigma}(0,1;Y^*)} \int_{\mathbb{R}} \langle x \rangle^{-2} \int_{4\tau(\zeta)}^{\tau(x)} \frac{\sigma^3}{\tau(x)^3} \\
\times \left( \sum_{j \in \mathbb{N}} \left\| h(\cdot, \sigma)^{\epsilon_j(z)} \right\|_{\gamma_{\Delta/\sigma}(0,1;Y^*)} \right) \frac{1}{2} \sigma^{-1} \left( \frac{x - \zeta}{\sigma} \right)^{-N/2+4} \frac{d\sigma}{\sigma} dx d\zeta \\
\lesssim \int_{\mathbb{R}} \langle x \rangle^{-2} \frac{\sigma^3}{\tau(x)^3} \frac{d\sigma}{\sigma} \\
\times \sup_{\sigma > 4\tau(\zeta)} \int_{B_1} \|h(\zeta, \sigma)\|_{\gamma_{\Delta/\sigma}(0,1;Y^*)} \sigma^{-1} \left( \frac{x - \zeta}{\sigma} \right)^{-N/2+4} \frac{d\sigma}{\sigma} \frac{dx}{x} \\
\lesssim \int_{\mathbb{R}} \langle x \rangle^{-2} \mathcal{M}\left( \|h(\cdot, \sigma)\|_{\gamma_{\Delta/\sigma}(0,1;Y^*)} \right) (x) dx \\
\lesssim \|h(\zeta, \sigma)\|_{L^{2\gamma}_{\Delta/\sigma}(\gamma_{\Delta/\sigma}(0,1;Y^*))}
$$

as required.

### 4.6 Bounding the second error term $\text{Err}^{(2)}_{t(z)}$

For each measurable function $t : \mathbb{R} \to (0, 1)$, we show the bound (4.18). Fix a measurable function $c : \mathbb{R} \to \Delta$ and let

$$
P^{j,z}_\sigma := \tilde{\Psi}_{\sigma}^{(c_j(z), c_{j+1}(z))} - \tilde{\Psi}_\sigma \in \mathcal{S}(\mathbb{R}),
$$

so that for all $z \in \mathbb{R}$,

$$
\text{Err}^{(2)}_{t(z)}(z) \leq \left( \sum_{j \in \mathbb{N}} \left\| \int_{\sigma_{c_j}^-}^{\sigma_{c_j}^+} (h(\cdot, \sigma) \ast P^{j,z}_\sigma)(z) \frac{d\sigma}{\sigma} \right\|_{Y^*} \right)^{1/r},
$$

where the sequences $(\sigma_{c_j}^\pm)_{j \in \mathbb{N}}$, constructed in (4.13), implicitly depend on $z$ and $t(z)$. By Lemma 2.18, using that $Y^*$ is UMD with cotype $r$ and that the intervals $I_{c_j} := (\sigma_{c_j}^-(z), \sigma_{c_j}^+(z))$
have finite overlap, for any measurable \( q : \mathbb{N} \times \mathbb{R} \times \mathbb{R}_+ \rightarrow (0, \infty) \), we have

\[
\| \text{Err}^{(2)}_{1,\cdot} \|_{L^2(\mathbb{R})} \lesssim \| h \|_{L^2(B_1;\gamma(0,1;Y^*))} \sup_{j,z,\sigma} \left\| \text{Dil}_{\sigma^{-1}} P^{j,z}_{\sigma} \right\|_{\mathcal{L}^2(\mathbb{R})} \sup_{j,z} \| q_{j,z}(\sigma) \|_{L^2_{\sigma/\sigma'}(I_{j,z})},
\]

where \( \| \cdot \|_\sigma \) is a Schwartz seminorm of sufficiently high order. We claim that

\[(4.22)\]

\[
\| \text{Dil}_{\sigma^{-1}} P^{j,z}_{\sigma} \|_\sigma \lesssim |\sigma \epsilon_j(z)| + \max(3 + \theta - \sigma \epsilon_{j+1}(z), 0);
\]

setting \( q_{j,z}(\sigma) \) to be the right-hand side of (4.22), it will then suffice to show

\[(4.23)\]

\[
\sup_{j,z} \| q_{j,z}(\sigma) \|_{L^2_{\sigma/\sigma'}(I_{j,z})} \lesssim C_\theta,
\]

with a constant \( C_\theta \) that is square-integrable over \( \{ \theta \in \Theta : \theta > 1 - \varepsilon \} \). We claim that this estimate also holds.

We are left with proving the claimed estimates. Let us start with (4.22). Writing out the definitions, we see that

\[
\text{Dil}_{\sigma^{-1}} P^{j,z}_{\sigma} = \text{Mod}_\theta \left( \Psi^{(\epsilon_j(z),\epsilon_{j+1}(z))}_{(\theta \sigma^{-1},\sigma)} - \Psi^{(0,\infty)}_{(\theta \sigma^{-1},\sigma)} \right),
\]

so we can estimate

\[
\| \text{Dil}_{\sigma^{-1}} P^{j,z}_{\sigma} \|_\sigma \lesssim \left\| \Psi^{(\epsilon_j(z),\epsilon_{j+1}(z))}_{(\theta \sigma^{-1},\sigma)} - \Psi^{(0,\infty)}_{(\theta \sigma^{-1},\sigma)} \right\|_\sigma
\]

\[
\leq \left\| \Psi^{(\epsilon_j(z),\epsilon_{j+1}(z))}_{(\theta \sigma^{-1},\sigma)} - \Psi^{(0,\epsilon_j(z))}_{(\theta \sigma^{-1},\sigma)} \right\|_\sigma + \left\| \Psi^{(0,\infty)}_{(\theta \sigma^{-1},\sigma)} - \Psi^{(0,\epsilon_{j+1}(z))}_{(\theta \sigma^{-1},\sigma)} \right\|_\sigma.
\]

Use the smoothness condition of families of left-truncated wave packets to bound the first summand by

\[
\sigma \int_0^{\epsilon_j(z)} \left\| (\sigma^{-1} \delta_{c^-}) \Psi^{(\lambda,\epsilon_{j+1}(z))}_{(\theta \sigma^{-1},\sigma)} \right\|_\sigma d\lambda \leq |\sigma \epsilon_j(z)|.
\]

As for the second summand, note by weak dependence on the right endpoint that the second summand vanishes if \( \epsilon_{j+1}(z) > \theta \sigma^{-1} + 3 \sigma^{-1} \). Assuming this is not the case, the second summand can be written as

\[
\left\| \Psi^{(0,\epsilon_j(z))}_{(\theta \sigma^{-1},\sigma)} - \Psi^{(0,\epsilon_{j+1}(z))}_{(\theta \sigma^{-1},\sigma)} \right\|_\sigma \leq \sigma \int_{\epsilon_{j+1}(z)}^{\theta \sigma^{-1} + 3 \sigma^{-1}} \left\| (\sigma^{-1} \delta_{c^-}) \Psi^{(0,\lambda)}_{(\theta \sigma^{-1},\sigma)} \right\|_\sigma d\lambda
\]

\[
\leq 3 + \theta - \sigma \epsilon_{j+1}(z),
\]

again using the smoothness condition. This proves (4.22).
Finally, we prove (4.23). For fixed \( j \) and \( z \), we have

\[
\int_{\sigma_{c_j}^{-}(z)}^{\sigma_{c_j}^{+}(z)} |\sigma c_j(z)|^2 \frac{d\sigma}{\sigma} \leq |c_j(z)|^2 \int_{\sigma_{c_j}^{-}(z)}^{\sigma_{c_j}^{+}(z)} \sigma \, d\sigma \\
\leq (\sigma_{c_j}^{+}(z))^2 |c_j(z)|^2 \leq (\theta - (1 - \epsilon))^2 \lesssim 1
\]

by (4.14).

\[
\int_{\sigma_{c_j}^{-}(z)}^{\sigma_{c_j}^{+}(z)} \max(3 + \theta - \sigma c_{j+1}(z), 0)^2 \frac{d\sigma}{\sigma} \lesssim \int_{\theta - (1 - \epsilon)}^{3 + \theta} \frac{d\sigma}{\sigma} = \log \left( \frac{3 + \theta}{\theta - (1 - \epsilon)} \right).
\]

The function on the right-hand side is square-integrable over \( \{\theta \in \Theta : \theta > 1 - \epsilon\} \), so we are done.

### 4.7 Bounding the third error term \( \text{Err}^{(3)} \)

We are estimating the function \( \text{Err}^{(3)} \) given by

\[
\text{Err}^{(3)}(z) := \sup_{\sigma} \left( \sum_{j \in \mathbb{N}} \| (H_{\sigma_{c_j}^{+}(z)} + H_1 * Y_{\sigma_{c_j}^{+}(z)}) - (H_{\sigma_{c_j}^{-}(z)} + H_1 * Y_{\sigma_{c_j}^{-}(z)}) \|_{Y'}^{1/r} \right),
\]

for all \( z \in \mathbb{R} \), with supremum taken over all sequences \( \sigma \) satisfying (4.16). Fix such a sequence \( \sigma(z) \) for each \( z \), let \( \psi = \Psi^{(0,\infty)}_{(\theta^{-1}, \sigma)} = \Psi^{(0,\infty)}_{(\theta, 1)} \) (using frequency-scale compatibility), and \( \psi_{\sigma} := \text{Dil}_{\sigma} \psi \), and write

\[
(H_{\sigma_j^{+}(z)} + H_1 * Y_{\sigma_j^{+}(z)}(z)) - (H_{\sigma_j^{-}(z)} + H_1 * Y_{\sigma_j^{-}(z)}(z)) = \int_{\sigma_j^{-}(z)}^{\sigma_j^{+}(z)} h(\cdot, \sigma) * \psi_{\sigma}(z) \frac{d\sigma}{\sigma} + \int_{0}^{1} \left( h(\cdot, \sigma) * \psi_{\sigma} * (Y_{\sigma_j^{+}(z)} - Y_{\sigma_j^{-}(z)}) \right)(z) \frac{d\sigma}{\sigma}
\]

\[(4.24)\]

\[
= \int_{\sigma_j^{-}(z)}^{\sigma_j^{+}(z)} h(\cdot, \sigma) * \psi_{\sigma}(z) \frac{d\sigma}{\sigma} + \int_{\frac{\delta - b}{\delta + b} \sigma_j^{-}(z)}^{\frac{\theta + b}{\theta - b} \sigma_j^{+}(z)} h(\cdot, \sigma) \psi_{\sigma} * (Y_{\sigma_j^{+}(z)} - Y_{\sigma_j^{-}(z)}) \frac{d\sigma}{\sigma},
\]

using that \( \psi_{\sigma} * (Y_{\sigma_j^{+}(z)} - Y_{\sigma_j^{-}(z)}) = 0 \) for \( \sigma \notin \left( \frac{\theta - b}{\theta + b} \sigma_j^{-}(z), \frac{\theta + b}{\theta - b} \sigma_j^{+}(z) \right) \). Let \( N_1 \) denote the set of \( j \in \mathbb{N} \) such that

\[
\frac{\theta + b}{\theta} \sigma_j^{-}(z) > \frac{\theta - b}{\theta + b} \sigma_j^{+}(z),
\]

and let \( N_2 := \mathbb{N} \setminus N_1 \). Write

\[
\text{Err}^{(3)}_{\sigma}(z) \leq \text{Err}^{(3,1)}_{\sigma}(z) + \text{Err}^{(3,2)}_{\sigma}(z),
\]
where the two functions on the right-hand side have the same form as $\text{Err}_\sigma^{(3)}(z)$ (meaning $\text{Err}_\sigma^{(3)}(z)$ with a particular choice of sequence $\sigma$ substituted in) but summing over only $N_1$ or $N_2$. We will prove the two estimates

\[
\|\text{Err}_\sigma^{(3,1)}\|_{L^2(\mathbb{R})} \lesssim \|h\|_{L^2(B_1; Y_{d\sigma}/\sigma(0,1;Y^*))},
\]

\[
\|\text{Err}_\sigma^{(3,2)}\|_{L^2(\mathbb{R})} \lesssim \|h\|_{L^2(B_1; Y_{d\sigma}/\sigma(0,1;Y^*))}
\]

with implicit constants independent of $\theta$. These will imply (4.19), and ultimately complete the proof of Theorem (3.6), which leads to the main theorem of this paper.

To estimate $\text{Err}_\sigma^{(3,1)}$, for $j \in N_1$ and $z \in \mathbb{R}$, write

\[
(H_{\sigma_j^+(z)} - H_1 \ast Y_{\sigma_j^+(z)}) + (H_{\sigma_j^-(z)} - H_1 \ast Y_{\sigma_j^-(z)})
\]

\[
\int_0^\infty h(\cdot, \sigma) \ast \text{Dil}_{\sigma_j} P_{\sigma_j}^{(1,\sigma)}(z) \frac{d\sigma}{\sigma},
\]

where

\[
\text{Dil}_{\sigma_j} P_{\sigma_j}^{(1,\sigma)}(w) = 1_{(\sigma_j^-(z), \sigma_j^+(z))}(\sigma) \psi_\sigma(w)
\]

\[
+ \frac{1}{(\partial \sigma \mathbf{b})^+} \frac{1}{(\partial \sigma \mathbf{b})^-} \left[ \psi_\sigma \ast (Y_{\sigma_j^+(z)} - Y_{\sigma_j^-(z)}) \right](w).
\]

Clearly, we have that $P_j : \mathbb{R} \times \mathbb{R}_+ \to \mathcal{S}(\mathbb{R})$ is piecewise smooth. Furthermore,

\[
\sup_{j,z,\sigma} \|P_{\sigma_j}^{(1,\sigma)}\|_\ast \lesssim 1
\]

for any Schwartz seminorm $\| \cdot \|_\ast$; this follows from a similar argument to the proof of (4.22), using that $\sigma_j^-(z) \simeq \sigma_j^+(z)$ for $j \in N_1$, and that $P_{\sigma_j}^{(1,\sigma)}$ vanishes unless $\sigma \simeq \sigma_j^\pm(z)$. Taking $\lambda$ large enough, we can ensure that the intervals $(\frac{\omega - \mathbf{b}}{\omega + \mathbf{b}} \sigma_j^-(z), \frac{\omega + \mathbf{b}}{\omega - \mathbf{b}} \sigma_j^+(z))$ have finite overlap, so since $Y^*$ is UMD with cotype $r$, Lemma 2.18 yields

\[
\|\text{Err}_\sigma^{(3,1)}\|_{L^2(\mathbb{R})} \lesssim \|h\|_{L^2(B_1; Y_{d\sigma}/\sigma(0,1;Y^*))} \sup_{j,z} \|1\|_{L^2(\frac{\omega - \mathbf{b}}{\omega + \mathbf{b}} \sigma_j^-(z), \frac{\omega + \mathbf{b}}{\omega - \mathbf{b}} \sigma_j^+(z))}
\]

\[
\lesssim \|h\|_{L^2(B_1; Y_{d\sigma}/\sigma(0,1;Y^*))}
\]

using that $\sigma_j^- \simeq \sigma_j^+$ for $j \in N_1$ in the last line. This proves (4.25).

Finally, we estimate $\text{Err}_\sigma^{(3,2)}$. For $z \in \mathbb{R}$ and $j \in N_2$, using that

\[
\psi_\sigma = \psi_\sigma \ast Y_{\sigma_j^-}(z)
\]
for $\sigma > \frac{\partial + b}{\partial} \sigma^{-}(z)$, the last line of (4.24) is equal to

$$
\int_{\sigma^{-}(z)}^{\sigma^{+}(z)} h(\cdot, \sigma) * \psi_{\sigma}(z) \frac{d\sigma}{\sigma} + \int_{\sigma^{-}(z)}^{\sigma^{+}(z)} h(\cdot, \sigma) \psi_{\sigma} * \left(Y_{\sigma^{+}(z)} - Y_{\sigma^{-}(z)}(z)\right) \frac{d\sigma}{\sigma}
$$

$$
= \int_{\sigma^{-}(z)}^{\sigma^{+}(z)} h(\cdot, \sigma) * \psi_{\sigma}(z) \frac{d\sigma}{\sigma} + \int_{\sigma^{-}(z)}^{\sigma^{+}(z)} \left( h(\cdot, \sigma) * \psi_{\sigma} * Y_{\sigma^{+}(z)}(z) \right) \frac{d\sigma}{\sigma}
$$

$$
- \int_{\sigma^{-}(z)}^{\sigma^{+}(z)} \left( h(\cdot, \sigma) * \psi_{\sigma} * Y_{\sigma^{-}(z)}(z) \right) \frac{d\sigma}{\sigma}
$$

$$
= \int_{\sigma^{-}(z)}^{\sigma^{+}(z)} h(\cdot, \sigma) * \psi_{\sigma}(z) \frac{d\sigma}{\sigma} + \int_{\sigma^{-}(z)}^{\sigma^{+}(z)} \left( h(\cdot, \sigma) * \psi_{\sigma} * Y_{\sigma^{+}(z)}(z) \right) \frac{d\sigma}{\sigma}
$$

$$
- \int_{\sigma^{-}(z)}^{\sigma^{+}(z)} \left( h(\cdot, \sigma) * \psi_{\sigma} * Y_{\sigma^{-}(z)}(z) \right) \frac{d\sigma}{\sigma} - \int_{\sigma^{+}(z)}^{\sigma^{-}(z)} h(\cdot, \sigma) * \psi_{\sigma}(z) \frac{d\sigma}{\sigma}.
$$

We can thus express $\text{Err}_{\sigma}^{(3,2)}$ as the sum of four functions ($\text{Err}_{\sigma}^{(3,2,1)}$), each corresponding to the summands above. Each of these can be bounded via Lemma 2.18, as we did for $\text{Err}_{\sigma}^{(3,1)}$, yielding

$$
\|\text{Err}_{\sigma}^{(3,2)}\|_{L^2(\mathbb{R})} \lesssim \|h\|_{L^2(B_1;|\gamma_{\sigma}|(0,1;Y^*))}
$$

and completing the proof.

5 | A DISCUSSION OF BANACH FUNCTION SPACES

If $T \in \mathcal{L}(L^p(\mathbb{R}; \mathbb{C}))$ is a bounded linear operator, then for every Banach space $X$, one can consider the tensor extension $T \otimes I_X : L^p(\mathbb{R}; \mathbb{C}) \otimes X \to L^p(\mathbb{R}; \mathbb{C}) \otimes X$, defined on elementary tensors by

$$(T \otimes I_X)(f \otimes x) := Tf \otimes x \quad \forall f \in L^p(\mathbb{R}; \mathbb{C}), x \in X,$$

and extended by linearity to the whole algebraic tensor product $L^p(\mathbb{R}; \mathbb{C}) \otimes X$ (which is the linear span of elementary tensors, and not a Banach space). This algebraic tensor product is dense in the Bochner space $L^p(\mathbb{R}; X)$, and if the bound

$$
\|(T \otimes I_X)g\|_{L^p(\mathbb{R}; X)} \lesssim_X \|g\|_{L^p(\mathbb{R}; X)}
$$

holds for all $g \in L^p(\mathbb{R}; \mathbb{C}) \otimes X$, then $T \otimes I_X$ has a bounded extension to $L^p(\mathbb{R}; X)$. The extension problem for $T$ asks for which $X$ the bound (5.1) holds. This problem has been solved for many
operators in harmonic analysis, most notably the Hilbert transform; the solution is that \( X \) satisfies (5.1) (for all \( p \in (1, \infty) \)) if and only if \( X \) is UMD.

The scalar-valued \( r \)-variational Carleson operator \( \mathcal{V}_r, \mathbb{C} \), that is, the operator defined in (1.2) in the case \( X = \mathbb{C} \), is bounded on \( L^p(\mathbb{R}; \mathbb{C}) \) for all \( p \in (1, \infty) \) provided that \( r > 2 \). However, we cannot formulate the abstract extension problem detailed above, as \( \mathcal{V}_r, \mathbb{C} \) is not linear — only sublinear. This is not because no extensions exist; on the contrary, when \( X \) is a Banach function space, there are two natural extensions of \( \mathcal{V}_r, \mathbb{C} \). One can be bounded by the extrapolation methods of the first author, Lorist, and Veraar [1], and the other is the operator \( \mathcal{V}_r \) considered above, which does not see the function space structure of \( X \). In this section, we clarify the difference between these extensions, and compare the results of this article with those of [1]. But first we have to introduce the language of Banach function spaces.

For a measure space \( \Omega \) (always assumed to be \( \sigma \)-finite), we let \( \Sigma(\Omega) \) denote the vector space of simple functions \( \Omega \to \mathbb{C} \) and \( \mathcal{L}_0(\Omega) \) the vector space of measurable functions \( \Omega \to \mathbb{C} \) (modulo equality almost everywhere).

**Definition 5.1.** Let \( \Omega \) be a \( \sigma \)-finite measure space. A subspace \( X \) of \( \mathcal{L}_0(\Omega) \), equipped with a norm \( \| \cdot \|_X \), is called a Banach function space (over \( \Omega \)) if it satisfies the following properties.

- If \( x \in \mathcal{L}_0(\Omega) \), \( y \in X \), and \(|x| \leq |y|\), then \( x \in X \) and \( \|x\|_X \leq \|y\|_X \).
- There exists \( \zeta \in X \) with \( \zeta > 0 \).
- If \( 0 \leq x_n \uparrow x \) with \((x_n)_{n \in \mathbb{N}} \) a sequence in \( X \), \( x \in \mathcal{L}_0(\Omega) \). \( \sup_{n \in \mathbb{N}} \|x_n\|_X < \infty \), then \( x \in X \) and \( \|x\|_X = \sup_{n \in \mathbb{N}} \|x_n\|_X \).

If \( X \) is a Banach function space and \( x \in X \), then the function \(|x| \in X \) is well defined. It is also possible to take powers \(|x|^p \), which are in \( \mathcal{L}_0(\Omega) \) but need not be in \( X \). Measuring such functions relates to the geometric concepts of convexity and concavity of Banach function spaces. It also leads to the construction of new Banach function spaces, called concavifications.

**Definition 5.2.** Let \( X \) be a Banach function space, \( p \in [1, \infty] \), and \( s \in (0, \infty) \).

- We say that \( X \) is \( p \)-convex if
  \[
  \left\| \left( \sum_j |x_j|^p \right)^{1/p} \right\|_X \lesssim \left( \sum_j \|x_j\|_X^p \right)^{1/p} \tag{5.2}
  \]
  for all finite sequences \((x_j)\) in \( X \), with the usual modification when \( p = \infty \). We say that \( X \) is \( p \)-concave if the reverse estimate holds.
- We define the \( s \)-concavification \( X^s \) of \( X \) by
  \[
  X^s := \{ x \in \mathcal{L}_0(\Omega) : |x|^{1/s} \in X \}
  \]
  and equip it with the quasi-norm
  \[
  \|x\|_{X^s} := \| |x|^{1/s} \|_X^s.
  \]
Every Banach function space $X$ is 1-convex and $\infty$-concave, and if $X$ is $p$-convex, $q$-concave, and infinite-dimensional, then $p \leq q$. If $X$ is $p_0$-convex and $q_0$-concave, then it is $p$-convex and $q$-concave for all $p \in [1, p_0]$ and $q \in [q_0, \infty]$. The Lebesgue space $L^r(\Omega)$ is both $r$-convex and $r$-concave. If $X$ is $p$-convex, for infinite sequences; likewise if $X$ is $q$-concave. The $s$-concavification $X^s$ is equivalent to a Banach space (that is, $\| \cdot \|_{X^s}$ is equivalent to a norm) if and only if $X$ is $p$-convex for some $p \geq s$. A key example is $L^p(\Omega)^s = L^{p/s}(\Omega)$.

Given a measure space $\Omega$, every function $f : \mathbb{R} \to L^0(\Omega)$ can be identified with both a function $f : \mathbb{R} \times \Omega \to \mathbb{C}$ (which we denote by the same letter) and a set of functions $\{ f_\omega : \mathbb{R} \to \mathbb{C} : \omega \in \Omega \}$. In this notation, we have
\[ f(t)(\omega) = f(t, \omega) = f_\omega(t). \]

Now we can define two different $r$-variational Carleson operators on $X$-valued functions.

**Definition 5.3.** Let $X$ be a Banach function space over a measure space $\Omega$, and let $r \in (2, \infty)$.

- For $f \in \mathcal{S}(\mathbb{R}; X)$, we define the norm $r$-variational Carleson operator $\mathcal{V}_r^{\cdot, nm} f : \mathbb{R} \to \mathbb{C}$ by
  \[ \mathcal{V}_r^{\cdot, nm} f(t) := \| \xi \mapsto \mathcal{C}_\xi f(t) \|_{V^r(\mathbb{R}; X)}. \]

  This is the operator considered throughout the rest of the article, which does not see the Banach function space structure of $X$.

- For $f \in \mathcal{S}(\mathbb{R}; X)$, we define the pointwise $r$-variational Carleson operator $\mathcal{V}_r^{\cdot, pt} f : \mathbb{R} \times \Omega \to \mathbb{C}$ by
  \[ \mathcal{V}_r^{\cdot, pt} f(t, \omega) := (\mathcal{V}_r^{\cdot, C} f_\omega)(t) = \| \xi \mapsto \mathcal{C}_\xi f_\omega(t) \|_{V^r(\mathbb{R}; \mathbb{C})}. \]

The operator $\mathcal{V}_r^{\cdot, pt} f$ is the lattice extension of the operator $\mathcal{V}_r^{\cdot, C}$ on scalar-valued functions; an analogous definition can be made for any operator (not necessarily linear) acting on scalar-valued functions. We can equivalently write
\[ \mathcal{V}_r^{\cdot, pt} f(t) = \sup_{\epsilon \in \Delta} \left( \sum_{j \in \mathbb{N}} | \mathcal{C}_{f_{j+1}} f(t) - \mathcal{C}_{f_j} f(t) |^r \right)^{1/r}, \]
with the supremum and $r$th powers on the right-hand side being taken in the Banach function space $X$. When $X = \mathbb{C}$, the operators $\mathcal{V}_r^{\cdot, nm}$ and $\mathcal{V}_r^{\cdot, pt}$ coincide, so both of these operators can be considered as extensions of $\mathcal{V}_r^{\cdot, C}$.

In [1, §5], the pointwise operator $\mathcal{V}_r^{\cdot, pt}$ is considered (denoted there by $C_r$). Since this is the lattice extension of $\mathcal{V}_r^{\cdot, C}$, and since $\mathcal{V}_r^{\cdot, C}$ satisfies appropriate weighted estimates (proved in [13]), the rescaled Rubio de Francia extrapolation theorem [1, Corollary 3.6] implies the following bounds for the pointwise operator (stated as [1, Theorem 5.2]).

† In this identification, we ignore issues of measurability.
‡ The very recent results of [30] can also be used to obtain these bounds directly from the sparse domination of $\mathcal{V}_r^{\cdot, C}$ established in [13], bypassing the use of weighted estimates.
**Theorem 5.4.** Suppose \( r \in (2, \infty) \), and let \( X \) be a Banach function space such that the \( r' \)-concavification \( X^{r'} \) is UMD. Then for all \( p \in (r', \infty) \), we have the estimates

\[
\|\mathcal{V}_{r, pt} f \|_{L^p(\mathbb{R}; X)} \lesssim_{X, p, r} \| f \|_{L^p(\mathbb{R}; X)}.
\]

When \( X \) is \( r \)-convex or \( r \)-concave, we have comparability between the two extensions.

**Proposition 5.5.** Let \( X \) be a Banach function space. If \( X \) is \( r \)-convex, then

\[
\|\mathcal{V}_{r, pt} f(t)\|_X \leq |\mathcal{V}_{r, nm} f(t)| \quad \forall f \in \mathcal{D}(\mathbb{R}; X), \forall t \in \mathbb{R}.
\]

The reverse estimate holds if \( X \) is \( r \)-concave.

**Proof.** For all measurable functions \( \epsilon : \mathbb{R} \to \Delta \) and all \( t \in \mathbb{R} \), we have

\[
\left\| \left( \sum_{j \in \mathbb{N}} |\mathcal{C}_{t+j}(t) f(t) - \mathcal{C}_{t+j}(t) f(t)|^{r'} \right)^{1/r} \right\|_X \leq \left( \sum_{j \in \mathbb{N}} \left\| \mathcal{C}_{t+j}(t) f(t) - \mathcal{C}_{t+j}(t) f(t) \right\|_X^{r} \right)^{1/r}
\]

if \( X \) is \( r \)-convex, and the reverse estimate holds if \( X \) is \( r \)-concave. Taking the supremum over all such \( \epsilon \) completes the proof. \( \square \)

By combining Proposition 5.5 with Theorems 5.4 and 1.1, it is possible to obtain new conditions for boundedness of the two variational Carleson operators. First let us obtain bounds for \( \mathcal{V}_{r, nm} \).

**Theorem 5.6.** Let \( r \in (2, \infty) \), and let \( X \) be an \( r \)-concave Banach function space such that the \( r' \)-concavification \( X^{r'} \) is UMD. Then for all \( p \in (r', \infty) \),

\[
\|\mathcal{V}_{r, nm} f \|_{L^p(\mathbb{R})} \lesssim_{p, r, X} \| f \|_{L^p(\mathbb{R}; X)} \quad \forall f \in \mathcal{D}(\mathbb{R}; X).
\]

The proof is a simple combination of the reverse estimate to (5.3) and the bound from Theorem 5.4. The Banach function space \( X = L^r(\Omega) \) satisfies the hypothesis of this theorem, but it does not satisfy those of Theorem 1.1, as \( X \) is not \( r_0 \)-intermediate UMD for any \( r_0 < r \). Furthermore, for all \( s \geq r \), the space \( X = L^s(\Omega) \) satisfies the hypotheses of the theorem, and we get bounds for \( \mathcal{V}_{r, nm} \) beyond the scope of Theorem 1.1 (as we get \( L^p \) bounds for all \( p > r' \), not just \( p > (r/(s-1))' \) as in Theorem 1.1). Thus, when restricted to Banach function spaces, extrapolating from weighted scalar estimates is stronger than Theorem 1.1 (but of course, Theorem 1.1 applies to intermediate UMD spaces that have nothing to do with function spaces).

Here is what happens when we try to obtain bounds for \( \mathcal{V}_{r, pt} \) in the same way.

**Theorem 5.7.** Let \( r \in (2, \infty) \), and suppose that \( X \) is an \( r \)-convex Banach function space. Suppose furthermore that \( X \) is \( r_0 \)-intermediate UMD for some \( r_0 \in [2, r) \). Then

\[
\|\mathcal{V}_{r, pt} f \|_{L^p(\mathbb{R}; X)} \lesssim_{p, r, X} \| f \|_{L^p(\mathbb{R}; X)} \quad \forall p \in (1, \infty).
\]

As before, the proof is a simple combination of (5.3) and Theorem 1.1. But there is a catch. Suppose that \( X \) is infinite-dimensional and satisfies the hypotheses of this theorem. Then \( X \) has cotype
which by \cite[Corollary 1.f.9]{29} implies that $X$ is $q$-concave for all $q > r_0$. Since $X$ is infinite-dimensional and $r$-convex, it follows that $r \leq q$ for all $q > r_0$, and hence that $r \leq r_0$. But this contradicts the assumption that $r_0 < r$, so the only spaces $X$ satisfying the hypotheses of Theorem 5.7 are finite dimensional, in which case everything reduces to the scalar case anyway. Thus in the restricted context of Banach function spaces, our results are weaker than the extrapolation results of \cite{1}. But for more general UMD Banach spaces such results do not apply, and at this point our result is the only one available.

6 \quad BOUNDS FOR THE WAVE PACKET EMBEDDING $\text{Emb}_\varphi$

Recall that for a Banach space $Y$ and a Schwartz function $\varphi \in \mathcal{S}(\mathbb{R})$, for all $f \in \mathcal{S}(\mathbb{R}; Y)$, we define $\text{Emb}_\varphi[f] : \mathbb{R}^3_+ \to Y$ by

\[ \text{Emb}_\varphi[f](\eta, y, t) := \langle f; \Lambda(\eta, y, t) \varphi \rangle = \int_{\mathbb{R}} f(x) t^{-1} e^{-2\pi i \eta(x-y)} \varphi \left( \frac{x-y}{t} \right) dx. \]

In this section, we provide a detailed sketch of the proof of \cite[Theorem 3.2]{3}. We reproduce the theorem here for convenience.

**Theorem** (Theorem 3.2). Let $\Theta$ and $\Theta_{in}$ be bounded intervals with $0 \in \Theta_{in} \subset \Theta$, and suppose $\varphi \in \mathcal{S}(\mathbb{R})$ has Fourier support in $\Theta_{in}$. Let $Y$ be an $r_0$-intermediate UMD space for some $r_0 \in [2, \infty)$. Then for all $p \in (1, \infty)$ and $q \in (\min(p, r_0)'(r_0 - 1), \infty)$,

\[ \| \text{Emb}_\varphi[f] \|_{L_p L_k^q Y} \lesssim \| f \|_{L_p(\mathbb{R}; Y)} \quad \forall f \in \mathcal{S}(\mathbb{R}; Y). \quad (3.2) \]

In this theorem, the embedding map $\text{Emb}_\varphi$ maps a Bochner space $L_p(\mathbb{R}; Y)$ into an iterated outer Lebesgue space $L_p L_k^q Y$. The reader should review the notions of (iterated) outer Lebesgue spaces that are illustrated in Section 2.2; specifics of this construction in the time-frequency space appear in Section 2.3.

By the outer Lebesgue space version of Marcinkiewicz interpolation, which we stated earlier as Proposition 2.24, it suffices to prove a range of weak estimates

\[ \| \text{Emb}_\varphi[f] \|_{L_p^\infty L_k^q Y} \lesssim \| f \|_{L_p(\mathbb{R}; Y)} \quad \forall f \in \mathcal{S}(\mathbb{R}; Y). \quad (6.1) \]

Let us unravel the definition of the bound (6.1) for fixed $p, q$. Fix $f \in L_p(\mathbb{R}; Y)$. For every $\lambda > 0$, we are required to exhibit an at-most-countable collection of strips (see Definition 2.27) $D_\lambda \subset \mathbb{D}$ such that

\[ \lambda^p \sum_{D \in D_\lambda} s_D \lesssim \| f \|_{L_p(\mathbb{R}; Y)}^p. \quad (6.2) \]

Furthermore, for any strip $D \in D_\lambda$, the bound

\[ \left\| 1_D \setminus \bigcup_{D \in D_\lambda} \text{Emb}_\varphi[f] \right\|_{L_k^q Y} \lesssim \lambda s_D^{1/q} \quad (6.3) \]
must hold. This is a localised version of a bound for the embedding $\text{Emb}_\varphi$ into the non-iterated outer Lebesgue space $L^q_\mu \mathcal{S}_\Theta$. Our proof begins with the non-localised, non-iterated embedding bound.

**Theorem 6.1.** Let $\Theta$ and $\Theta_{in}$ be bounded intervals with $0 \in \Theta_{in} \subset \Theta$, and suppose $\varphi \in \mathcal{S}(\mathbb{R})$ has Fourier support in $\Theta_{in}$. Let $Y$ be an $r_0$-intermediate UMD space for some $r_0 \in [2, \infty)$. Then for all $q \in (r_0, \infty)$,

$$
\|\text{Emb}_\varphi[f]\|_{L^q_\mu \mathcal{S}_\Theta} \lesssim \|f\|_{L^q(\mathbb{R};Y)} \quad \forall f \in \mathcal{S}(\mathbb{R};Y). \tag{6.4}
$$

The proof of Theorem 6.1 contains most of the classical time-frequency arguments, in particular the tree selection argument. The disadvantage of the bound (6.4) is that it only holds for the restricted range of exponents $q \in (r_0, \infty)$, which is only strong enough to prove Theorem 1.1 for $p > r_0$. Further restrictions on $p$ would arise since the bounds in terms of non-iterated outer Lebesgue quasi-norms for the dualising functions $A^\pm_\varphi[g]$ are also weaker.

Once we have Theorem 6.1, whose proof we will soon sketch, the iterated embedding bounds of Theorem 3.2 are established via a localisation argument. For every $\lambda \in \mathbb{R}^+$, we need to construct the aforementioned collection of strips $D_\lambda$. This selection is straightforward: one finds a collection of open intervals $B_{\lambda}$ such that

$$
\{x \in \mathbb{R} : (M_p \|f\|_Y)(x) > \lambda\} = \bigcup_{B_{\lambda}(x) \in B_{\lambda}} B_\lambda(x),
$$

and then defines

$$
D = \{D_{\lambda}(x) \in D : B_\lambda(x) \in B_{\lambda}\}.
$$

The measure bound (6.2) then follows directly by the weak-type $(p, p)$ of the Hardy–Littlewood $p$-maximal operator. Proving the localised outer $L^q$ bounds (6.3) involves a sequence of localisation lemmata [3, Lemmata 7.2–7.4], which we will not explicitly state here. These localisation statements essentially state that on the complement of $\bigcup_{D \in D_\lambda} D$, the function $\text{Emb}_\varphi[f]$ only depends on the ‘good’ part of $f$ obtained from a modulation-independent variant of a Calderón–Zygmund decomposition; this ‘good’ part behaves as a function which is in $L^q$ for large $q$, allowing the use of the non-iterated bound (6.4).

Thus the problem is reduced to proving Theorem 6.1, whose proof we now sketch. First, by a size domination principle similar to that used in Theorem 3.6, we show that

$$
\|\text{Emb}_\varphi[f]\|_{L^q_\mu \mathcal{S}_\Theta} \lesssim \|\text{Emb}_\varphi[f]\|_{L^q_\mu \mathcal{R}_\Theta}
$$

for all $q \in [1, \infty]$. Thus it suffices to prove embedding bounds with respect to the randomised local size $\mathcal{R}_\Theta$ rather than the full local size $\mathcal{S}_\Theta = \mathcal{R}_\Theta + \mathcal{L}_\infty \Theta$. The details of this part of the argument are found in [3, Proposition 5.2] (the abstract size domination principle) and [3, Lemma 5.5] (controlling $\mathcal{L}_\infty \Theta$ by $\mathcal{R}_\Theta$). Thus we are left with proving

$$
\|\text{Emb}_\varphi[f]\|_{L^q_\mathcal{R}_\Theta} \lesssim \|f\|_{L^q(\mathbb{R};Y)}
$$
for \( q \in (r_0, \infty) \). As before, by interpolation this reduces to the endpoint bounds

\[
\| \Emb_\varphi [f] \|_{L_\infty^\infty \Theta} \lesssim \| f \|_{L_\infty^\infty (\mathbb{R}; Y)}
\]

(6.5)

and

\[
\| \Emb_\varphi [f] \|_{L_{r_0, \infty} \Theta} \lesssim \| f \|_{L_{r_0}^0 (\mathbb{R}; Y)}.
\]

(6.6)

By unpacking the definitions, the estimate (6.5) is equivalent to proving

\[
\| \Emb_\varphi [f] \|_{\mathcal{R}_\Theta (T)} \lesssim \| f \|_{L_\infty^\infty (\mathbb{R}; Y)}
\]

(6.7)

for all trees \( T \in \mathcal{T}_\Theta \). By our definition of \( \mathcal{R}_\Theta (T) \) as a \( \gamma \)-norm over \( T^{\text{out}} \), the left-hand size of (6.7) can be seen as the modulation of a Littlewood–Paley square function of a localisation of \( f \), plus a tail term. Using the UMD property to handle the Littlewood–Paley square function, and rapid decay of \( \varphi \) to control the tail term, the estimate (6.7) can be established. This is done in [3, Proposition 6.2].

The weak endpoint (6.6) is harder, and this is where a tree selection argument is used. Full details are given in [3, Proposition 6.3]; here we provide only a short overview. We need to show that for all \( \tau > 0 \), there exists an at-most-countable collection of trees \( \mathcal{T}_\tau \) such that

\[
\tau^{r_0} \sum_{T \in \mathcal{T}_\tau} s_T \lesssim \| f \|_{L_0^0 (\mathbb{R}; Y)},
\]

(6.8)

with

\[
\| \mathbbm{1}_{\mathbb{R}^3 \setminus \bigcup_{T \in \mathcal{T}_\tau} T} \Emb_\varphi [f] \|_{\mathcal{R}_\Theta} \lesssim \tau.
\]

(6.9)

The collection of trees is constructed by a greedy algorithm, forcing the estimate (6.9) to hold. When \( Y \) is a Hilbert space, the estimate (6.8) (with \( r_0 = 2 \)) follows by an orthogonality argument as in [17].\(^\dagger\) For general, \( r_0 \)-intermediate UMD spaces \( Y \equiv [X, H]_{2/r_0} \), Hilbert space orthogonality is replaced by a ‘tree orthogonality’ estimate, deduced in [3, Theorem 6.5] by an interpolation argument using orthogonality at the Hilbert endpoint \( H \) and Littlewood–Paley estimates for single trees at the UMD endpoint \( X \).\(^\ddagger\) Thus Theorem 6.1 is proven, leading to Theorem 3.2 by the localisation argument described earlier.

Remark 6.2. In [3], we use a more intricate version of the wave packet embedding and more complicated local sizes. The embedding maps \( \Emb \) defined in [3, §3.2] do not depend on a wave packet \( \varphi \): rather, they are valued in a space of linear maps \( \mathcal{L}(\Phi, Y) \), where \( \Phi \) is a Fréchet space of wave packets, which we think of as generalised \( Y \)-valued distributions. The \( \varphi \)-dependent embedding maps \( \Emb_\varphi \) are recovered through the identity \( \Emb_\varphi [f] [\varphi] = \Emb [f] [\varphi] \), provided that \( \varphi \in \Phi \).

\(^\dagger\) We do not reproduce the argument in [3], and it is not explicitly written in [17] for Hilbert spaces, but since the proof only relies on orthogonality properties, the only necessary changes to the proof of [17] are notational.

\(^\ddagger\) This is analogous to the tile-type estimates used in [24, 25]. We have not investigated further connections between ‘tree orthogonality’ and tile-type, but this would be an interesting line of research.
Furthermore, the local sizes used in [3] involve extra terms (defect local sizes), as well as pointwise suprema over $\Phi$. These additional embellishments capture higher multilinear phenomena which do not arise in the study of the (linearised) variational Carleson operator, and which create substantial difficulty in proving the outer Hölder inequality.†

In proving bounds for $\text{Emb}$ into outer Lebesgue spaces [3, Theorem 7.1], which are linear rather than multilinear, these aforementioned embellishments can be removed: the size domination arguments of [3, § 5] allow for the removal of the extra terms in the local sizes, and [3, Proposition 3.16] lets the supremum over $\Phi$ be replaced by the action of a single wave packet $\varphi$. Since the embedding maps $\text{Emb}$ are strictly larger than the embedding maps $\text{Emb}_\varphi$ discussed here (when measured with respect to the appropriate local sizes), the outer Lebesgue space estimates for $\text{Emb}_\varphi$ used in this paper are a direct consequence of [3, Theorem 7.1].‡

ACKNOWLEDGEMENTS
We thank Francesco Di Plinio, Christoph Thiele, and Mark Veraar for their encouragement throughout this project. We also thank Emiel Lorist and Zoe Nieraeth for useful comments on Section 5, and the anonymous referee for their careful reading and suggestions. The first author was supported by a Fellowship for Postdoctoral Researchers from the Alexander von Humboldt Foundation.

JOURNAL INFORMATION
The Journal of the London Mathematical Society is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

REFERENCES
1. A. Amenta, E. Lorist, Mark Veraar, et al., Rescaled extrapolation for vector-valued functions, Publ. Mat. 63 (2019), no. 1, 155–182.
2. A. Amenta and G. Uraltsev, Banach-valued modulation invariant carleson embeddings and outer-$L^p$ spaces: The walsh case, J. Fourier Anal. Appl. 26 (2020), no. 4, 1–54.
3. A. Amenta and G. Uraltsev, The bilinear Hilbert transform in UMD spaces, Math. Ann. 378 (2020), no. 3, 1129–1221.
4. C. Benea and C. Muscalu, Sparse domination via the helicoidal method. Rev. Mat. Iberoam. 37 (2021), no. 6, 2037–2118.
5. J. Bergh and J. Löfström, Interpolation spaces: an introduction, vol. 223, Springer Science & Business Media, Berlin, 2012.
6. J. Bourgain, Some remarks on Banach spaces in which martingale difference sequences are unconditional, Ark. Mat. 21 (1983), no. 1, 163–168.
7. D. L. Burkholder, A geometric condition that implies the existence of certain singular integrals of Banach-space-valued functions, Conference on harmonic analysis in honor of Antoni Zygmund, vol. 1, 1983, pp. 270–286.
8. D. L. Burkholder, Martingales and singular integrals in Banach spaces, Handbook of the geometry of Banach spaces, Vol. I, North-Holland, Amsterdam, 2001, pp. 233–269.

†In this paper, the corresponding result is Proposition 2.32, which was reasonably straightforward to establish; the difficulties come from having three factors instead of two.
‡To be really pedantic, the hypotheses of that theorem implicitly require that the inner frequency interval $\Theta_{in}$ is symmetric and that $B_1 \subseteq \Theta$, but the same proof can be carried out for general bounded intervals $\Theta_{in}$ and $\Theta$ provided that $0 \in \Theta_{in} \subseteq \Theta$. Only the notation changes.
9. J. T. Campbell et al., *Oscillation and variation for the Hilbert transform*, Duke Math. J. **105** (2000), no. 1, 59–83.

10. L. Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 135–157.

11. J. L. R. de Francia, *Martingale and integral transforms of Banach space valued functions*, Probability and Banach spaces (Zaragoza, 1985), Lecture Notes in Mathematics, vol. 1221, Springer, Berlin, 1986, pp. 195–222.

12. C. Demeter, *A guide to Carleson’s theorem*, Rocky Mountain J. Math. **45** (2015), no. 1, 169–212.

13. F. Di Plinio, Y. Q. Do, and G. N. Ural’tsev, *Positive sparse domination of variational Carleson operators*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **XVII** (2018), no. 4, 1443–1458.

14. F. Di Plinio, L. Kangwei, M. Henri, and V. Emil, *Banach-Valued Multilinear Singular Integrals with Modulation Invariance*, Int. Math. Res. Not. (2020). https://doi.org/10.1093/imrn/rnaa234.

15. Y. Do and M. Lacey, *Weighted bounds for variational Fourier series*, Stud. Math. **211** (2012), no. 2, 153–190.

16. Y. Do and M. Lacey, *Weighted bounds for variational Walsh-Fourier series*, J. Fourier Anal. Appl. **18** (2012), no. 6, 1318–1339.

17. Y. Do and C. Thiele, *$L^p$ theory for outer measures and two themes of Lennart Carleson united*, Bull. Amer. Math. Soc. (N.S.) **52** (2015), no. 2, 249–296.

18. C. Fefferman, *Pointwise convergence of Fourier series*, Ann. of Math. (2) **98** (1973), 551–571.

19. L. Grafakos, *Classical Fourier analysis*, 3rd edn, Graduate Texts in Mathematics, vol. 249, Springer, Berlin, 2014, pp. xvii + 638.

20. G. Hong, W. Liu, and T. Ma, *Vector-valued $q$-variational inequalities for averaging operators and Hilbert transform*, Preprint, arXiv:1907.11633, 2019.

21. G. Hong, W. Liu, and T. Ma, *Vector-valued $q$-variational inequalities for averaging operators and Hilbert transform*, Arch. Math. **115** (2020), 423–433.

22. G. Hong and T. Ma, *Vector valued $q$-variation for differential operators and semigroups I*, Math. Z. **286** (2017), no. 1-2, 89–120.

23. R. A. Hunt, *On the convergence of Fourier series*, Orthogonal Expansions and their Continuous Analogues (Proc. Conf., Edwardsville, Ill., 1967), Southern Illinois Univ. Press, Carbondale, Ill., 1968, pp. 235–255.

24. T. P. Hytönen and M. T. Lacey, *Pointwise convergence of vector-valued Fourier series*, Math. Ann. **357** (2013), no. 4, 1329–1361.

25. T. P. Hytönen et al., *Analysis in Banach spaces. Vol. I: Martingales and Littlewood-Paley theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge., vol. 63, Springer, Cham, 2016, pp. xvi + 614.

26. T. P. Hytönen et al., *Analysis in Banach spaces. Vol. II: Probabilistic methods and operator theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge., vol. 67, Springer, Cham, 2017, pp. xxi + 616.

27. M. Lacey and C. Thiele, *A proof of boundedness of the Carleson operator*, Math. Res. Lett. **7** (2000), no. 4, 361–370.

28. J. Lindenstrauss and L. Tzafriri, *Classical Banach spaces. II*, Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas], Function spaces, vol. 97, Springer, Berlin–New York, 1979, pp. x+243.

29. E. Lorist and Z. Nieraeth, *Sparse domination implies vector-valued sparse domination*, Preprint, arXiv:2003.02233, 2020.

30. R. Oberlin et al., *A variation norm Carleson theorem*, J. Eur. Math. Soc. **14** (2012), no. 2, 421–464.

31. J. Parcet, F. Soria, and Q. Xu, *On the growth of vector-valued Fourier series*, Indiana Univ. Math. J. **62** (2013), no. 6, 1765–1784.

32. G. Pisier, *Martingales in Banach spaces*, Cambridge Studies in Advanced Mathematics, vol. 155, Cambridge Univ. Press, Cambridge, 2016, pp. xxviii + 561.

33. G. Pisier and Q. Xu, *The strong $p$-variation of martingales and orthogonal series*, Probab. Theory Related Fields **77** (1988), no. 4, 497–514.

34. C. Thiele, *Wave packet analysis*, CBMS Regional Conference Series in Mathematics, vol. 105, Amer. Math. Soc., Providence, R.I., 2006, pp. vi + 86.

35. G. Ural’tsev, *Variational Carleson embeddings into the upper 3-space*, arXiv:1610.07657, 2016.