SPANNING AND SAMPLING IN LEBESGUE AND SOBOLEV SPACES

H.-Q. BUI AND R. S. LAUGESEN

ABSTRACT. We establish conditions on ψ under which the small-scale affine system $\{\psi(a_j x - k) : j \geq J, k \in \mathbb{Z}^d\}$ spans the Lebesgue space $L^p(\mathbb{R}^d)$ and the Sobolev space $W^{m,p}(\mathbb{R}^d)$, for $1 \leq p < \infty$ and $J \in \mathbb{Z}$. The dilation matrices a_j are expanding (meaning $\lim_{j\to\infty} \|a_j^{-1}\| = 0$) but they need not be diagonal.

For spanning L^p our result assumes $\int_{\mathbb{R}^d} \psi \, dx \neq 0$ and, when $p > 1$, that the periodization of $|\psi|$ or of $\mathbb{1}_{\{\psi \neq 0\}}$ is bounded. But the periodization of ψ need not be constant; in other words, the functions $\{\psi(x-k) : k \in \mathbb{Z}^d\}$ need not form a partition of unity like B-splines do. For spanning $W^{m,p}$ we impose the Strang–Fix condition on ψ, but only to order $m-1$ whereas earlier authors required order m.

These spanning results follow from explicitly approximating an arbitrary function f by linear combinations of the $\psi(a_j x - k)$, with the coefficients being local averages of f.

1. Introduction

1.1. **Spanning by sampling.** Under what conditions on $\psi(x)$ will the *small-scale dyadic affine system* $\{\psi(2^j x - k) : j \geq J, k \in \mathbb{Z}^d\}$ span the Lebesgue space $L^p = L^p(\mathbb{R}^d)$ or the Sobolev space $W^{m,p} = W^{m,p}(\mathbb{R}^d)$, when $1 \leq p < \infty$ and $J \in \mathbb{Z}$ is given? We normalize $\int_{\mathbb{R}^d} \psi \, dx = 1$ and investigate this spanning question by means of sampling formulas that show how to write an arbitrary function as a limit of linear combinations of translates and dilates of the single function ψ.

For example in dimension $d = 1$ we prove for $\psi \in L^1(\mathbb{R})$ with bounded variation that the small-scale affine system spans $L^p(\mathbb{R})$ for all $1 \leq p < \infty$; see the Remarks on Theorem 1. This spanning result follows from sampling formula (56) after Corollary 11, which says in the current situation that

$$f(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=J+1}^{J+N} \sum_{k\in\mathbb{Z}} f(2^{-j}k)\psi(2^j x - k) \qquad \text{in } L^p(\mathbb{R}) \tag{1}$$

whenever $f \in L^p(\mathbb{R})$ has bounded variation and is left or right continuous. Notice ψ and f need not have compact support. The limit in (1) holds pointwise a.e., if ψ additionally has a radially decreasing L^1-majorant.

The distinctive feature of formula (1) is its averaging over dilation or frequency scales, from $j = J+1$ to $j = J+N$. Without such averaging the formula holds only in the special case where the functions $\{\psi(x-k) : k \in \mathbb{Z}\}$ form a partition of unity, in other words where ψ has constant periodization 1.

Date: August 16, 2004; revised version.

Laugesen was supported by N.S.F. Award DMS–0140481 and a Maclaurin Fellowship from the New Zealand Institute of Mathematics and its Applications. *2000 AMS Classification:* Primary 42C30, 41A25; Secondary 41A15, 42C40, 46E35. *Key words:* spanning, completeness, density, sampling, Strang–Fix, approximation order, quasi-interpolation, approximate approximations.

More general than (1) is sampling formula (47), which says for arbitrary $f \in L^p$ that

$$f(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \sum_{k\in\mathbb{Z}^d} c_{n,k}\psi(a_{j(n)}x - k) \qquad \text{in } L^p(\mathbb{R}^d);$$

here ψ need not have bounded variation, the dilations a_j are matrices that *expand* in a sense defined later, and $j(n)$ is an increasing integer sequence that depends on ψ. The coefficients $c_{n,k} = \int_{\mathbb{R}^d} f(a_{j(n)}^{-1}y)\phi(y-k)\,dy$ are sampled averages of f. Pointwise sampling would not make sense here because f is defined only up to sets of measure zero. The function ϕ is the analyzer and ψ is the synthesizer.

We would be delighted if practical applications were found for these sampling formulas, to complement their evident theoretical usefulness. One should seek an application that naturally specifies a generating function ψ with nonconstant periodization (the constant periodization case is well understood already).

Notes.

1. The Shannon–Whittaker sampling theorem [17, §2.1] says $f(x) = \sum_{k\in\mathbb{Z}} f(2^{-j}k)\,\text{sinc}(2^j x - k)$ for all sufficiently large j, for each band limited f. Our sampling formulas provide an extension of Shannon sampling to more-or-less arbitrary functions f and ψ, in particular with no band limitation on $\text{spt}(\hat{f})$.

2. The sampling throughout this paper differs qualitatively from wavelet sampling, where one samples the deviation of f from its local mean value (with $\int_{\mathbb{R}^d}\psi\,dx = \widehat{\psi}(0) = 0$ and $\phi = \psi$). Our sampling fits rather into the quasi-interpolation tradition. And note our spanning results aim only for spanning, not for additional structure such as orthonormal bases or frames.

Overview of the paper. Section 2 specifies standing assumptions on the dilation and translation matrices. These dilation matrices a_j need not be diagonal, or dyadic.

Section 3 describes our *spanning* results for Lebesgue and Sobolev spaces, with the two Corollaries being particularly concrete. Corresponding *sampling* formulas are developed later, in Sections 6 and 7. In particular Section 6.3 obtains sampling formulas when ψ is a Schwartz function such as a Gaussian, or more generally when ψ has bounded variation. Section 7.4 proves a rate of approximation to the sampled function.

Sections 4 and 5 lay the technical groundwork for all sampling formulas in the paper. Appendices A and B collect basic facts on periodizations and on the local supremum operator Q that we use throughout the paper. Appendix C proves a Riemann–Lebesgue limit, and Appendix D develops measure theoretic results on norm and pointwise convergence of arithmetic means of sequences of functions (following Banach–Saks, Szlenk and Komlós).

Open problems are discussed in Section 3.4, especially the Mexican hat spanning problem of Meyer which has motivated our current work.

A briefer account of our methods and results can be found in [12], where we restrict attention to L^1 and L^2 in one dimension.

2. Standing assumptions, and some definitions

2.1. Standing assumptions.

1. Given an exponent $1 \le p \le \infty$ we define the conjugate exponent q by $\frac{1}{p} + \frac{1}{q} = 1$.

2. Fix the *dimension* $d \in \mathbb{N}$ and write $\mathcal{C} = [0,1)^d$ for the unit cube in $\mathbb{R}^d$. Let $L^p = L^p(\mathbb{R}^d)$ and $W^{m,p} = W^{m,p}(\mathbb{R}^d)$.

3. Let the $\boxed{\text{dilation matrices } a_j}$ for $j \in \mathbb{Z}$ be invertible $d \times d$ real matrices that are *expanding*, in the sense that

$$|a_j x| \geq \lambda_j |x| \qquad \text{for all } x \in \mathbb{R}^d, \quad j \in \mathbb{Z}, \tag{2}$$

for some constants λ_j that approach infinity as $j \to \infty$. That is

$$\|a_j^{-1}\| \to 0 \qquad \text{as } j \to \infty, \tag{3}$$

where $\|a\|$ denotes the norm of a matrix a as an operator from the column vector space $\mathbb{R}^d$ to itself.

In one dimension, real numbers $a_j \neq 0$ are expanding if and only if $|a_j| \to \infty$ as $j \to \infty$. In all dimensions, if $a_j = M^j$ for some invertible matrix M whose eigenvalues all have magnitude greater than 1 then the a_j are expanding (see for example [15, Remark 2.2]).

4. Fix a $\boxed{\text{translation matrix } b,}$ again assumed to be an invertible $d \times d$ real matrix. Many of our results, operators and constants will depend implicitly on b and on the dimension d.

5. Given a function $\psi(x)$ on $\mathbb{R}^d$, define

$$\boxed{\psi_{j,k}(x) = \psi(a_j x - bk)}, \qquad j \in \mathbb{Z}, \quad k \in \mathbb{Z}^d, \quad x \in \mathbb{R}^d.$$

(The traditional dyadic $\psi_{j,k}(x) = \psi(2^j x - k)$ comes from choosing $a_j = 2^j I$ and $b = I$.) Clearly if $\psi \in W^{m,p}$ then $\psi_{j,k} \in W^{m,p}$. Notice we do not normalize $\psi_{j,k}$ with any prefactor.

2.2. **Some definitions.**

1. A subset U of a topological vector space V over the complex numbers is said to *span* V if the finite linear combinations of elements of U form a dense subset of V, that is if V is equal to

$$V\text{-span}(U) = \text{closure in } V \text{ of } \{\sum_{m=1}^{n} c_m u_m : c_m \in \mathbb{C}, u_m \in U, n \in \mathbb{N}\}.$$

2. Write $BV = BV(\mathbb{R}^d)$ for the class of complex-valued functions whose real and imaginary parts have *bounded variation* on $\mathbb{R}^d$, as defined in [18, Chapter 5] or [51, Chapter 5]. (Except we do not assume BV functions belong to L^1, just to L^1_{loc}: for example $1 + e^{-|x|}$ belongs to BV. Whenever global L^1 integrability is required, we will state it explicitly.)

In one dimension, a BV function can be redefined on a set of measure zero to satisfy the usual classical definition [21, §3.5]. In all dimensions, Sobolev functions have bounded variation: $W^{1,1} \subset BV$ by [18, p. 170].

3. Some of our results (notably on Sobolev sampling) will require the dilation matrices a_j to *expand nicely*, meaning the a_j are expanding and

$$|\det a_j| \leq C\lambda_j^d \tag{4}$$

for all $j \in \mathbb{Z}$ and some constant $C > 0$. Geometrically, this means the volume of the image of the unit ball under a_j is bounded by a multiple of the volume of the largest ball inscribed in that image and centered at the origin. Note if the a_j expand nicely then

$$\|a_j^{-1}\| \|a_j\| \leq C \qquad \text{for all } j \in \mathbb{Z} \tag{5}$$

by the inequality "$H \leq H_I$" in [47, eq. (14.3)].

In one dimension, every sequence a_j of nonzero real numbers with $|a_j| \to \infty$ expands nicely, because (4) just says $|a_j| \le C|a_j|$ when $d = 1$.

4. Write $X(x) = x$ for the identity function on $\mathbb{R}^d$, and let $\chi_r(x) = 1 + |x|^r$ when $r \ge 0$.

5. When μ is a multiindex, write $f^{(\mu)} = D^\mu f$ for the μ-th derivative of f.

6. Define the Fourier transform using 2π in the exponent: $\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i \xi x}\, dx$, where ξ = row vector $\in \mathbb{R}^d$.

3. **Spanning results**

3.1. **Spanning L^p.** Our first theorem gives sufficient conditions for a function ψ to generate a small-scale affine system spanning L^p. In the theorem, we write

$$\mathbb{1}_{\{\psi \neq 0\}}(x) = \begin{cases} 1 & \text{if } \psi(x) \neq 0 \\ 0 & \text{if } \psi(x) = 0 \end{cases}$$

for the characteristic function of the set where ψ is nonzero. We also need the operators P and Q. Define the *periodization* operator P by

$$\boxed{(Pf)(x) = |\det b| \textstyle\sum_{k \in \mathbb{Z}^d} f(x - bk)} \qquad \text{for } x \in \mathbb{R}^d.$$

If $f : \mathbb{R}^d \to [0, \infty]$ is measurable then $Pf : \mathbb{R}^d \to [0, \infty]$ is well defined and also measurable, and of course Pf is periodic with respect to the lattice $b\mathbb{Z}^d$. Also define a *local supremum* operator Q by

$$\boxed{(Qf)(x) = \text{ess. sup}_{|y-x|<\sqrt{d}} |f(y)| = \|f\|_{L^\infty(B(x,\sqrt{d}))}} \qquad \text{for } x \in \mathbb{R}^d,$$

when f is measurable. Then $Qf : \mathbb{R}^d \to [0, \infty]$ is lower semicontinuous and hence measurable.

Theorem 1. *Let $J \in \mathbb{Z}$ and write $A_J(\psi) = \{\psi_{j,k} : j \ge J, k \in \mathbb{Z}^d\}$.*

(a) Let $p = 1$ and suppose $\psi \in L^1$. If $\int_{\mathbb{R}^d} \psi\, dx \neq 0$ then $A_J(\psi)$ spans L^1.

(b) Let $1 < p < \infty$. Suppose $\psi \in L^p$ and that either $P(\mathbb{1}_{\{\psi \neq 0\}}) \in L^\infty$ or $P(|\psi|) \in L^\infty$. If $\int_{\mathbb{R}^d} \psi\, dx \neq 0$ then $A_J(\psi)$ spans L^p.

More generally if $\psi = \psi_0 + \psi_1$ where $\psi_0, \psi_1 \in L^p$ and $P(\mathbb{1}_{\{\psi_0 \neq 0\}}) \in L^\infty$ and $P(|\psi_1|) \in L^\infty$, with $\int_{\mathbb{R}^d} \psi\, dx \neq 0$, then $A_J(\psi)$ spans L^p.

(c) Let $p = \infty$ and suppose $\psi \in L^\infty$ with $Q\psi \in L^1$, and that $P\psi$ is constant a.e. If $\int_{\mathbb{R}^d} \psi\, dx \neq 0$ then the L^∞-span$(A_J(\psi))$ contains $UC \cap L^1$.

The class UC consists of all uniformly continuous functions on $\mathbb{R}^d$. It is easy to show $UC \cap L^1 \subset L^\infty$, and that functions in $UC \cap L^1$ must vanish at infinity.

Corollary 2. *Let $\psi \in L^p, 1 \le p < \infty$, and when $1 < p < \infty$ assume ψ decays according to $|\psi(x)| \le C|x|^{-d-\epsilon}$ for all large $|x|$, for some constants $C, \epsilon > 0$.*

If $\int_{\mathbb{R}^d} \psi\, dx \neq 0$ then $A_J(\psi)$ spans L^p for all $J \in \mathbb{Z}$.

Filippov and Oswald [19, Theorem 3] proved the corollary in the special case of dyadic dilations, $a_j = 2^j I$. See our remarks on the literature, below.

We prove the theorem and corollary at the end of the section.

Remarks on Theorem 1 and Corollary 2.

1. The hypothesis $P\mathbb{1}_{\{\psi \neq 0\}} \in L^\infty$ in the theorem is a kind of "finite intersection" property of the support of ψ, and it certainly holds whenever ψ has compact support. But it might

hold when ψ has noncompact support, as shown by the counterexample in Appendix B, which is the characteristic function of a certain unbounded set. Theorem 1 is stronger than Corollary 2, when $p > 1$, because ψ does not decay at infinity in this counterexample.

2. The other hypothesis $P|\psi| \in L^\infty$ is certainly satisfied if ψ is bounded with compact support, or if ψ is Schwartz class, such as a Gaussian. More generally, if $\psi \in L^\infty$ has a radially decreasing L^1-majorant then $P|\psi| \in L^\infty$ (Lemma 19).

And in one dimension the hypothesis $P|\psi| \in L^\infty(\mathbb{R})$ is assured whenever $\psi \in L^1(\mathbb{R})$ has bounded variation (for example when $\psi \in W^{1,1}(\mathbb{R})$), because then $P|\psi| \in BV_{loc}(\mathbb{R})$ by Lemma 20 and $BV_{loc} \subset L^\infty_{loc}$ in one dimension, so that $P|\psi| \in L^\infty(\mathbb{R})$ by periodicity.

3. $A_J(\psi)$ will generally not span L^∞. For example, if ψ has compact support then finite linear combinations of the $\psi_{j,k}$ cannot well approximate the constant function $f \equiv 1$ in the L^∞-norm. For another example (albeit one where $P\psi$ is nonconstant), if ψ is continuous with compact support and ψ happens to vanish at every point bk for $k \in \mathbb{Z}^d$, then $\psi_{j,k}(0) = 0$ for all j, k. Thus every (continuous) function in the uniform closure of the span of the $\psi_{j,k}$ must vanish at $x = 0$, preventing $A_J(\psi)$ from spanning all of L^∞.

Remarks on the L^p spanning literature. The best previous L^p spanning result with $\int_{\mathbb{R}^d} \psi \, dx \neq 0$ was proved by Filippov and Oswald [19, Theorem 3]. They obtained precisely Corollary 2 in the special case of dyadic dilations $a_j = 2^j I$. Their methods are very different from ours and provide no sampling formulas. Also, their methods are highly dyadic and so it is not clear whether they could handle more general dilations.

The next best previous spanning results for L^p are of "Strang–Fix" type. These make three assumptions: that ψ has *constant* periodization $P\psi = \text{const} \neq 0$, that ψ has *compact support*, and that the dilations are expanding and *isotropic* ($a_j = \lambda_j I$). See [6, Theorem 4.1] for $p = 2$, and [44, Theorems I,III] for $p = 2, \infty$, and [16] for $1 \leq p \leq \infty$ (with refinements in [10]). A few papers relax the compact support assumption to a polynomial decay condition on $\psi \in L^\infty$: see [31, Theorem 3.1] and [24] for $p = \infty$, and [27, Theorem 1.1] for $1 \leq p \leq \infty$. In fact di Guglielmo [23, Théorème 2′] had the first results in this area, handling $p \geq 2$ with nonisotropic (though still diagonal) dilation matrices and with ψ having compact support. But di Guglielmo required ψ to have a special convolution form that is strictly stronger than having constant periodization, when $d > 1$, as we discuss in Section 3.2. (An exception is [23, Théorème 5], which assumes constant periodization, $p = 2$, and a vanishing moment condition on ϕ and ψ, and then proves a sampling formula.)

Theorem 1 improves on this Strang–Fix literature in every respect for $p < \infty$: instead of constant periodization it assumes only *bounded* periodization ($P|\psi| \in L^\infty$ or $P\mathbb{1}_{\{\psi \neq 0\}} \in L^\infty$), and it imposes no support or decay conditions on ψ, and the dilations can expand arbitrarily. Matters are even better when $p = 1$, for then ψ need not even have bounded periodization in Theorem 1. And for all p, our ψ need not be bounded, as Corollary 2 demonstrates.

In fairness to the "Strang–Fix" authors, they were not aiming to prove spanning results. They were proving approximation formulas with explicit "big-O" error terms, and spanning results were simply a byproduct. When we establish analogous L^p approximation rates, in Section 7.4, we too will assume ψ has constant periodization (although we will weaken the support and decay conditions of earlier authors, and handle a wider class of dilations).

Clearly $p = \infty$ is a very special case in Theorem 1, and this part of the theorem improves on the literature only in that it relaxes the compact support or decay assumption on ψ to the weaker hypothesis $Q\psi \in L^1$.

A final point in favor of Theorem 1 is that we prove it by a general approximation formula whose coefficients are sampled average values of f. Of all the authors in the Strang-Fix tradition mentioned above, only di Guglielmo gave such explicit sampling formulas, when $2 \leq p < \infty$; and di Guglielmo required ψ to have a special convolution form. All other authors used coefficients depending on the Fourier transform $\hat{f}$, when $1 \leq p < \infty$. See our further remarks on sampling formulas at the end of subsection 6.1.

In a different vein from Theorem 1, for $p = 2$ the class of $\psi \in L^2$ such that the $\psi_{j,k}$ span L^2 by approximations of the specific form $\lim_{j\to\infty} \sum_{k\in\mathbb{Z}^d} c_{j,k}\psi(\lambda_j x - k)$ ("only one j at a time") has been fully characterized without decay assumptions, in a delightful work by de Boor, DeVore and Ron [9]. When $\int_{\mathbb{R}^d} \psi \, dx \neq 0$ this characterization reduces to the constant periodization condition again, but the characterization is new when $\int_{\mathbb{R}^d} \psi \, dx = 0$. Section 1 of [9] gives a wide survey of the literature up to 1993, mentioning a number of papers treating noncompact support. See also the later paper [37], for results on $L^p, 2 \leq p \leq \infty$.

Working still with $p = 2$, the wavelet literature offers two spanning results with nonconstructive proofs, for dyadic dilations. First, if $\psi \in L^2(\mathbb{R})$ is a scaling function and $|\widehat{\psi}|$ is continuous at the origin with $\widehat{\psi}(0) \neq 0$, then $A(\psi)$ spans $L^2(\mathbb{R})$. See [26, Theorem 2.1.7] and references therein. Of course for ψ to satisfy a scaling relation is a significant restriction. And if it happens that $\psi \in L^1(\mathbb{R})$ then $P\psi$ is constant (see [50, Proposition 2.17]), and so $A_J(\psi)$ spans L^2 anyway by Strang–Fix type results.

The second result says that if $\psi \in L^2$ and $|\widehat{\psi}|$ is continuous at the origin with $\widehat{\psi}(0) \neq 0$, and if $P(|\widehat{\psi}|^2) \in L^\infty$, then $A_J(\psi)$ spans L^2. This is a simple adaptation of Daubechies' spanning result [17, Proposition 5.3.2]. The proof is special to L^2. It is unclear to us what relation there might be between the hypothesis $P(|\widehat{\psi}|^2) \in L^\infty$ here and the hypotheses $P(|\psi|) \in L^\infty$ or $P(\mathbb{1}_{\{\psi \neq 0\}}) \in L^\infty$ in Theorem 1 for $p = 2$.

Contrast with Wiener's Tauberian theorems. The famous Tauberian theorems of Wiener [48] say that the collection of all translates $\{\psi(\cdot - y) : y \in \mathbb{R}^d\}$ spans L^1 if and only if $\psi \in L^1$ and $\widehat{\psi}$ is nonzero everywhere, and spans L^2 if and only if $\psi \in L^2$ and $\widehat{\psi}$ is nonzero a.e. (For some recent sharp constructions in L^p for $p \neq 1, 2$, see [38].)

Our affine spanning results differ from Wiener's theorems because we restrict ourselves to a discrete subset of translations, but allow ourselves also a discrete sequence of dilations. The contrast with Wiener's theorem is particularly stark for spanning L^1, because if $\widehat{\psi}$ is nonzero merely on a neighborhood of the origin then $A_J(\psi)$ spans L^1 by Theorem 1(a).

Incidentally, it is possible to span L^p for $2 < p < \infty$ using just integer translates of a single function ψ (and no dilates at all). The function ψ must be rather special. See [4] and references therein. One can further generalize this problem to considering arbitrary discrete sets of translates (not just $k \in \mathbb{Z}^d$), and the recent paper [11] completely characterizes when L^1 can be spanned this way.

Now we can prove the theorem and corollary for spanning L^p.

Proof of Theorem 1. The hypotheses ensure $\psi \in L^1$, in all three parts of the theorem. This is immediate when $p = 1$. When $1 < p < \infty$, if $P\mathbb{1}_{\{\psi \neq 0\}} \in L^\infty$ then the set $\{x : \psi(x) \neq 0\}$ has finite measure and so $\psi \in L^p$ implies $\psi \in L^1$ by Hölder's inequality, while if $P|\psi| \in L^\infty$ then $\psi \in L^1$ just by integrating $P|\psi|$ over the set $b\mathcal{C}$. When $p = \infty$ we assume $Q\psi \in L^1$ in part (c), and $|\psi| \leq Q\psi$ by Lemma 22 so that $\psi \in L^1$.

Since $\psi \in L^1$ with $\int_{\mathbb{R}^d} \psi \, dx \neq 0$, we can normalize $\int_{\mathbb{R}^d} \psi \, dx = 1$. Define $\phi = \mathbb{1}_{\mathcal{C}}$.

To prove part (a) and the first paragraph of (b), take an arbitrary $f \in L^p$. Then by combining Proposition 9(d) with sampling formula (47) in Proposition 9(b) (using case (i) and $\delta = 1$ in the proposition), we conclude $f \in L^p\text{-span}(A_J(\psi))$.

Now we sketch a proof for the second paragraph of (b). We may assume $\int_{\mathbb{R}} \psi_0 \, dx = 1/2$ and $\int_{\mathbb{R}} \psi_1 \, dx = 1/2$, by adding a suitable smooth, compactly supported function to ψ_0 and subtracting it from ψ_1. Then the two functions $2\psi_0$ and $2\psi_1$ satisfy the hypotheses of case (i) of Proposition 9, with $\varepsilon = 0$ and $\varepsilon = 1$ respectively. Applying that proposition to the two functions yields two sampling formulas, one based on $2\psi_0$ and one on $2\psi_1$, with both converging to f. We choose the sequence $j(n)$ to work for both $2\psi_0$ and $2\psi_1$ simultaneously (see the remark after Proposition 9).

After adding the two sampling formulas and dividing by 2 we obtain a new sampling formula that also converges to f. Since adding $2\psi_0(a_j x - k)$ to $2\psi_1(a_j x - k)$ and dividing by 2 gives exactly $\psi(a_j x - k)$, we conclude that f lies in the L^2-span of $A_J(\psi)$.

For part (c) of this theorem, where $p = \infty$, simply take $f \in UC \cap L^1 \subset L^\infty$ and use case (iii) of Proposition 9(a)(d). Note the constant function $P\psi$ must equal its mean value a.e., and this mean value equals

$$\frac{1}{|b\mathcal{C}|} \int_{b\mathcal{C}} (P\psi)(y) \, dy = \int_{b\mathcal{C}} \sum_{k \in \mathbb{Z}^d} \psi(y - bk) \, dy = \int_{\mathbb{R}^d} \psi(y) \, dy = 1. \tag{6}$$

□

Proof of Corollary 2.

When $p = 1$, this is just Theorem 1.

Suppose $1 < p < \infty$ and choose $R > 0$ with $|\psi(x)| \leq C|x|^{-d-\epsilon}$ for all $|x| \geq R$. Define $\psi_0(x) = \psi(x)$ when $|x| < R$ and $\psi_0(x) = 0$ otherwise, and let $\psi_1 = \psi - \psi_0$.

Notice ψ_0 has compact support and so $P\mathbb{1}_{\{\psi_0 \neq 0\}} \in L^\infty$, while $P|\psi_1| \in L^\infty$ by Lemma 19 because ψ_1 has a radially decreasing L^1 majorant of the form $C \max\{R, |x|\}^{-d-\epsilon}$.

Since $\psi = \psi_0 + \psi_1$, Theorem 1(b) tells us that $A_J(\psi)$ spans L^p. □

3.2. **Spanning Sobolev spaces.** We will develop sufficient conditions for spanning the Sobolev space $W^{m,p}$. Recall $A_J(\psi) = \{\psi_{j,k} : j \geq J, k \in \mathbb{Z}^d\}$, that $\chi_r(x) = 1 + |x|^r$, and that $\psi^{(\mu)} = D^\mu \psi$ means the μ-th derivative of ψ, when μ is a multiindex.

Theorem 3. *Assume the dilations a_j expand nicely. Take $J \in \mathbb{Z}, m \in \mathbb{N}, 1 \leq p \leq \infty$ and suppose $\psi \in W^{m,p}$ with*

$$\chi_{|\mu|}\psi^{(\mu)} \in L^p \qquad \text{for all multiindices } \mu \text{ of order } |\mu| \leq m, \text{ and} \tag{7}$$

$$(D^\mu \widehat{\psi})(\ell b^{-1}) = 0 \qquad \text{for all } |\mu| < m \text{ and all row vectors } \ell \in \mathbb{Z}^d \setminus \{0\}. \tag{8}$$

(a) If $p = 1$ and $\int_{\mathbb{R}^d} \psi \, dx \neq 0$ then $A_J(\psi)$ spans $W^{m,1}$.

(b) Let $1 < p < \infty$ and suppose either

$$P(\mathbb{1}_{\{\psi^{(\mu)} \neq 0\}}) \in L^\infty \qquad \text{for all } |\mu| \leq m \text{ or else} \tag{9}$$

$$P(|\chi_{|\mu|}\psi^{(\mu)}|) \in L^\infty \qquad \text{for all } |\mu| \leq m. \tag{10}$$

If $\int_{\mathbb{R}^d} \psi \, dx \neq 0$ then $A_J(\psi)$ spans $W^{m,p}$.

More generally, let $1 < p < \infty$ and suppose the functions $\psi_0, \psi_1 \in W^{m,p}$ each satisfy (7), *that the sum $\psi = \psi_0 + \psi_1$ satisfies* (8), *and that ψ_0 and ψ_1 satisfy* (9) *and* (10) *respectively. If $\int_{\mathbb{R}^d} \psi \, dx \neq 0$ then $A_J(\psi)$ spans $W^{m,p}$.*

(c) Let $p = \infty$ and suppose (8) *holds also for all $|\mu| = m$, and that $Q(\chi_{|\mu|}\psi^{(\mu)}) \in L^1$ for all $|\mu| \leq m$. If $\int_{\mathbb{R}^d} \psi \, dx \neq 0$ then the $W^{m,\infty}$-span$(A_J(\psi))$ contains $W^{m,\infty} \cap UC^m \cap L^1$.*

UC^m denotes the class of functions whose derivatives of order $\leq m$ are all uniformly continuous.

Next we observe a simple decay condition near infinity suffices for ψ to span $W^{m,p}$, in conjunction with the vanishing of the Fourier transform at the lattice points.

Corollary 4. *Assume the dilations a_j expand nicely. Let $m \in \mathbb{N}, 1 \leq p < \infty$, and suppose $\psi \in W^{m,p}$ decays according to*

$$|\psi^{(\mu)}(x)| \leq C|x|^{-d-|\mu|-\epsilon} \qquad \text{for all } |\mu| \leq m \text{ and all large } |x|, \tag{11}$$

for some constants $C, \epsilon > 0$. Suppose $(D^\mu \widehat{\psi})(\ell b^{-1}) = 0$ for all $|\mu| < m$ and all $\ell \in \mathbb{Z}^d \setminus \{0\}$.

If $\int_{\mathbb{R}^d} \psi \, dx \neq 0$ then $A_J(\psi)$ spans $W^{m,p}$ for all $J \in \mathbb{Z}$.

We prove the theorem and corollary at the end of the section.

Remarks on Theorem 3 and Corollary 4.

1. If $\psi \in W^{m,\infty}$ has compact support then the P- and Q-hypotheses in Theorem 3 are all satisfied, because if $f \in L^\infty$ has compact support then $P(\mathbb{1}_{\{f \neq 0\}}), P(|f|) \in L^\infty$ and $Qf \in L^1$.

2. The theorem and corollary reduce when $m = 0$ back to the L^p spanning results (Theorem 1 and Corollary 2), except that the Sobolev results assume the dilations expand *nicely* and Corollary 4 assumes slightly more than Corollary 2 when $p = 1$. We do not know whether our Sobolev spanning results still hold when the a_j are expanding without expanding nicely.

Examples for Theorem 3. The easiest way to construct a ψ whose Fourier transform vanishes to order m at the nonzero lattice points, as required in Theorem 3 and Corollary 4, is to put

$$\psi = \psi_0 * u * \cdots * u \qquad \text{(with } m \text{ factors of } u\text{)} \tag{12}$$

where u has constant periodization $Pu = 1$ a.e. This works because $\widehat{\psi} = \widehat{\psi_0}\widehat{u} \cdots \widehat{u}$ while $Pu = 1$ a.e. implies $\widehat{u}(0) = 1$ and $\widehat{u}(\ell b^{-1}) = 0$ for all row vectors $\ell \in \mathbb{Z}^d \setminus \{0\}$ (see Section 3.3).

For example in one dimension this construction yields the triangular function $\psi(x) = 1 - |x|$ for $-1 \leq x \leq 1$, by choosing $\psi_0 = u = \mathbb{1}_{[-1/2,1/2)}$ and using $m = 1, b = 1$. By Theorem 3, the resulting system $A_J(\psi)$ will span $W^{1,p}(\mathbb{R}), 1 \leq p < \infty$.

Remarks on the Sobolev spanning literature. Prior spanning results for $W^{m,p}$ with $\int_{\mathbb{R}^d} \psi \, dx \neq 0$ make three assumptions: that $(D^\mu \widehat{\psi})(\ell b^{-1}) = 0$ for all $|\mu| \leq m$ (not just for $|\mu| < m$) and all $\ell \in \mathbb{Z}^d \setminus \{0\}$, that ψ has *compact support*, and that the dilations are expanding and *isotropic* ($a_j = \lambda_j I$). For $p = 2$ see [6, Theorem 4.1] (which even treats all fractional derivatives $m \in (0, \infty)$), and for $p = 2, \infty$ see [44, Theorems I,III].

Indeed Strang and Fix's comprehensive work in [20, 43, 44] led to the condition

$$(D^\mu \widehat{\psi})(\ell b^{-1}) = 0 \qquad \text{for all } |\mu| \leq m \text{ and } \ell \in \mathbb{Z}^d \setminus \{0\} \tag{13}$$

becoming known as the *Strang–Fix condition.* (Although historically, Schoenberg [41, Theorem 2] seems to have been the first to use the condition, in the context of polynomial interpolation and smoothing in one dimension.)

Di Guglielmo had earlier proved a spanning result [23, Théorème 2′] for $p \geq 2$ with anisotropic (though still diagonal) dilation matrices and with ψ having compact support. But for spanning $W^{m,p}$ in that paper, ψ was required to be an m-fold convolution like in (12) with u being the characteristic function of a cube. This means $\widehat{u}$ vanishes on the union of hyperplanes $\{\xi \in \mathbb{R}^d : \xi_i \in \mathbb{Z} \setminus \{0\}$ for some $i = 1, \ldots, d\}$, assuming $b = I$, and so di Guglielmo's transform $\widehat{\psi}$ vanishes on all these hyperplanes instead of just at the lattice points (where hyperplanes intersect) like in the Strang–Fix condition.

Also note the work in Mikhlin's monograph [35], where Strang–Fix type spanning results are obtained using "primitive functions". Unfortunately the required number of such functions grows with m, whereas here we need just one function, ψ.

In any event, Theorem 3 improves on all the literature because it assumes the Strang–Fix condition only for derivatives of order $|\mu| < m$, and it allows ψ to have noncompact support (so long as for example the weighted periodizations of ψ and its derivatives belong to L^∞), and the dilations can expand arbitrarily so long as they remain *nicely* expanding. Theorem 3 is also the first result to treat all $1 \leq p \leq \infty$. Corollary 4 and its decay condition are new as well.

Now we prove the theorem and corollary for spanning Sobolev space.

Proof of Theorem 3. The hypotheses ensure $\psi \in W^{m,1}$, in all three parts of the theorem. This is immediate when $p = 1$. When $1 < p < \infty$, if (9) holds then the set $\{x : \psi^{(\mu)}(x) \neq 0\}$ has finite measure and so $\psi^{(\mu)} \in L^p$ implies $\psi^{(\mu)} \in L^1$ by Hölder's inequality. On the other hand if (10) holds then $\psi^{(\mu)} \in L^1$ just by integrating $P(|\psi^{(\mu)}|)$ over the set $b\mathcal{C}$. When $p = \infty$ we know $Q(\psi^{(\mu)}) \in L^1$ in part (c), and $|\psi^{(\mu)}| \leq Q(\psi^{(\mu)})$ by Lemma 22 so that $\psi^{(\mu)} \in L^1$.

Since in particular $\psi \in L^1$ with $\int_{\mathbb{R}^d} \psi \, dx \neq 0$, we can normalize $\int_{\mathbb{R}^d} \psi \, dx = 1$.

And note that the hypothesis (8) on $D^\mu \widehat{\psi}$ is well defined because the Fourier transform is C^m-smooth away from the origin (by the first paragraph of Lemma 13, with $n = |\rho| = m$).

Now we begin the proof. Define $\phi = \mathbb{1}_{\mathcal{C}}$. By the density of C_c^m in $W^{m,p}$ for $1 \leq p < \infty$, to prove parts (a) and (b) we need only show that the $W^{m,p}$-span of $A_J(\psi)$ contains C_c^m. For part (a) and the first paragraph of part (b) this follows directly from case (i)′ of Proposition 14(b)(d) (take $\delta = 1$ in Proposition 14, and take either $\varepsilon = 0$ if (9) holds or $\varepsilon = 1$ if (10) holds; the value of ε is irrelevant when $p = 1$). For the second paragraph of (b), instead use Corollary 15 to show the $W^{m,p}$-span of $A_J(\psi)$ contains C_c^m.

For proving part (c) of the theorem, when $p = \infty$, simply let $f \in W^{m,\infty} \cap UC^m \cap L^1$ and use case (iii) of Proposition 14(a)(d) (note the values of ε and δ are irrelevant). □

Proof of Corollary 4. When $p = 1$, the spanning of $W^{m,1}$ by $A_J(\psi)$ is an immediate consequence of Theorem 3(a), with the decay condition (11) ensuring $\chi_{|\mu|}\psi^{(\mu)} \in L^1$.

Now suppose $1 < p < \infty$ and choose $R > 0$ with $|\psi^{(\mu)}(x)| \leq C|x|^{-d-|\mu|-\epsilon}$ whenever $|\mu| \leq m$ and $|x| \geq R$. Let $\zeta(x)$ be a smooth bump function that equals 1 on $\{|x| \leq R\}$ and has compact support.

Define $\psi_0 = \zeta\psi \in W^{m,p}$ and $\psi_1 = (1 - \zeta)\psi \in W^{m,p}$, so that $\psi = \psi_0 + \psi_1$. Notice ψ_0 has compact support and so $\chi_{|\mu|}\psi_0^{(\mu)} \in L^p$ and $P(\mathbb{1}_{\{\psi_0^{(\mu)} \neq 0\}}) \in L^\infty$ for each $|\mu| \leq m$. Thus ψ_0 satisfies (9). And the decay condition (11) ensures $\chi_{|\mu|}\psi_1^{(\mu)} \in L^p$, indeed with $|\chi_{|\mu|}\psi_1^{(\mu)}| \leq C \max\{R, |x|\}^{-d-\epsilon}$. This gives a radially decreasing L^1 majorant for $|\chi_{|\mu|}\psi_1^{(\mu)}|$, and so $P(|\chi_{|\mu|}\psi_1^{(\mu)}|) \in L^\infty$ by Lemma 19. Hence ψ_1 satisfies (10).

Theorem 3(b) now tells us that $A_J(\psi)$ spans $W^{m,p}$. □

3.3. **The zeroth Strang–Fix condition.** Taking $m = 0$ in the Strang–Fix condition (13) says for $\psi \in L^1$ that $\widehat{\psi}(\ell b^{-1}) = 0$ for all row vectors $\ell \in \mathbb{Z}^d \setminus \{0\}$. This is equivalent to ψ having constant periodization $P\psi \equiv$ (const.), as one sees by computing the ℓ^{th} Fourier coefficient of the $\mathbb{Z}^d$-periodic function $x \mapsto (P\psi)(bx)$.

3.4. **Open problems.** Write

$$A(\psi) = \{\psi_{j,k} : j \in \mathbb{Z}, k \in \mathbb{Z}^d\}$$

for the *affine system* generated by ψ, and recall $A_J(\psi)$ is the analogous small-scale system containing only the high frequencies $j \geq J$.

- *If $\psi \in L^1 \cap L^p$ and $\int_{\mathbb{R}^d} \psi\, dx \neq 0$, then does $A_J(\psi)$ span L^p?* This would render the periodization hypotheses in Theorem 1 unnecessary, which of course we already know for $p = 1$.
- *If $A(\psi)$ spans L^p and $\int_{\mathbb{R}^d} \psi\, dx \neq 0$, then is L^p spanned also by $A_J(\psi)$?* In other words, are the small scales (high frequencies) sufficient for spanning, provided spanning is known using all frequencies? (This is weaker than the previous question.)

  The answer is "No" when $\int_{\mathbb{R}^d} \psi\, dx = 0$, because for example if $A(\psi)$ forms an orthonormal basis for L^2, such as the Haar system does, then the small scale system $A_J(\psi)$ obviously does not span L^2 by itself.
- *Criteria for spanning when $\int_{\mathbb{R}^d} \psi\, dx = 0$?* Sufficient conditions are known for $A(\psi)$ to span $L^p, 1 < p < \infty$, in the wavelet setting [26, §5.3] and in the more general frame setting (cf. [25]), both of which require $\int_{\mathbb{R}^d} \psi\, dx \neq 0$. But there seems no good understanding of the pure spanning question, when $\psi \in L^1 \cap L^p$ and $\int_{\mathbb{R}^d} \psi\, dx = 0$.

  For example, Meyer has raised the following specific question in one dimension [34, p. 137]: does the affine system $A(\psi)$ span L^p for all $1 < p < \infty$, when $\psi = (1 - x^2)e^{-x^2/2}$ is the Mexican hat function and the dilations $a_j = 2^j$ are dyadic? This is known to be true when $p = 2$ (indeed the system forms a frame), but the problem is open for all other p-values. Notice $\int_{\mathbb{R}} \psi\, dx = 0$, since the Mexican hat is the second derivative of $-e^{-x^2/2}$.

  We hope in a future paper to extend our sampling methods to prove spanning in L^p for at least some functions ψ with $\int_{\mathbb{R}} \psi\, dx = 0$.

Next we take a broader view and consider the collections of affine and small-scale affine generators,

$$AG(W^{m,p}) = \{\psi : A(\psi) \text{ spans } W^{m,p}\} \quad \text{and} \quad AG_J(W^{m,p}) = \{\psi : A_J(\psi) \text{ spans } W^{m,p}\},$$

where $m \in \mathbb{N} \cup \{0\}$ and $1 \leq p < \infty$. We examine some known properties and open problems.

- *Multiplicative invariance.* If $\psi \in AG(W^{m,p})$ then $c\psi \in AG(W^{m,p})$ for all $c \neq 0$. Similarly for AG_J.
- *Translation invariance.* If $\psi \in AG(W^{m,p})$ then $\psi(\cdot - by) \in AG(W^{m,p})$ for all $y \in \mathbb{Z}^d$. Similarly for AG_J.

  We do not know whether this translation invariance holds for all $y \in \mathbb{R}^d$, though it certainly does if ψ satisfies the hypotheses of Theorems 1 or 3, because then $\psi(\cdot - by)$ satisfies those hypotheses also.

- *Invariance under differentiation.* In one dimension, if $\psi \in AG(W^{m,p})$ and $m \geq 1, 1 < p < \infty$, then $\psi' \in AG(W^{m-1,p})$. Similarly for AG_J. *Proof:* Consider $f \in C_c^\infty$ with $\int_{\mathbb{R}} f(x)\,dx = 0$ (such functions f are dense in $W^{m-1,p}$, since $1 < p < \infty$). Then $F(x) = \int_{-\infty}^x f(y)\,dy$ belongs to C_c^∞, with $F' = f$. Since F can be approximated in $W^{m,p}$ by a linear combination of the $\psi_{j,k}$, we see that f can be approximated in $W^{m-1,p}$ by a linear combination of the $(\psi')_{j,k}$.
- *$AG(L^p)$ is dense in L^p.* Similarly for AG_J. *Proof:* The class of Schwartz functions with $\int_{\mathbb{R}^d} \psi\,dx \neq 0$ is dense in L^p, and every such function satisfies the hypotheses of Corollary 2 and so belongs to $AG_J(L^p) \subset AG(L^p)$.
- *Is $AG(L^p)$ topologically open?* The subclass of $AG_J(L^p)$ that we identify in Theorem 1 is *not* open in L^p for $1 < p < \infty$. But it is difficult to think how one might find a counterexample to openness for the full collections $AG_J(L^p)$ or $AG(L^p)$, since we currently don't have good necessary conditions on them.
- *Is $AG(W^{m,p})$ pathwise connected?* We do not know. The subclasses of $AG_J(W^{m,p})$ that we develop for $1 \leq p < \infty$ in Theorems 1 and 3 are certainly pathwise connected. *Proof:* The linear variation $(1-t)\psi + t\widetilde{\psi}$ for $0 \leq t \leq 1$ provides a path from ψ to $\widetilde{\psi}$ that is valid for Theorems 1 and 3 provided $\int_{\mathbb{R}^d}[(1-t)\psi + t\widetilde{\psi}]\,dx \neq 0$ for all $0 \leq t \leq 1$. By combining two such linear paths, we can connect together any valid pair $\psi, \tilde{\psi}$.
- *Independence from the dilations.* Is membership in $AG(L^p)$ independent of the expanding dilation matrices $\{a_j\}_{j \in \mathbb{Z}^d}$, for fixed translation matrix b?

  We do not know, though the hypotheses of Theorems 1 and 3 are certainly independent of the particular dilations a_j employed.

3.5. **Interchange of dilation and translation?** As a sidelight, we ask: can ψ still generate a spanning set if the dilation and translation operations are interchanged? In particular, can $\{\psi(2^j(x-k)) : j \in \mathbb{Z}, k \in \mathbb{Z}^d\}$ span L^p? We would expect the answer to be *No* in most cases. For example, with $\psi = \mathbb{1}_{[0,1/2)}$ in one dimension we see each function $\psi(2^j(x-k))$ is constant on the interval $1/2 < x < 1$, and this "limit on the resolution" prevents any possibility of spanning L^p.

4. Averaging of rapidly oscillating functions

For the sampling formulas in Section 5 we need convergence of averages of rescaled periodizations of ψ. The rescaled periodization $(P\psi)(a_j x)$ oscillates rapidly when j is large and hence converges *weakly* to its mean value as $j \to \infty$. We will further obtain *norm* and *pointwise* convergence to the mean value, by taking a subsequence and then an arithmetic mean with respect to j.

Lemma 5. *Let $1 \leq p < \infty$ and suppose $g \in L^p_{loc}$ is periodic with respect to the lattice $b\mathbb{Z}^d$ and has mean value zero. Let $\{j_1(n)\}_{n=1}^\infty$ be a strictly increasing sequence of integers.*

Then a subsequence $\{j_2(n)\}_{n=1}^\infty$ exists such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N g(a_{j_3(n)}x) = 0 \qquad \text{in } L^p_{loc} \text{ and pointwise a.e.} \tag{14}$$

for each subsequence j_3 of j_2.

In particular, (14) *holds if also* $g \in BV_{loc} \cap L^\infty$ *and* j_2 *is chosen to ensure the dilations* $a_{j_2(n)}$ *grow exponentially (meaning* $|a_{j_2(n+1)}x| \geq \gamma |a_{j_2(n)}x|$ *for all* $x \in \mathbb{R}^d$ *and all* $n \in \mathbb{N}$*, for some growth factor* $\gamma > 1$*).*

Here $BV_{loc} = BV_{loc}(\mathbb{R}^d)$ is the class of functions having bounded variation on every bounded open set $\mathcal{O} \subset \mathbb{R}^d$ (see Section 2.2). In one dimension, BV_{loc} functions are automatically locally bounded by [18, §5.10], in which case the assumption $g \in L^\infty$ is superfluous in the last paragraph of the Proposition because also g is periodic. Note in all dimensions that Sobolev functions have locally bounded variation: $W^{1,1}_{loc} \subset BV_{loc}$ by [18, p. 170].

In the last paragraph of the lemma, exponential growth of the dilations $a_{j_2(n)}$ can indeed be attained by suitable choice of the subsequence j_2, because the matrices a_j are expanding with $|a_j x| \geq \lambda_j |x|$ and $\lambda_j \to \infty$. Of course if the dilations are exponentially growing in the first place (such as the dyadic dilations $a_j = 2^j I$) then we can just take $j_2(n) = n$.

Example. The following trigonometric example demonstrates the automatic cancellation of errors at different scales that underpins Lemma 5.

Suppose in one dimension that $g(x) = \cos(2\pi x)$ and the dilations a_j for $j > 0$ form a strictly increasing sequence of positive integers. Then the arithmetic means (over $j > 0$) of the functions $g(a_j x)$ converge to zero in L^2_{loc}, because

$$\Big\| \frac{1}{N} \sum_{j=1}^N g(a_j x) \Big\|^2_{L^2[0,1]} = \Big\| \frac{1}{N} \sum_{j=1}^N \cos(2\pi \cdot a_j x) \Big\|^2_{L^2[0,1]}$$

$$= \frac{1}{N^2} \sum_{j=1}^N \| \cos(2 a_j \pi x) \|^2_{L^2[0,1]}$$

by trigonometric orthogonality, since the a_j are distinct integers,

$$= \frac{1}{2N} \to 0 \qquad \text{as } N \to \infty.$$

Clearly the same argument works when using any subsequence of the dilations a_j.

It's interesting to plot the graphs of $\cos(2\pi x), \cos(4\pi x), \cos(6\pi x), \ldots$, and see the cancellations that occur when these functions are added.

Proof of Lemma 5. Write $B(t)$ for the open ball of radius t centered at the origin. Then $g(a_j x) \rightharpoonup 0$ weakly in $L^p(B(t))$ as $j \to \infty$, by the Riemann–Lebesgue Lemma 26 applied to an arbitrary $h \in L^q(B(t))$. (This lemma uses the mean value zero hypothesis on g.) In particular, $g(a_{j_1(n)}x) \rightharpoonup 0$ weakly in $L^p(B(t))$ as $n \to \infty$.

Hence Corollary 28 applied to these functions yields a subsequence j_2 of j_1 such that $\frac{1}{N} \sum_{n=1}^N g(a_{j_3(n)}x) \to 0$ in $L^p(B(t))$ and pointwise a.e. in $B(t)$, for every subsequence j_3 of j_2 and every $t \in \mathbb{N}$. Condition (14) follows immediately, since every compact set lies in one of the balls $B(t)$.

It remains only to prove the last paragraph of Lemma 5. This paragraph is used only in Corollary 11 and in Remark 3 after Proposition 14, and so the rest of the proof can safely be skipped on first reading.

Assume $g \in BV_{loc} \cap L^\infty$ is periodic with respect to the lattice $b\mathbb{Z}^d$ and has mean value zero. Assume the subsequence $j_2(n)$ has been chosen to ensure that the dilations $a_{j_2(n)}$ grow exponentially. Write $J = j_2(1)$.

We need only prove (14) holds pointwise a.e., for each subsequence j_3 of j_2, because then it holds in L^p_{loc} by dominated convergence, using the boundedness of g. We can suppose g is real valued.

Our first observation is that the exponential growth assumption implies $|a_{j_2(n)}a^{-1}_{j_2(n+1)}x| \leq \gamma^{-1}|x|$. Iterating this inequality yields the matrix norm estimate

$$\|a_{j_2(m)}a^{-1}_{j_2(n)}\| \leq \gamma^{m-n} \qquad \text{whenever } 1 \leq m \leq n. \tag{15}$$

The exponential growth assumption also implies $|a_{j_2(n)}x| \geq \gamma^{n-1}|a_{j_2(1)}x| \geq \gamma^{n-1}\lambda_J|x|$, so that $\|a^{-1}_{j_2(n)}\| \leq \gamma^{1-n}\lambda_J^{-1}$.

Next we will show certain inner products of the functions $g(a_{j_2(n)}x)$ are close to zero:

$$\left| \int_{B(t)} g(a_{j_2(m)}x)g(a_{j_2(n)}x)\,dx \right| \leq A\gamma^{m-n} = A\gamma^{-|m-n|} \tag{16}$$

for all $1 \leq m \leq n$, for some positive constant $A = A(t, J, \|g\|_\infty, \|\nabla g\|(b\mathcal{C}))$ that depends in an increasing fashion on its last two parameters. (Here $\|\nabla g\|$ is a positive Radon measure, since g has locally bounded variation.) Once (16) has been established, it will clearly also hold for each subsequence j_3 of j_2. Then Lemma 29 with $\beta(\ell) = A\gamma^{-|\ell|}$ will imply the desired pointwise a.e. convergence $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N g(a_{j_3(n)}x) = 0$ in the ball $B(t)$, for each t, completing the proof of Lemma 5.

So we have only to prove (16). We can further suppose g is smooth, as follows. Let $g_\varepsilon = g * \eta_\varepsilon$ for $\varepsilon > 0$ be a smooth mollification of g by some nonnegative bump function η having integral 1. Notice $g_\varepsilon \in BV_{loc} \cap L^\infty \cap C^\infty$ is $b\mathbb{Z}^d$-periodic and has mean value zero, with norm estimates $\|g_\varepsilon\|_\infty \leq \|g\|_\infty$ and $\|\nabla g_\varepsilon\|(b\mathcal{C}) \leq C(d)\|\nabla g\|(b\mathcal{C})$. Since $g_\varepsilon \to g$ in L^2_{loc} as $\varepsilon \to 0$, we can choose $\varepsilon = \varepsilon(m, n, t, g) > 0$ to be so small that

$$\left| \int_{B(t)} g(a_{j_2(m)}x)g(a_{j_2(n)}x)\,dx - \int_{B(t)} g_\varepsilon(a_{j_2(m)}x)g_\varepsilon(a_{j_2(n)}x)\,dx \right| \leq \gamma^{m-n}.$$

Hence (16) will follow once we prove it for g_ε instead of g. Thus we might as well suppose g is smooth.

Since g has mean value zero, Poisson's equation $\Delta u = g$ has a solution $u \in C^\infty$ that is $b\mathbb{Z}^d$-periodic and satisfies the normalization $\int_{b\mathcal{C}} u(x)\,dx = 0$. And $\|u\|_\infty \leq C\|g\|_\infty$ by the solution formula for Poisson's equation on the compact torus $\mathbb{R}^d/b\mathbb{Z}^d$, using that the Green function on this torus has L^1-norm that is constant with respect to the variable not integrated; see [5, Theorem 4.13]. Now a classical maximum principle argument on a neighborhood of $b\mathcal{C} \subset \mathbb{R}^d$ gives the gradient estimate [22, (3.16)]

$$\|\nabla u\|_\infty \leq C(\|u\|_\infty + \|g\|_\infty) \leq C\|g\|_\infty. \tag{17}$$

Rewrite the integral in (16) as

$$\int_{B(t)} g(a_{j_2(m)}x)g(a_{j_2(n)}x)\,dx \;= \\ -\int_{B(t)} a_{j_2(m)}a^{-1}_{j_2(n)}\nabla u(a_{j_2(n)}x)\cdot\nabla g(a_{j_2(m)}x)\,dx + \int_{\partial B(t)} [g(a_{j_2(m)}x)a^{-1}_{j_2(n)}\nabla u(a_{j_2(n)}x)]\cdot\nu(x)\,dS(x), \tag{18}$$

by the divergence theorem applied to the vector field $g(a_{j_2(m)}x)a_{j_2(n)}^{-1}\nabla u(a_{j_2(n)}x)$. The first integral in (18) is bounded by

$$\begin{aligned}
&\|a_{j_2(m)}a_{j_2(n)}^{-1}\|\,\|\nabla u\|_\infty \int_{B(t)} |\nabla g(a_{j_2(m)}x)|\,dx \\
&= \|a_{j_2(m)}a_{j_2(n)}^{-1}\|\,\|\nabla u\|_\infty \frac{1}{|\det a_{j_2(m)}|}\int_{a_{j_2(m)}B(t)} |\nabla g(x)|\,dx \qquad \text{by } x \mapsto a_{j_2(m)}^{-1}x \\
&\le \gamma^{m-n}\|g\|_\infty\, C(t,J)\|\nabla g\|(b\mathcal{C}) \qquad (19)
\end{aligned}$$

by the exponential growth condition (15), by (17), and by Lemma 25 applied to the measure $|\nabla g|\,dx$.

For the second integral in (18), estimate similarly to obtain

$$\|g\|_\infty\, \|a_{j_2(n)}^{-1}\|\|\nabla u\|_\infty\, |\partial B(t)| \le C(t)\gamma^{1-n}\lambda_J^{-1}\|g\|_\infty^2. \qquad (20)$$

Now estimate (16) follows from (19) and (20), because $\gamma^{1-n} \le \gamma^{m-n}$.

□

5. **Discretized approximations to the identity**

The core sampling results of the paper are developed in this section. These results are somewhat technical, and readers might prefer to first locate the most relevant sampling formulas for L^p in Section 6 or for $W^{m,p}$ in Section 7, and then work backwards to the needed parts of this section.

The key object of our study is a "discretized approximation to the identity" operator $I_j[\psi,\phi]$, defined by

$$(I_j[\psi,\phi]h)(x) = |\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} h(x, a_j^{-1}y - x)\phi(y - bk)\,dy \right) \psi_{j,k}(x), \qquad j \in \mathbb{Z}. \qquad (21)$$

Lemma 6 will specify properties of the synthesizer ψ and of the analyzer ϕ under which I_j is well defined. We require $h(x,y)$ to belong to the mixed-norm space

$$L^{(p,\infty)} = \{h : h \text{ is measurable on } \mathbb{R}^d \times \mathbb{R}^d \text{ and } \|h\|_{(p,\infty)} < \infty\}$$

where $\|h\|_{(p,\infty)} = \text{ess. sup}_{y\in\mathbb{R}^d}(\int_{\mathbb{R}^d} |h(x,y)|^p\,dx)^{1/p}$. That is, $\|h\|_{(p,\infty)}$ takes the L^p norm of h with respect to x, and then the L^∞ norm with respect to y. The definition is analogous when $p = \infty$, and it turns out that $L^{(\infty,\infty)} = L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$.

For example if $h(x,y) = f(x+y)$ and $f \in L^p$ then $h \in L^{(p,\infty)}$ with $\|h\|_{(p,\infty)} = \|f\|_p$. Notice the definition of I_j in (21) simplifies considerably when $h(x,y) = f(x+y)$.

Motivation for I_j. To motivate the operator I_j, suppose $h(x,y) = f(x+y)$, $\int_{\mathbb{R}^d}\psi(x)\,dx = 1$, $\phi = \mathbb{1}_{\mathcal{C}}$ and $b = I$, and relate $I_j[\psi,\phi]h$ to a classical approximation to the identity as follows. We have

$$\begin{aligned} f(x) &= \lim_{\varepsilon\to 0}(f*\psi_\varepsilon)(x) \qquad \text{in } L^p \\ &= \lim_{\varepsilon\to 0}\int_{\mathbb{R}^d} f(z)\varepsilon^{-d}\psi(\varepsilon^{-1}(x-z))\,dz \\ &= \lim_{\varepsilon\to 0}\int_{\mathbb{R}^d} f(\varepsilon y)\psi(\varepsilon^{-1}x-y)\,dy \qquad \text{by } z=\varepsilon y \\ &\approx \lim_{\varepsilon\to 0}\sum_{k\in\mathbb{Z}^d}\left(\int_{k+\mathcal{C}} f(\varepsilon y)\,dy\right)\psi(\varepsilon^{-1}x-k) \end{aligned} \tag{22}$$

by a Riemann sum approximation. If we replace ε with a_j^{-1}, then this last line is exactly $\lim_{j\to\infty} I_j[\psi,\phi]h$, as we wanted. But extreme caution is required in the Riemann sum approximation step, because it discretizes with a fixed step size, indeed step size 1. Nonetheless the approximation is valid under suitable conditions on ψ *provided we average over different dilation levels*, as we will show in Lemma 7(c).

Now we develop properties of the I_j operator. Recall the periodization operator P defined in Section 2. And introduce

$$|\psi|^0 = \mathbb{1}_{\{\psi\neq 0\}} \qquad \text{and} \qquad |\phi|^0 = \mathbb{1}_{\{\phi\neq 0\}}$$

as shorthand notations for the characteristic functions where ψ and ϕ are nonzero. Define

$$p(\varepsilon) = \begin{cases} p & \text{if } \varepsilon = 0, \\ 1 & \text{if } \varepsilon = 1, \end{cases} \qquad \text{and} \qquad q(\delta) = \begin{cases} q & \text{if } \delta = 0, \\ 1 & \text{if } \delta = 1. \end{cases}$$

Lemma 6.

Take $\varepsilon,\delta\in\{0,1\}$. Assume one of the following conditions holds:

- (i) $1\le p<\infty$ *and* $\psi\in L^p, (p-1)P(|\psi|^\varepsilon)\in L^\infty, \phi\in L^q, P(|\phi|^\delta)\in L^\infty$ *and* $h(x,y)=f(x+y)$ *for some* $f\in L^p$*;*
- (i)′ $1\le p<\infty$ *and* $\psi\in L^p,(p-1)P(|\psi|^\varepsilon)\in L^\infty$, $\phi\in L^{q(\delta)}$ *with* $|\phi|^\delta\in L^1$, *and* $h(x,y)=\int_{[0,1]} f(x+ty)\,d\omega(t)$ *for some* $f\in C_c$ *and some Borel probability measure* ω *on* $[0,1]$*;*
- (ii) $1\le p<\infty$ *and* $\psi\in L^p,(p-1)P(|\psi|^\varepsilon)\in L^\infty, Q(|\psi|^{p(\varepsilon)})\in L^1$, *and* $\phi\in L^q$ *with* ϕ *having compact support, and* $h\in L^{(p,\infty)}$*;*
- (iii) $p=\infty$ *and* $\psi\in L^\infty, P(|\psi|)\in L^\infty,\phi\in L^1$ *and* $h\in L^{(\infty,\infty)}$*;*
- (iv) $p=\infty$ *and* $\psi\in L^\infty,\phi\in L^1, P(|\phi|)\in L^\infty$ *and* $h(x,y)=f(x+y)$ *for some* $f\in L^1$.

Then the series (21) *defining* $I_j[\psi,\phi]h$ *converges pointwise absolutely a.e. to an* L^p *function. The series further converges unconditionally in* L^p *if (i), (i)′, (ii) or (iv) holds.*

And in cases (i), (ii) and (iii) a norm stability estimate holds independent of $j\in\mathbb{Z}$*:*

$$\|I_j[\psi,\phi]h\|_p \le \|h\|_{(p,\infty)}\cdot\begin{cases} \|\psi\|_{p(\varepsilon)}^{1-\varepsilon/q}\|P|\psi|^\varepsilon\|_\infty^{1/q}\|\phi\|_{q(\delta)}^{1-\delta/p}\|P|\phi|^\delta\|_\infty^{1/p} & \textit{in case (i),} \\ C(\mathrm{spt}(\phi))^{1/p}\|Q|\psi|^{p(\varepsilon)}\|_1^{1/p}\|P|\psi|^\varepsilon\|_\infty^{1/q}\|\phi\|_q & \textit{in case (ii),} \\ \|P|\psi|\|_\infty\|\phi\|_1 & \textit{in case (iii).} \end{cases} \tag{23}$$

Remarks on Lemma 6.

1. Cases (i) and (ii) will be used for sampling in L^p, and cases (i)$'$ and (ii) for sampling in Sobolev space (where ultimately we will take $h(x, y)$ to be the Taylor polynomial of f based at x with increment y).

2. The lemma restricts ε and δ to take values 0 or 1 because this seems to capture the most interesting results. But the method of proof allows $\varepsilon \in [0, q]$ and $\delta \in [0, p]$, when care is taken.

The assumption $(p-1)P(|\psi|^\varepsilon) \in L^\infty$ is vacuous when $p = 1$, and the corresponding factor $\|P|\psi|^\varepsilon\|_\infty^{1/q}$ in the norm estimate should be replaced by 1 in that case.

3. Clearly case (i) assumes less about ψ than case (ii) does, but on the other hand it assumes more about h. Case (ii) is the only one to consider the general function $h \in L^{(p,\infty)}$.

4. Case (i) is a special case of case (i)$'$ when $f \in C_c$, as one sees by taking the measure ω to be a delta mass at $t = 1$. But in case (i) the additional assumption $P|\phi|^\delta \in L^\infty$ allows us to prove the norm estimate (23) that is independent of j, which later yields stability of our L^p sampling formulas. We have no stability estimate in case (i)$'$.

5. The series defining $I_j[\psi, \phi]h$ will generally not converge in L^∞, for example if $h \equiv 1$ and ψ has compact support. Thus we do not expect unconditional convergence in L^∞, in case (iii).

6. The assumption $Q(|\psi|^{p(\varepsilon)}) \in L^1$ in case (ii) allows us to bound the values of ψ at nearby points, so that we can estimate certain Riemann sums involving ψ with integrals involving $Q\psi$. We do this formally in (22) and rigorously in (29).

7. The norm estimate (23) in case (i) involves the $L^{p(\varepsilon)}$-norm of ψ, which is finite because when $\varepsilon = 0$ we have $p(\varepsilon) = p$ and $\psi \in L^p$ by hypothesis, while when $\varepsilon = 1$ we have $p(\varepsilon) = 1$ and $(p-1)P|\psi| \in L^\infty$ so that $\psi \in L^1 = L^{p(\varepsilon)}$. Similarly the $L^{q(\delta)}$-norm of ϕ is finite.

8. The behavior of I_j was studied by di Guglielmo [23] for $p \geq 2$ under the restrictions that ψ have constant periodization (indeed that ψ equal a convolution with the characteristic function of a cube) and have compact support, that ϕ be bounded with compact support, and that $h(x, y) = f(x + y)$. Case (i) in Lemma 6 (and in the next lemma) builds on di Guglielmo's work [23, p. 288]. The other cases of the lemmas are new.

We discuss other relevant literature after the sampling formulas in Sections 6 and 7.

Proof of Lemma 6.

The integral $\int_{\mathbb{R}^d} h(x, a_j^{-1}y - x)\phi(y - bk)\, dy$ occurring in the definition of I_j is well defined, as follows. In case (i), $h(x, a_j^{-1}y - x) = f(a_j^{-1}y)$ belongs to L^p as a function of y, while $\phi(y - bk) \in L^q$. In case (i)$'$, h is bounded and our assumptions ensure $\phi \in L^1$. In case (ii) we see $y \mapsto h(x, y)$ belongs to L^p_{loc} for almost every x and that $\phi \in L^q$ has compact support. In case (iii), h is bounded and $\phi \in L^1$. In case (iv), $h(x, a_j^{-1}y - x) = f(a_j^{-1}y)$ belongs to L^1 and $|\phi| \leq P|\phi|$ is bounded. So in every case, the integral is well defined.

Consider $1 \le p < \infty$. To start with,

$$|(I_j[\psi,\phi]h)(x)|^p \le \left(|\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |h(x, a_j^{-1}y - x)||\phi(y-bk)|\, dy\, |\psi(a_j x - bk)| \right)^p$$
$$\le |\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} |h(x, a_j^{-1}y - x)||\phi(y-bk)|\, dy \right)^p |\psi(a_j x - bk)|^{p(\varepsilon)}$$
$$\cdot \left(|\det b| \sum_{k\in\mathbb{Z}^d} |\psi(a_j x - bk)|^{\varepsilon} \right)^{p-1}$$

by Hölder's inequality on the sum, when $p > 1$. (When $p = 1$ the last inequality is vacuous, because then $p(\varepsilon) = 1$ for both $\varepsilon = 0$ and $\varepsilon = 1$.)

By applying Hölder's inequality to the y-integral we find

$$|(I_j[\psi,\phi]h)(x)|^p$$
$$\le |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |h(x, a_j^{-1}y - x)|^p |\phi(y-bk)|^{\delta}\, dy\, \|\phi\|_{q(\delta)}^{p-\delta}\, |\psi(a_j x - bk)|^{p(\varepsilon)} \cdot \|P|\psi|^{\varepsilon}\|_{\infty}^{p-1}. \tag{24}$$

Case (i). If $h(x,y) = f(x+y)$ then integrating (24) with respect to x yields the norm estimate

$$\|I_j[\psi,\phi]h\|_p^p \le |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |f(y)|^p |\phi(a_j y - bk)|^{\delta}\, dy\, \|\psi\|_{p(\varepsilon)}^{p(\varepsilon)} \cdot \|P|\psi|^{\varepsilon}\|_{\infty}^{p-1} \|\phi\|_{q(\delta)}^{p-\delta} \tag{25}$$
$$\le \|f\|_p^p \|P|\phi|^{\delta}\|_{\infty}\, \|\psi\|_{p(\varepsilon)}^{p(\varepsilon)} \cdot \|P|\psi|^{\varepsilon}\|_{\infty}^{p-1} \|\phi\|_{q(\delta)}^{p-\delta}$$

as claimed in norm estimate (23) for case (i). Or to argue more carefully, the finiteness of this last estimate tells us the series defining $I_j[\psi,\phi]h$ converges absolutely a.e. to an L^p function that satisfies the norm estimate (23).

We follow this same method in the rest of the proof, that is, we apply absolute values and then prove an L^p estimate, so that the pointwise convergence of the series defining $I_j[\psi,\phi]h$ follows automatically.

Case (i)′. By integrating (24) with respect to x then substituting $h(x,y) = \int_{[0,1]} f(x + ty)\, d\omega(t)$ and making the changes of variable $x \mapsto a_j^{-1}(x+bk)$ and $y \mapsto y + bk$, we deduce

$$\|I_j[\psi,\phi]h\|_p^p \le \int_{[0,1]} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} R_j(x,y,t)\, |\psi(x)|^{p(\varepsilon)} |\phi(y)|^{\delta}\, dx dy d\omega(t) \cdot \|P|\psi|^{\varepsilon}\|_{\infty}^{p-1} \|\phi\|_{q(\delta)}^{p-\delta} \tag{26}$$

where

$$R_j(x,y,t) = |\det a_j^{-1} b| \sum_{k\in\mathbb{Z}^d} |f(a_j^{-1}(x+bk) + t a_j^{-1}(y-x))|^p \tag{27}$$
$$\le \|f\|_{\infty}^p \cdot |\{\xi \in \mathbb{R}^d : \operatorname{dist}(\xi, \operatorname{spt}(f)) \le \operatorname{diam}(a_j^{-1} b\mathcal{C})\}|. \tag{28}$$

Hence R_j is bounded independently of x, y, t. Since also our assumptions in case (i)′ imply $\psi \in L^{p(\varepsilon)}$, $|\phi|^{\delta} \in L^1$ and $\phi \in L^{q(\delta)}$, we see $I_j[\psi,\phi]h$ belongs to L^p by the estimate (26).

Case (ii). Using $\delta = 0$ in formula (24) shows

$$\begin{aligned}
&|(I_j[\psi,\phi]h)(x)|^p \\
&\leq |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |h(x, a_j^{-1}y - x)|^p |\phi(y-bk)|^0 (\widetilde{Q}|\psi|^{p(\varepsilon)})(a_j x - y)\, dy \cdot \|P|\psi|^\varepsilon\|_\infty^{p-1} \|\phi\|_q^p \qquad (29) \\
&\qquad\qquad \text{for almost every } x, \text{ by (125) in Lemma 22 with } E = \operatorname{spt}(\phi), \\
&\leq \int_{\mathbb{R}^d} |h(x, -a_j^{-1}y)|^p (\widetilde{Q}|\psi|^{p(\varepsilon)})(y)\, dy \cdot \|P|\phi|^0\|_\infty \|P|\psi|^\varepsilon\|_\infty^{p-1} \|\phi\|_q^p \qquad \text{by } y \mapsto a_j x - y. \qquad (30)
\end{aligned}$$

Integrating with respect to x gives the norm estimate

$$\begin{aligned}
\|I_j[\psi,\phi]h\|_p^p &\leq \int_{\mathbb{R}^d} \|h(\cdot\,, -a_j^{-1}y)\|_p^p (\widetilde{Q}|\psi|^{p(\varepsilon)})(y)\, dy \cdot \|P|\phi|^0\|_\infty \|P|\psi|^\varepsilon\|_\infty^{p-1} \|\phi\|_q^p \qquad (31) \\
&\leq C \|h\|_{(p,\infty)}^p \|Q|\psi|^{p(\varepsilon)}\|_1 \cdot \|P|\psi|^\varepsilon\|_\infty^{p-1} \|\phi\|_q^p
\end{aligned}$$

where $C = C(\operatorname{spt}(\phi))$, using here that $\|\widetilde{Q}\cdot\|_1 \leq C(E)\|Q\cdot\|_1$ by definition of $\widetilde{Q}$ in (125). Thus we have proved estimate (23) in case (ii).

Unconditional convergence. The series defining $I_j[\psi,\phi]h$ converges unconditionally in L^p in cases (i), (i)′ and (ii), because

$$\lim_{K\to\infty} \sum_{|k|\geq K} \left| \left(|\det b| \int_{\mathbb{R}^d} h(x, a_j^{-1}y - x)\phi(y - bk)\, dy \right) \psi_{j,k}(x) \right| = 0$$

in L^p by dominated convergence (using the pointwise absolute convergence proved above).

Consider $p = \infty$.

Case (iii). With $\psi \in L^\infty, \phi \in L^1, P|\psi| \in L^\infty$ and $h \in L^{(\infty,\infty)}$ we find

$$\begin{aligned}
|(I_j[\psi,\phi]h)(x)| &\leq |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |h(x, a_j^{-1}y - x)|\, |\phi(y-bk)|\, dy\, |\psi(a_j x - bk)| \\
&\leq \|h\|_{(\infty,\infty)} \|\phi\|_1 \|P|\psi|\|_\infty
\end{aligned}$$

for almost every x, implying the norm estimate (23) for case (iii).

Case (iv). Suppose $\psi \in L^\infty, \phi \in L^1, P|\phi| \in L^\infty$ and $h(x,y) = f(x+y)$ for some $f \in L^1$ (yes, we really mean L^1). Then for each $K \geq 0$,

$$\begin{aligned}
&\sum_{|k|\geq K} \left\| \left(|\det b| \int_{\mathbb{R}^d} h(x, a_j^{-1}y - x)\phi(y-bk)\, dy \right) \psi_{j,k}(x) \right\|_\infty \\
&\qquad \leq \int_{\mathbb{R}^d} |f(a_j^{-1}y)| \left(|\det b| \sum_{|k|\geq K} |\phi(y - bk)| \right) dy\, \|\psi\|_\infty \\
&\qquad \to 0 \qquad \text{as } K \to \infty
\end{aligned}$$

by dominated convergence with the dominating function involving $f \in L^1$ and $P|\phi| \in L^\infty$. Hence the series defining $I_j[\psi,\phi]h$ converges unconditionally in L^∞ for each fixed j.

(*Aside.* The defect of case (iv) is that its norm estimate depends on j, as we see by taking $K = 0$ above.) □

The next lemma justifies our calling $I_j[\psi,\phi]$ a discretized approximation to the identity.

Lemma 7. *Take $\varepsilon,\delta \in \{0,1\}$. Assume one of the following conditions holds:*

(i) $1 \le p < \infty$ *and* $\psi \in L^p, (p-1)P(|\psi|^\varepsilon) \in L^\infty, \phi \in L^q, P(|\phi|^\delta) \in L^\infty$, *and* $h(x,y) = f(x+y)$ *for some* $f \in L^p$;

(i)′ $1 \le p < \infty$ *and* $\psi \in L^p, (p-1)P(|\psi|^\varepsilon) \in L^\infty$, $\phi \in L^{q(\delta)}$ *with* $|\phi|^\delta \in L^1$, *and* $h(x,y) = \int_{[0,1]} f(x+ty)\, d\omega(t)$ *for some* $f \in C_c$ *and some Borel probability measure* ω *on* $[0,1]$;

(ii) $1 \le p < \infty$ *and* $\psi \in L^p, (p-1)P(|\psi|^\varepsilon) \in L^\infty, Q(|\psi|^{p(\varepsilon)}) \in L^1$ *and* $\phi \in L^q$ *with* ϕ *having compact support, and* $h \in L^{(p,\infty)}$ *with*

$$\lim_{y\to 0} h(\cdot,y) = h(\cdot,0) \qquad \text{in } L^p; \tag{32}$$

(iii) $p = \infty$ *and* $\psi \in L^\infty, Q\psi \in L^1$ *and* $\phi \in L^\infty$ *with* ϕ *having compact support, and* $h \in L^{(\infty,\infty)}$ *with* $\lim_{y\to 0} h(\cdot,y) = h(\cdot,0)$ *in* L^∞.

Then (a)–(d) hold:

(a) [Upper bound]

$$\limsup_{j\to\infty} \|I_j[\psi,\phi]h\|_p \le \|h(\cdot,0)\|_p \cdot \begin{cases} \|\psi\|_{p(\varepsilon)}^{1-\varepsilon/q} \|P|\psi|^\varepsilon\|_\infty^{1/q} \|\phi\|_{q(\delta)}^{1-\delta/p} \||\phi|^\delta\|_1^{1/p} & \text{in cases (i), (i)}', \\ C(\mathrm{spt}(\phi))^{1/p} \|Q|\psi|^{p(\varepsilon)}\|_1^{1/p} \|P|\psi|^\varepsilon\|_\infty^{1/q} \|\phi\|_q & \text{in case (ii)}, \\ C(\mathrm{spt}(\phi)) \|Q\psi\|_1 \|P|\phi|\|_\infty & \text{in case (iii)}. \end{cases} \tag{33}$$

(b) [Constant periodization] If $(P\psi)(x) = \int_{\mathbb{R}^d} \psi(y)\, dy$ for almost every x then

$$\lim_{j\to\infty} (I_j[\psi,\phi]h)(x) = h(x,0) \int_{\mathbb{R}^d} \psi(y)\, dy \int_{\mathbb{R}^d} \phi(z)\, dz \qquad \text{in } L^p. \tag{34}$$

(c) [Arbitrary periodization] Suppose $1 \le p < \infty$ and $J \in \mathbb{Z}$. A strictly increasing integer sequence $\{j(n)\}_{n=1}^\infty$ exists (independent of h) such that $j(1) \ge J$ and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (P\psi)(a_{j(n)}x) = \int_{\mathbb{R}^d} \psi(y)\, dy \qquad \text{in } L^p_{loc} \text{ and pointwise a.e.} \tag{35}$$

For any such sequence,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (I_{j(n)}[\psi,\phi]h)(x) = h(x,0) \int_{\mathbb{R}^d} \psi(y)\, dy \int_{\mathbb{R}^d} \phi(z)\, dz \qquad \text{in } L^p. \tag{36}$$

(d) [Pointwise convergence] Instead of (i), (i)′, (ii) or (iii), assume $1 \le p < \infty$ and $\psi \in L^1 \cap L^\infty$ and suppose $\phi \in L^q$ has compact support. Suppose either ψ has compact support and $h \in L^{(p,\infty)}$ or else ψ has a radially decreasing L^1-majorant and $h \in L^{(\infty,\infty)}$. Assume further (cf. Remark 3 below) that

$$\lim_{r\to 0} \frac{1}{|B(r)|} \int_{B(r)} |h(x,y) - h(x,0)|^p\, dy = 0 \quad \text{for almost every } x \in \mathbb{R}^d. \tag{37}$$

If the a_j expand nicely (as defined in Section 2), then parts (b) and (c) above hold with pointwise convergence a.e. in the limits (34) *and* (36) *(instead of L^p convergence), and we* *also have a pointwise analogue of* (33) *in part (a):*

$$\limsup_{j\to\infty} |(I_j[\psi,\phi]h)(x)| \leq |h(x,0)| \|P\psi\|_\infty \left| \int_{\mathbb{R}^d} \phi(z)\, dz \right| \qquad \text{for almost every } x. \tag{38}$$

Remarks on Lemma 7.

1. Hypothesis (32) says $y \mapsto h(\cdot\,, y)$ is continuous at $y = 0$, as a map $\mathbb{R}^d \to L^p$.

2. Conditions (i) and (i)$'$ are restricted to $p < \infty$ in this lemma because they *fail* to imply L^∞ convergence in Lemma 7(b), by the following counterexample. Take $\psi(x)$ to be the function defined in (126), in one dimension with $a_j = 2^j, b = 1$. This ψ has constant periodization $P\psi \equiv 1$. Let $\phi(x) = \mathbb{1}_{(0,1]}(x)$ and suppose $f \geq 0$ is continuous with compact support and with $f = 0$ on $[0, \infty)$ and $f \geq 1$ on $[-1/2, -1/4]$. Let $h(x, y) = f(x + y)$. For each $j \geq 2$, put $\ell = 2^{j-1}$ and consider $x \in (2^{-j-\ell-1}, 2^{-j-\ell}]$. Then $f(x) = 0$. By keeping only the term $k = -\ell$ in the sum defining $I_j[\psi,\phi]h$ we find

$$(I_j[\psi,\phi]h)(x) \geq \int_{-\ell}^{-\ell+1} f(2^{-j}y)\, dy \cdot \psi(2^j x + \ell) \geq 1$$

because $\psi(2^j x + \ell) = 1$ by the definition of ψ in (126) and $f(2^{-j}y) \geq 1$ because $2^{-j}y \in [-1/2, -1/2 + 2^{-j}]$. Thus $\|I_j[\psi,\phi]h - f\|_\infty \geq 1$ for all $j \geq 2$, and so (b) can fail when $p = \infty$, under condition (i) or (i)$'$.

3. Regarding the hypotheses in part (d), note $y \mapsto h(x, y)$ belongs to L^p_{loc} for almost every x, by Fubini's theorem applied to h. The hypothesis (37), saying that $y = 0$ is a Lebesgue point of $y \mapsto |h(x, y) - h(x, 0)|^p$ for almost every x, can be verified directly for the specific types of h used later in the paper, namely $h(x, y) =$

- $f(x + y)$,
- $f(x + y) - f(x)$,
- $\int_{[0,1]} f(x + ty)\, d\omega(t)$ where ω is a Borel probability measure defined on $[0, 1]$,
- $\int_{[0,1]} |f(x + ty) - f(x)|\, d\omega(t)$,

where $f \in L^p_{loc}$. The point is that almost every x is a p-Lebesgue point for $f \in L^p_{loc}$, meaning $\lim_{r\to 0} \int_{B(r)} |f(x + y) - f(x)|^p\, dy/|B(r)| = 0$, and for every such x one can show (37) holds when h has one of the types just mentioned.

Proof of Lemma 7. First we show $\phi \in L^1$, so that the integral of ϕ in the lemma does make sense. In case (i), $P|\phi|^\delta \in L^\infty$ is assumed. When $\delta = 1$ this means $P|\phi| \in L^\infty$, and integrating $P|\phi|$ over the set $b\mathcal{C}$ gives $\phi \in L^1$. When $\delta = 0$ it means ϕ has the "finite intersection" property $P|\phi|^0 \in L^\infty$, which implies the measure of $\{\phi \neq 0\}$ is finite so that $\phi \in L^q$ implies $\phi \in L^1$ by Hölder's inequality. In case (i)$'$, if $\delta = 1$ then $\phi \in L^1$ is immediate, while if $\delta = 0$ then we have $\phi \in L^q$ and $|\phi|^0 \in L^1$, so that $\phi \in L^1$ by Hölder again. In cases (ii) and (iii) it is easy to see $\phi \in L^1$, by the compact support assumption.

Observe $\psi \in L^1$ by similar arguments (it is immediate when $p = 1$, of course), and noting for case (iii) that $Q\psi \in L^1$ implies $\psi \in L^1$ by Lemma 22.

Next, $P\psi \in L^p_{loc}$ as follows. If $p = 1$ then $\psi \in L^1$ and so $P\psi \in L^1_{loc}$ by Lemma 18. If $1 < p < \infty$ then $P|\psi|^\varepsilon \in L^\infty$ is assumed in cases (i), (i)$'$ and (ii). When $\varepsilon = 1$ this means $P|\psi| \in L^\infty \subset L^p_{loc}$. When $\varepsilon = 0$ it means $P|\psi|^0 \in L^\infty$, and this finite intersection property together with $P(|\psi|^p) \in L^1_{loc}$ (since $|\psi|^p \in L^1$) gives $|P\psi|^p \in L^1_{loc}$, or $P\psi \in L^p_{loc}$ as we wanted. Lastly if $p = \infty$ then case (iii) assumes $Q\psi \in L^1$ and so $P\psi \in L^\infty$ by Lemma 23.

The mean value of $P\psi$ equals $\int_{\mathbb{R}^d} \psi(y)\,dy$ by the calculation (6), and so we can construct the sequence $j(n)$ claimed in part (c) as follows. The $b\mathbb{Z}^d$-periodic function $g(x) = (P\psi)(x) - \int_{\mathbb{R}^d} \psi(y)\,dy$ has mean value zero and belongs to L^p_{loc}. So if $1 \le p < \infty$ and $J \in \mathbb{Z}$ is given, then Lemma 5 yields a strictly increasing integer sequence $j(n) \ge J$ such that the averaging relation (35) holds. Clearly this sequence $j(n)$ is independent of h and ϕ, and depends only on ψ, p and J (and of course on the translations and dilations: b and the a_j).

Note for later use. If ψ_0 and ψ_1 are two functions with $P\psi_0, P\psi_1 \in L^p_{loc}$, then a strictly increasing integer sequence exists with $j(n) \ge J$ such that (35) holds for both ψ_0 and ψ_1. The point here is that Lemma 5 first yields a sequence whose every subsequence satisfies (35) for ψ_0, then Lemma 5 can be applied again to obtain a particular subsequence for which (35) also holds for ψ_1.

With these preliminaries taken care of, we begin to prove parts (a)–(d).

Part (a).

Case (i). The estimate (25) gives

$$
\begin{aligned}
\|I_j[\psi,\phi]h\|_p^p &\le \int_{\mathbb{R}^d} |f(y)|^p (P|\phi|^\delta)(a_j y)\,dy\, \|\psi\|_{p(\varepsilon)}^{p(\varepsilon)} \cdot \|P|\psi|^\varepsilon\|_\infty^{p-1} \|\phi\|_{q(\delta)}^{p-\delta} \\
&\to \int_{\mathbb{R}^d} |f(y)|^p \| |\phi|^\delta \|_1\,dy\, \|\psi\|_{p(\varepsilon)}^{p(\varepsilon)} \cdot \|P|\psi|^\varepsilon\|_\infty^{p-1} \|\phi\|_{q(\delta)}^{p-\delta} \qquad \text{as } j \to \infty,
\end{aligned}
$$

by the Riemann–Lebesgue Lemma 26 applied with "$p = \infty$" and $g = P(|\phi|^\delta) - \||\phi|^\delta\|_1$ (which is bounded and has mean value zero) and $h = |f|^p \in L^1$. This proves (33) in case (i).

Case (i)$'$. The estimate (28) implies that R_j is bounded by a constant independent of x, y, t and j, for all large j (using that $\|a_j^{-1}\| \to 0$). Since also $R_j(x,y,t) \to \int_{\mathbb{R}^d} |f(z)|^p\,dz$ for each x, y, t as $j \to \infty$ (by interpreting the definition of R_j in (27) as a Riemann sum, and using $f \in C_c$), we may apply dominated convergence to formula (26) to obtain that

$$
\limsup_{j\to\infty} \|I_j[\psi,\phi]h\|_p^p \le \int_{[0,1]} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(z)|^p\,dz\, |\psi(x)|^{p(\varepsilon)} |\phi(y)|^\delta\,dx dy d\omega(t) \|P|\psi|^\varepsilon\|_\infty^{p-1} \|\phi\|_{q(\delta)}^{p-\delta},
$$

which proves the estimate (33) in case (i)$'$.

Case (ii). By dominated convergence, as $j \to \infty$ the righthand side of (31) approaches the limiting value $\int_{\mathbb{R}^d} \|h(\cdot\,,0)\|_p^p (\widetilde{Q}|\psi|^{p(\varepsilon)})(y)\,dy \cdot \|P|\phi|^0\|_\infty \|P|\psi|^\varepsilon\|_\infty^{p-1} \|\phi\|_q^p$, because $\widetilde{Q}|\psi|^{p(\varepsilon)} \in L^1$ and $h \in L^{(p,\infty)}$ while $h(\cdot, y) \to h(\cdot, 0)$ in L^p as $y \to 0$ by assumption and $a_j^{-1} y \to 0$ since the a_j are expanding. This proves (33) in case (ii), since we can now replace $\widetilde{Q}$ with Q like we did after (31).

Case (iii). When $p=\infty$ we argue directly:

$$
\begin{aligned}
&|(I_j[\psi,\phi]h)(x)| \\
&\leq |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |h(x,a_j^{-1}y-x)||\phi(y-bk)|\,dy\,|\psi(a_jx-bk)| \\
&\leq |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |h(x,a_j^{-1}y-x)||\phi(y-bk)|(\widetilde{Q}\psi)(a_jx-y)\,dy \\
&\qquad\qquad\qquad \text{for almost every } x, \text{ by (125) in Lemma 22 with } E=\operatorname{spt}(\phi), \\
&\leq \int_{\mathbb{R}^d} \|h(\cdot\,,-a_j^{-1}y)\|_\infty |(\widetilde{Q}\psi)(y)|\,dy\cdot \|P|\phi|\|_\infty \qquad \text{by } y\mapsto a_jx-y.
\end{aligned}
$$

The righthand side of this last inequality is independent of x, and as $j\to\infty$ it approaches $\int_{\mathbb{R}^d}\|h(\cdot\,,0)\|_\infty(\widetilde{Q}\psi)(y)\,dy\cdot\|P|\phi|\|_\infty$ by dominated convergence. Note $\|\widetilde{Q}\psi\|_1\leq C(E)\|Q\psi\|_1$ by the definition of $\widetilde{Q}$ in (125), and this completes the proof of (33).

Before considering parts (b) and (c) of the lemma, we detour to prove (33) for a useful variant of h from case (i)$'$.

Lemma 8. *Take $\varepsilon,\delta\in\{0,1\}$. Assume*

(i)$'$ *$1\leq p<\infty$ and $\psi\in L^p, (p-1)P(|\psi|^\varepsilon)\in L^\infty$, $\phi\in L^{q(\delta)}$ with $|\phi|^\delta\in L^1$, and $H^*(x,y)=\int_{[0,1]}|f(x+ty)-f(x)|\,d\omega(t)$ for some $f\in C_c$ and some Borel probability measure ω on $[0,1]$.*

Then clearly $H^*(x,0)=0$, and $\lim_{j\to\infty}\|I_j[\psi,\phi]H^*\|_p=0$.

Proof of Lemma 8. We have

$$
\|I_j[\psi,\phi]H^*\|_p^p \leq \int_{[0,1]}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} R_j^*(x,y,t)\,|\psi(x)|^{p(\varepsilon)}|\phi(y)|^\delta\,dxdyd\omega(t)\cdot\|P|\psi|^\varepsilon\|_\infty^{p-1}\|\phi\|_{q(\delta)}^{p-\delta} \quad (39)
$$

by applying (26) to H^* instead of to h, where

$$
R_j^*(x,y,t)=|\det a_j^{-1}b|\sum_{k\in\mathbb{Z}^d}|f(a_j^{-1}(x+bk)+ta_j^{-1}(y-x))-f(a_j^{-1}(x+bk))|^p.
$$

Clearly $R_j^*(x,y,t)$ is a Riemann sum, converging pointwise to $\int_{\mathbb{R}^d}|f(z)-f(z)|^p\,dz=0$ as $j\to\infty$, since $f\in C_c$ and $\|a_j^{-1}\|\to 0$. And by the triangle inequality and the proof of Lemma 7(a) in case (i)$'$ above, one finds $R_j^*(x,y,t)$ is bounded by a constant independent of x,y,t,j, for all large j. Thus dominated convergence applied to (39) gives $\|I_j[\psi,\phi|]H^*\|_p\to 0$ as $j\to\infty$. □

Now we return to proving Lemma 7.

Parts (b) and (c). The existence of the sequence $j(n)$ satisfying (35) was established at the beginning of this proof.

Define

$$
H(x,y)=h(x,y)-h(x,0)\in L^{(p,\infty)}.
$$

Then the definition of I_j in (21) implies

$$
(I_j[\psi,\phi]h)(x)=(I_j[\psi,\phi]H)(x)+h(x,0)(P\psi)(a_jx)\int_{\mathbb{R}^d}\phi(z)\,dz. \quad (40)
$$

Case (i). Assume case (i) holds, with $h(x,y) = f(x+y)$. For proving parts (b) and (c) of the lemma we need only consider $f \in C_c$, by the stability estimate $\|I_j[\psi,\phi]h\|_p \leq C\|h\|_{(p,\infty)} = C\|f\|_p$ proved in case (i) of Lemma 6, and in view of the density of C_c in L^p for $1 \leq p < \infty$.

But when $f \in C_c$, case (i) is covered by case (i)′ below (taking ω to be a delta mass at $t=1$ and noting that case (i) implies $\phi \in L^1 \cap L^q$ with $|\phi|^\delta \in L^1$).

Case (i)′. Suppose $h(x,y) = \int_{[0,1]} f(x+ty)\,d\omega(t)$ for some $f \in C_c$ and some Borel probability measure ω, so that $H(x,y) = \int_{[0,1]}[f(x+ty) - f(x)]\,d\omega(t)$. Then

$$\lim_{j\to\infty} I_j[\psi,\phi]H = 0 \qquad \text{in } L^p \tag{41}$$

by Lemma 8, since $|I_j[\psi,\phi]H| \leq I_j[|\psi|,|\phi|]H^*$ pointwise and $|\psi|,|\phi|$ and f satisfy the hypotheses of Lemma 8.

Now to prove part (b) of the lemma, observe if $(P\psi)(x) = \int_{\mathbb{R}^d}\psi(y)\,dy$ for almost every x that the desired limit (34) now follows immediately from (41) and decomposition (40).

And for part (c) we just use (41) and (40) and observe that

$$\lim_{N\to\infty} h(x,0)\frac{1}{N}\sum_{n=1}^{N}(P\psi)(a_{j(n)}x) = h(x,0)\int_{\mathbb{R}^d}\psi(y)\,dy \qquad \text{in } L^p,$$

by the boundedness and compact support of $h(x,0) = f(x) \in C_c$ and using the L^p_{loc} convergence of the periodizations in (35).

Cases (ii) and (iii). In these cases $\lim_{j\to\infty} I_j[\psi,\phi]H = 0$ in L^p by part (a) of the lemma, because $H \in L^{(p,\infty)}$ and $H(\cdot\,,y) \to H(\cdot\,,0) = 0$ as $y \to 0$ by hypothesis (32).

So part (b) of the lemma again follows from the decomposition (40).

Part (c) assumes $p < \infty$ which rules out case (iii), so assume case (ii) holds. Part (c) follows like in the proof of part (i)′ above if $h(x,0)$ is bounded with compact support. But we can reduce to this situation by the stability estimate $\|I_j[\psi,\phi]h\|_p \leq C\|h\|_{(p,\infty)}$ proved in case (ii) of Lemma 6, in view of the following approximation to h. Given $\epsilon > 0$, choose $\tilde{h} \in C_c$ with $\|h(\cdot\,,0) - \tilde{h}\|_p < \epsilon$, and then define

$$h_\epsilon(x,y) = \begin{cases} \tilde{h}(x) & \text{if } |y| \leq \epsilon, \\ h(x,y) & \text{otherwise.} \end{cases}$$

Then trivially $h_\epsilon(\cdot\,,y) \to h_\epsilon(\cdot,0)$ in L^p as $y \to 0$, while

$$\|h - h_\epsilon\|_{(p,\infty)} \leq \max_{|y|\leq\epsilon}\|h(\cdot\,,y) - h(\cdot\,,0)\|_p + \|h(\cdot\,,0) - \tilde{h}\|_p \to 0$$

as $\epsilon \to 0$. That is, we can approximate h arbitrarily closely in $L^{(p,\infty)}$ by a function satisfying the same hypotheses as h but which is also bounded with compact support at $y=0$.

Part (d). First we show $I_j[\psi,\phi]h$ is well defined. Our hypotheses in part (d) ensure $\psi \in L^\infty$ has a bounded radially decreasing L^1-majorant. Hence $P|\psi| \in L^\infty$ by Lemma 19, and $Q\psi$ has a bounded radially decreasing L^1-majorant by Lemma 21.

The hypotheses in part (d) further specify that either $h \in L^{(p,\infty)}$ or $h \in L^{(\infty,\infty)}$. If $h \in L^{(p,\infty)}$ then case (ii) in Lemma 6 is satisfied with $\varepsilon = 1$, while if $h \in L^{(\infty,\infty)}$ then case

(iii) in Lemma 6 is satisfied with $\varepsilon = 1$. Thus in any event, Lemma 6 guarantees that the series defining $I_j[\psi, \phi]h$ converges pointwise absolutely a.e., to a function in either L^p or L^∞.

Next we examine the pointwise limit of $I_j[\psi, \phi]H$ as $j \to \infty$. Write $\widetilde{Q}\psi$ for the function defined from finitely many translates of $Q\psi$ in Lemma 22, for the bounded set $E = \mathrm{spt}(\phi)$. Then $\widetilde{Q}\psi$ also has a bounded radially decreasing L^1-majorant by an easy argument. That is, $\widetilde{Q}\psi \leq \eta$ for some bounded radially decreasing function $\eta(|x|) \in L^1$. If ψ has compact support then η can be taken to have compact support, and if ψ does not have compact support then the hypotheses of part (d) tell us $h \in L^{(\infty,\infty)}$.

For almost every y, the majorant estimate gives that

$$\begin{aligned}(\widetilde{Q}\psi)_{a_j^{-1}}(y) &= |\det a_j|(\widetilde{Q}\psi)(a_j y) \\ &\leq |\det a_j|\eta(|a_j y|) \\ &\leq C\lambda_j^d \eta(\lambda_j|y|) \qquad \text{by (2) and (4), since the } a_j \text{ expand nicely,} \\ &= C\eta_{\epsilon_j}(|y|) \end{aligned} \tag{42}$$

where $\epsilon_j = \lambda_j^{-1} \to 0$ as $j \to \infty$. Now we can prove $\lim_{j\to\infty}(I_j[\psi, \phi]H)(x) = 0$ a.e. Indeed

$$\begin{aligned}|(I_j[\psi, \phi]H)(x)|^p &\leq C \int_{\mathbb{R}^d} |H(x, -a_j^{-1}y)|^p (\widetilde{Q}\psi)(y)\, dy \qquad \text{by (30) applied to } H \text{ with } \varepsilon = 1, \\ &= C \int_{\mathbb{R}^d} |H(x, y)|^p (\widetilde{Q}\psi)_{a_j^{-1}}(-a_j y)\, dy \qquad \text{by } y \mapsto -a_j y \\ &\leq C \int_{\mathbb{R}^d} |H(x, y)|^p \eta_{\epsilon_j}(|y|)\, dy \qquad \text{by (42)} \\ &\to C|H(x, 0)|^p \int_{\mathbb{R}^d} \eta(|y|)\, dy = 0 \qquad \text{as } j \to \infty \end{aligned} \tag{43}$$

whenever $y = 0$ is a Lebesgue point for $y \mapsto |H(x, y)|^p$, that is whenever (37) holds. (This limit (43) is by a standard result [42, Theorem I.1.25] on pointwise convergence at Lebesgue points, for approximations to the identity. The result applies directly if $h \in L^{(\infty,\infty)}$ because then $y \mapsto |H(x, y)|^p$ belongs to L^∞. The result can be modified to apply when ψ has compact support and $h \in L^{(p,\infty)}$ because then η has compact support, which compensates for the function $y \mapsto |H(x, y)|^p$ belonging only to L^1_{loc}.)

Hence decomposition (40) implies (38), the desired pointwise estimate for part (a).

Pointwise convergence in (34) for part (b) is similarly immediate from the decomposition (40) and the assumption that $(P\psi)(x) = \int_{\mathbb{R}^d} \psi(y)\, dy$ for almost every x.

For proving pointwise convergence in part (c) we first recall the $b\mathbb{Z}^d$-periodic function $g(x) = (P\psi)(x) - \int_{\mathbb{R}^d} \psi(y)\, dy$ has mean value zero and belongs to L^p_{loc} (since $P\psi \in L^\infty$ here in part (d)). Hence Lemma 5 yields a sequence $j(n)$ such that the averaging relation (35) holds, in particular with pointwise convergence. Pointwise convergence in (36) now follows from decomposition (40). □

6. Sampling in L^p

In Sections 6.1 and 6.2 we develop average and pointwise sampling formulas for L^p. Then in Section 6.3 we simplify the sequence $j(n)$ over which our sampling formulas are taken, supposing either ψ has bounded variation and the dilations grow exponentially, or else that the Fourier transform satisfies $\sum_{\ell\in\mathbb{Z}^d} |\widehat{\psi}(\ell)| < \infty$ and the dilations are integer matrices.

6.1. **Average sampling.** The next proposition combines the basic sampling formula from Lemma 7 for $h(x, y) = f(x + y)$ with the stability and spanning properties from Lemma 6. For simplicity we denote $f_j = I_j[\psi, \phi]h$. In other words

$$f_j(x) = |\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} f(a_j^{-1}y)\phi(y - bk)\, dy \right) \psi_{j,k}(x), \qquad j \in \mathbb{Z}. \tag{44}$$

Recall $\frac{1}{p} + \frac{1}{q} = 1$ and that $p(\varepsilon) = \varepsilon + (1 - \varepsilon)p$ and $q(\delta) = \delta + (1 - \delta)q$. Write $|\phi|^0 = \mathbb{1}_{\{\phi\neq 0\}}$. Obviously if ϕ has compact support then $P(|\phi|^0) \in L^\infty$.

Proposition 9. *Take $\varepsilon, \delta \in \{0, 1\}$. Assume one of the following conditions holds:*

(i) *$1 \le p < \infty$ and $\psi \in L^p, (p-1)P(|\psi|^\varepsilon) \in L^\infty, \phi \in L^q, P(|\phi|^\delta) \in L^\infty, f \in L^p$;*

(iii) *$p = \infty$ and $\psi \in L^\infty, Q\psi \in L^1$ and $\phi \in L^\infty$ with ϕ having compact support, and $f \in L^\infty$.*

Assume $\int_{\mathbb{R}^d} \psi\, dx = 1$ and $\int_{\mathbb{R}^d} \phi\, dx = 1$. Then (a)–(e) hold:

(a) [Constant periodization sampling] If $P\psi = 1$ a.e. then

$$f = \lim_{j\to\infty} f_j \qquad \text{in } L^p, \tag{45}$$

provided when $p = \infty$ we also assume f is uniformly continuous.

(b) [Dilation-averaged sampling] Suppose $1 \le p < \infty$ and $J \in \mathbb{Z}$. A strictly increasing integer sequence $\{j(n)\}_{n=1}^\infty$ exists (independent of f) such that $j(1) \ge J$ and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (P\psi)(a_{j(n)}x) = 1 \qquad \text{in } L^p_{loc} \text{ and pointwise a.e.} \tag{46}$$

For any such sequence,

$$f = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_{j(n)} \qquad \text{in } L^p. \tag{47}$$

(c) [Stability] $\|f_j\|_p \le C(\psi, \phi, p, \varepsilon, \delta)\|f\|_p$ for all $j \in \mathbb{Z}$.

(d) [Spanning] The series (44) defining f_j converges pointwise absolutely a.e. to an L^p function. It also converges unconditionally in L^p (provided when $p = \infty$ we know $f \in L^1$).

Hence $f_j \in L^p$-span$(\{\psi_{j,k} : k \in \mathbb{Z}^d\})$ if $1 \le p < \infty$ or if $p = \infty$ and $f \in L^1$.

(e) [Pointwise convergence] Instead of (i) or (iii), assume $1 \le p < \infty$ and $\psi \in L^1 \cap L^\infty$ and suppose $\phi \in L^q$ has compact support. Suppose either ψ has compact support and $f \in L^p$, or else ψ has a radially decreasing L^1-majorant and $f \in L^\infty$.

If the a_j expand nicely, then parts (a) and (b) above hold with pointwise convergence a.e. in the limits (45) and (47) (instead of L^p convergence).

Proof of Proposition 9. Let $h(x, y) = f(x + y)$, so that $h \in L^{(p,\infty)}$ with $\|h\|_{(p,\infty)} = \|f\|_p$.

Proposition 9(a)(b)(e) follows from Lemma 7(b)(c)(d), since $f_j = I_j[\psi, \phi]h$ and $\int_{\mathbb{R}^d} \psi(y)\, dy = 1$ and $\int_{\mathbb{R}^d} \phi(z)\, dz = 1$ by assumption. Note for part (a) that when $p = \infty$, the assumed uniform continuity of f ensures $\lim_{y\to 0} h(\cdot\,, y) = h(\cdot\,, 0)$ in L^∞.

To prove Proposition 9(c), just call on the stability estimates in Lemma 6, noting in case (iii) here that $Q\psi \in L^1$ implies $P|\psi| \in L^\infty$ by Lemma 23.

For Proposition 9(d), refer to the pointwise and unconditional L^p convergence in Lemma 6. Then the L^p convergence of the series for f_j implies $f_j \in L^p\text{-span}(\{\psi_{j,k} : k \in \mathbb{Z}^d\})$. $\square$

Remarks on Proposition 9.

1. In part (a), the constant periodization hypothesis $P\psi \equiv 1$ can be restated in terms of zeros of the Fourier transform of ψ; see Section 3.3.

2. If we are given two functions ψ_0 and ψ_1 satisfying the requirements of case (i) in Proposition 9, when $1 \le p < \infty$, then the sequence $j(n)$ in part (b) can be chosen to satisfy the averaging relation (46) for both ψ_0 and ψ_1 simultaneously. *Proof:* See the Note in the proof of Lemma 7.

3. The analyzing and synthesizing roles of ϕ and ψ are clarified by rewriting f_j as

$$f_j = \sum_{k\in\mathbb{Z}^d} \langle f, \phi^*_{j,k}\rangle \psi^*_{j,k} \tag{48}$$

where $\psi^*_{j,k}(x) = |\det a_j b|^{1/p}\psi(a_j x - bk)$ and $\phi^*_{j,k}(y) = |\det a_j b|^{1/q}\overline{\phi(a_j y - bk)}$ are normalized to have L^p and L^q norms independent of j, respectively. The coefficients in (48) are stable (meaning their ℓ^p-norm is controlled by the L^p-norm of f) because Hölder's inequality gives

$$\begin{aligned}\left(\sum_{k\in\mathbb{Z}^d} \left|\langle f, \phi^*_{j,k}\rangle\right|^p\right)^{1/p} &= |\det a_j b|^{1/q}\left(\sum_{k\in\mathbb{Z}^d}\left|\int_{\mathbb{R}^d} f(y)\phi(a_j y - bk)\,dy\right|^p\right)^{1/p} \\ &\le |\det b|^{1/q}\left(\sum_{k\in\mathbb{Z}^d}\int_{\mathbb{R}^d}|f(y)|^p|\phi(a_j y - bk)|^\delta\,dy\right)^{1/p}\|\phi\|_{q(\delta)}^{1-\delta/p} \\ &\le |\det b|^{1/q}\|P|\phi|^\delta\|_\infty^{1/p}\|\phi\|_{q(\delta)}^{1-\delta/p}\|f\|_p = C(p,\phi)\|f\|_p\end{aligned}$$

in case (i), while in case (iii) we get the ℓ^∞-estimate $\left|\langle f, \phi^*_{j,k}\rangle\right| \le |\det b|\|\phi\|_1\|f\|_\infty$.

4. Proposition 14 parts (b) and (d) are used in proving our L^p spanning result, Theorem 1.

L^p sampling literature relevant to Proposition 9 (and Proposition 10). As discussed earlier in "Remarks on the L^p spanning literature" (after Theorem 1), the best previous affine approximation formulas for L^p are of Strang–Fix type under the assumptions that ψ has constant periodization $P\psi \equiv 1$, that ψ has compact support (or decays at a polynomial rate), and that the dilations are expanding and isotropic ($a_j = \lambda_j I$). See the references in that earlier section.

These approximation formulas have the form $f = \lim_{j\to\infty}\sum_{k\in\mathbb{Z}^d} c_{j,k}\psi(\lambda_j x - bk)$, which is similar to our "constant periodization" approximation $f = \lim_{j\to\infty} f_j$ in Proposition 9(a). However our f_j is defined explicitly in terms of sampled averages of f, while the coefficients $c_{j,k}$ are not explicit (when $1 \le p < \infty$) because the authors proceed on the Fourier transform side and use sampled values of $\hat{f}$ rather than of f itself. The only exceptions seem to be di Guglielmo and Jia–Lei. Di Guglielmo [23, Théorème 2′] essentially proved Proposition 9(a)(c) for $p \ge 2$ when ψ, ϕ are bounded with compact support and $f \in W^{1,p}$ and the dilation a_j is a diagonal matrix. (Unfortunately di Guglielmo's ψ has a special convolution form as discussed in Section 3.2, except in the L^2 sampling formula [23, Théorème 5] which assumes only constant periodization and a vanishing moment condition on ψ, ϕ.) And Jia and Lei [27, Theorem 4.1] proved something like Proposition 9(a)(c) for $1 \le p \le \infty$ when ψ is bounded and decays at infinity and has constant periodization, ϕ is smooth with compact support, $f \in W^{1,p}$ and $a_j = \lambda_j I$.

In contrast, our sampling formulas in Proposition 9 are explicitly in terms of sampled average values of f. Later, in Proposition 10, we even sample pointwise values of f, leading for example to the sampling formula (1) stated in the Introduction for ψ and f of bounded variation. In both these propositions, the flexibility of the sampling method (*i.e.* the choice of ϕ) and the generality of ϕ and ψ (*e.g.* ψ having only bounded periodization instead of constant periodization, in Proposition 9(b)) are much greater than in previous works. The situation is better still when $p = 1$, for then case (i) of the propositions requires no periodization bound on ψ whatsoever.

The dilation-averaged sampling technique in Proposition 9(b) seems qualitatively new.

Another point of interest is that our L^p sampling formulas also converge pointwise a.e. This seems to be new for $1 \le p < \infty$.

When $p = \infty$ (uniform approximation) our contribution is less because part (a) of the proposition assumes constant periodization, and explicit L^∞ sampling formulas in this situation go back to Strang and Fix [44, Theorem III] and earlier to Schoenberg [41, Theorem 2], for example. See also the next paragraph. Thus the average and pointwise sampling formulas in Propositions 9(a) and 10(a) are new mainly in their technical details, when $p = \infty$, for we are requiring just that $Q\psi \in L^1$ and that $f \in L^\infty$ be uniformly continuous.

Finally we note two developments in sampling theory which are relevant to Proposition 9 and Proposition 10. In the case $p = \infty$, Butzer and his group obtained a sampling formula for bounded and uniformly continuous functions on $\mathbb{R}$, under the assumptions that $P\psi \equiv 1$ (constant periodization) and $P|\psi|(x) = \sum_{k \in \mathbb{Z}} |\psi(x-k)|$ converges uniformly on $[0,1]$ (note that the last condition is stronger than $P|\psi| \in L^\infty$). Their sampling formula uses pointwise values of the sampled function and also converges in the $L^\infty(\mathbb{R})$ norm. Their result is comparable to our Proposition 10(a) (for $d = 1, p = \infty, z_j(k) = k$), and moreover, these authors also proved their sampling formula at each point of continuity. We refer to the survey articles [13] and [14] (and the references therein) for the research in this direction.

The second development concerns sampling algorithms in a special class of closed subspaces of $L^p, 1 \le p < \infty$. These subspaces are spanned by a Riesz basis or an L^p frame, which consists of integer translates of one function or a finite number of functions. The iterative sampling algorithm, obtained by Aldroubi and Feichtinger, and by Aldroubi, Sun and Tang, involves a projection operator and sampled values ([1]) or sampled averages of the function ([3]) in these subspaces.

More literature on sampling theory can be found in the survey articles [2], [8] and [46].

Note. The authors in the Strang-Fix tradition mentioned above all proved precise approximation *rates*, approximating a $W^{m,p}$ function at rate $O(\|a_j^{-1}\|^m)$ in the L^p norm. See Proposition 17 with $r = 0$ for our analogous result.

6.2. **Pointwise sampling.** A pointwise sampling formula can be formally obtained by taking ϕ to be a delta mass at the origin in Proposition 9, so that f_j samples the values of f at points $a_j^{-1}bk$ for $k \in \mathbb{Z}^d$. We make this pointwise sampling idea rigorous in the next proposition.

Fix $J \in \mathbb{Z}$ and choose "sampling points" $z_j(k) \in \mathbb{R}^d$ for each $j \geq J, k \in \mathbb{Z}^d$, and require them to stay near the integer lattice points in the sense that

$$\Delta := \sup_{j \geq J, k \in \mathbb{Z}^d} \sup_{y \in k + \mathcal{C}} |z_j(k) - y| + \sqrt{d} < \infty.$$

(The choice $z_j(k) = k$ gives uniform sampling.)

Call x a *partial Lebesgue point* for f if there exists a sequence $\{X_m\}_{m=1}^{\infty}$ of measurable sets shrinking to x (each with positive measure) such that $|X_m|^{-1} \int_{X_m} |f(y) - f(x)| \, dy \to 0$ as $m \to \infty$. For instance, for the characteristic function $f = \mathbb{1}_{[0,1]}$ in one dimension, the jump point at $x = 0$ is a partial Lebesgue point by using intervals $X_m = [0, 1/m]$ to the right of $x = 0$. Points of continuity are automatically partial Lebesgue points.

We introduce the function

$$f_j^{\bullet}(x) = |\det b| \sum_{k \in \mathbb{Z}^d} f(a_j^{-1} b z_j(k)) \psi_{j,k}(x), \qquad j \geq J, \tag{49}$$

with the "$\bullet$" indicating the pointwise nature of the sampling.

Proposition 10. *Take $\varepsilon \in \{0, 1\}$. Let $J \in \mathbb{Z}$ and consider sampling points $z_j(k)$ as above. Assume one of the following conditions holds:*

(i) $1 \leq p < \infty, \psi \in L^p, (p-1)P(|\psi|^{\varepsilon}) \in L^{\infty}$;

(iii) $p = \infty, \psi \in L^{\infty}, Q\psi \in L^1$.

Assume $\int_{\mathbb{R}^d} \psi \, dx = 1$. Let $f \in L^p$ with $Qf \in L^p$, f continuous a.e. and with $x = a_j^{-1} b z_j(k)$ being a partial Lebesgue point for f, for each $j \geq J, k \in \mathbb{Z}^d$. Then (a)–(e) hold:

(a) [Constant periodization sampling] If $P\psi = 1$ a.e. then

$$f = \lim_{j \to \infty} f_j^{\bullet} \qquad \text{in } L^p, \tag{50}$$

provided when $p = \infty$ we also assume f is uniformly continuous.

(b) [Dilation-averaged sampling] Suppose $1 \leq p < \infty$. A strictly increasing integer sequence $\{j(n)\}_{n=1}^{\infty}$ exists (independent of f) such that $j(1) \geq J$ and (46) holds. For any such sequence,

$$f = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f_{j(n)}^{\bullet} \qquad \text{in } L^p. \tag{51}$$

(c) [Stability] $\|f_j^{\bullet}\|_p \leq C(\psi, p, \varepsilon, J, \Delta) \|Qf\|_p$ for all $j \geq J$.

(d) [Spanning] The series (49) defining $f_j^{\bullet}$ converges pointwise absolutely a.e. to an L^p function. It also converges unconditionally in L^p (provided when $p = \infty$ we know $Qf \in L^1$).

Hence $f_j^{\bullet} \in L^p$-span$(\{\psi_{j,k} : k \in \mathbb{Z}^d\})$ if $1 \leq p < \infty$ or if $p = \infty$ and $Qf \in L^1$.

(e) [Pointwise convergence] Instead of (i) or (iii), assume $1 \leq p < \infty, \psi \in L^1 \cap L^{\infty}$, and ψ has a radially decreasing L^1-majorant.

If the a_j expand nicely, then parts (a) and (b) above hold with pointwise convergence a.e. in the limits (50) and (51) (instead of L^p convergence).

Remarks on Proposition 10.

1. The assumption $Qf \in L^p$ in case (i) means that at any two nearby points, the values of f are in some sense close. This allows us to estimate $f_j^{\bullet}$ with f_j, hence reduce from pointwise sampling to the average sampling already treated in Proposition 9. See estimate (54) in the proof below.

2. Regarding the stability statement in part (c) for $p = \infty$, notice $\|Qf\|_\infty = \|f\|_\infty$.

3. The stability and spanning properties of $f_j^\bullet$ in parts (c) and (d) of the proposition depend ultimately on Lemma 6, and thus they hold whether or not $\int_{\mathbb{R}^d} \psi \, dx = 1$.

4. See remarks on the sampling literature, earlier in this section.

Proof of Proposition 10. For each invertible $d \times d$ real matrix a, we define measurable functions

$$(Q_a f)(x) = \|f\|_{L^\infty(B(x,\|a\|\Delta))}, \quad (S_a f)(x) = \|f(x) - f(\cdot)\|_{L^\infty(B(x,\|a\|\Delta))}.$$

(The operators Qf and Sf in Appendix B come from choosing $a = (\sqrt{d}/\Delta)I$.)

We claim $Q_a f, S_a f \in L^p$. This is obvious when $p = \infty$, because $\|Q_a f\|_\infty \leq \|f\|_\infty$ and $\|S_a f\|_\infty \leq 2\|f\|_\infty$. When $1 \leq p < \infty$ we will show $Qf \in L^p$ implies $Q_a f, S_a f \in L^p$. First, $S_a f \leq |f| + Q_a f \leq 2Q_a f$ a.e. by arguing like in Lemma 22. Second, given the matrix a we let $\mathcal{K}(a)$ be a finite collection of lattice points $k \in \mathbb{Z}^d$ such that $B(0, \|a\|\Delta)$ lies in the union of the balls $B(k, \sqrt{d})$ for $k \in \mathcal{K}(a)$. Then

$$(S_a f)(x) \leq 2(Q_a f)(x) \leq 2 \sum_{k \in \mathcal{K}(a)} (Qf)(x + k), \tag{52}$$

which belongs to L^p as desired because $Qf \in L^p$ by assumption. So $Q_a f, S_a f \in L^p$.

For the matrix $a = a_j^{-1} b$, the set $\mathcal{K}(a_j^{-1} b)$ consists of just the origin when j is large enough, because $\|a_j^{-1} b\| \to 0$ by the expanding property of the dilations a_j.

<u>Part (d).</u> Let $\phi(x) = \mathbb{1}_{b\mathcal{C}}(x)/|b\mathcal{C}|$ so that $\int_{\mathbb{R}^d} \phi \, dx = 1$. For each $j \geq J, k \in \mathbb{Z}^d$,

$$\begin{aligned} \left| f(a_j^{-1} b z_j(k)) - \int_{\mathbb{R}^d} f(a_j^{-1} y) \phi(y - bk) \, dy \right| &\leq \int_{\mathbb{R}^d} |f(a_j^{-1} b z_j(k)) - f(a_j^{-1} y)| \phi(y - bk) \, dy \\ &\leq \int_{\mathbb{R}^d} (S_{a_j^{-1} b} f)(a_j^{-1} y) \phi(y - bk) \, dy \end{aligned} \tag{53}$$

since $|a_j^{-1} b z_j(k) - a_j^{-1} y| \leq \|a_j^{-1} b\| |z_j(k) - b^{-1} y| < \|a_j^{-1} b\| \Delta$ for all $y - bk \in \operatorname{spt}(\phi) = b\mathcal{C}$, by definition of Δ and because $a_j^{-1} b z_j(k)$ is a partial Lebesgue point for f.

Writing $T_j(x, y) = (S_{a_j^{-1} b} f)(x + y)$, we find

$$\begin{aligned} \left| f_j^\bullet(x) - f_j(x) \right| &= \left| |\det b| \sum_{k \in \mathbb{Z}^d} \left[f(a_j^{-1} b z_j(k)) - \int_{\mathbb{R}^d} f(a_j^{-1} y) \phi(y - bk) \, dy \right] \psi_{j,k}(x) \right| \\ &\leq (I_j[|\psi|, \phi] T_j)(x) \qquad \text{by (53).} \end{aligned} \tag{54}$$

Recall from Proposition 9(d) that the series defining f_j converges pointwise absolutely a.e. to an L^p function, and converges unconditionally in L^p (provided $f \in L^1$ when $p = \infty$).

The series defining $I_j[|\psi|, \phi] T_j$ converges similarly by Lemma 6 cases (i), (iii) and (iv) (except that for unconditional convergence when $p = \infty$ we should assume $Qf \in L^1$, which ensures $f \in L^1$ and $S_a f \in L^1$ by (52)). The same convergence properties must hold for the series defining $f_j^\bullet$, in view of estimate (54). This proves part (d) of the proposition, and in particular shows $f_j^\bullet \in L^p$.

<u>Parts (a),(b).</u> Proposition 9(a) gives L^p convergence of f_j to f, in part (a). Proposition 9(b) gives existence of the sequence $j(n)$ for part (b), and gives L^p convergence of $\frac{1}{N} \sum_{n=1}^N f_{j(n)}$ to f , in part (b).

Thus for parts (a) and (b), by (54) it is enough to show $\lim_{j\to\infty} I_j[|\psi|,\phi]T_j = 0$ in L^p (provided when $p=\infty$ that f is uniformly continuous). We have

$$\|I_j[|\psi|,\phi]T_j\|_p \leq C(\psi,p,\varepsilon)\|S_{a_j^{-1}b}f\|_p \tag{55}$$

by the stability estimate in Lemma 6. When $p=\infty$, $\|S_{a_j^{-1}b}f\|_\infty \to 0$ as $j\to\infty$ by the uniform continuity of f because $\|a_j^{-1}b\|\to 0$. So suppose $p<\infty$. Then $S_{a_j^{-1}b}f \to 0$ pointwise a.e. by the almost everywhere continuity of f. Hence $S_{a_j^{-1}b}f \to 0$ in L^p by dominated convergence (with the dominating function constructed from (52), noting that $\mathcal{K}(a_j^{-1}b)=\{0\}$ for all large j by the paragraph after (52)). Therefore $I_j[|\psi|,\phi]T_j \to 0$ in L^p by (55), as desired.

Part (c). We already have a stability estimate on f_j from Proposition 9(c). And the norm $\|S_{a_j^{-1}b}f\|_p$ that occurs in (55) is bounded by $C(J,\Delta)\|Qf\|_p$ for all $j\geq J$, by (52) and the paragraph after (52). Hence the desired stability estimate on $f_j^\bullet$ follows from (54) and (55).

Part (e). Notice $f\in L^\infty$, by the hypothesis $Qf\in L^p$ and Lemma 23. So Proposition 9(e) yields pointwise convergence of f_j to f in part (a), and pointwise convergence of $\frac{1}{N}\sum_{n=1}^N f_{j(n)}$ to f in part (b).

Thus in view of (54), it is enough to show $\lim_{j\to\infty} I_j[|\psi|,\phi]T_j = 0$ pointwise a.e. Write

$$d(\ell) = \max_{j\geq \ell}\|a_j^{-1}b\|, \qquad \ell\in\mathbb{Z},$$

for the maximal stretching of the matrices $a_j^{-1}b$ for $j\geq\ell$. Notice $d(\ell)\to 0$ as $\ell\to\infty$ because the a_j are expanding. Putting $U_\ell(x,y) = (S_{d(\ell)I}f)(x+y)$, we deduce $T_j\leq U_\ell$ for all $j\geq\ell$. Hence for each fixed ℓ,

$$\begin{aligned}\limsup_{j\to\infty}(I_j[|\psi|,\phi]T_j)(x) &\leq \limsup_{j\to\infty}(I_j[|\psi|,\phi]U_\ell)(x)\\ &\leq U_\ell(x,0)\|P\psi\|_\infty = (S_{d(\ell)I}f)(x)\|P\psi\|_\infty\end{aligned}$$

pointwise a.e. by Lemma 7(d) (noting that $U_\ell\in L^{(\infty,\infty)}$ because $\|S_{d(\ell)I}f\|_\infty\leq 2\|f\|_\infty$, and that $P\psi\in L^\infty$ by Lemma 19 since $\psi\in L^\infty$ has a radially decreasing L^1 majorant).

But $d(\ell)\to 0$ and so $\lim_{\ell\to\infty}(S_{d(\ell)I}f)(x)=0$ a.e. by the almost everywhere continuity of f. We conclude $I_j[|\psi|,\phi]T_j\to 0$ pointwise a.e., completing the proof of part (e). □

6.3. **Applications to L^p sampling when ψ has bounded variation, or has Fourier transform in ℓ^1 on the lattice.** The least satisfactory feature of Propositions 9 and 10 is their reliance on the sequence $j(n)$, when the periodization of ψ is not constant. This sequence $j(n)$ depends on ψ, and in principle can be determined inductively by requiring $g = P\psi - 1$ to satisfy the Banach–Saks norm estimate (133). But in practice the determination would be difficult.

Fortunately we have discovered two classes of ψ for which Propositions 9(b) and 10(b) hold with the simplest possible dilation averaging, namely averaging over *all* dilation scales $j>J$ with no need for a special subsequence $j(n)$. In other words, one can take $j(n)=J+n$ independently of ψ.

The resulting sampling formula is $f=\lim_{N\to\infty}\frac{1}{N}\sum_{j=J+1}^{J+N} f_j$ in L^p, and similarly for $f_j^\bullet$.

Our first corollary considers ψ with bounded variation, for dilations growing exponentially.

Corollary 11. *Assume $\psi \in L^1 \cap BV$ with $\int_{\mathbb{R}^d} \psi\, dx = 1$. Take $1 \le p \le \infty$. If $d > 1$ and $1 \le p < \infty$ then assume $P(|\psi|) \in L^\infty$. If $d > 1$ and $p = \infty$ then assume $Q\psi \in L^1$. Let $J \in \mathbb{Z}$.*

Assume the dilations a_j grow exponentially (meaning $|a_{j+1}x| \ge \gamma |a_j x|$ for all $x \in \mathbb{R}^d$ and all $j \ge J$, for some growth factor $\gamma > 1$).

(I) [Average sampling] Let $\delta \in \{0, 1\}$. If $1 \le p < \infty$ then suppose $\phi \in L^q, P(|\phi|^\delta) \in L^\infty$ and let $f \in L^p$. If $p = \infty$ then suppose $\phi \in L^\infty$ with ϕ having compact support, and $f \in L^\infty$. For all p, suppose $\int_{\mathbb{R}^d} \phi\, dx = 1$.

Then parts (a)–(d) of Proposition 9 hold, with $j(n) = J + n$ in part (b).

(II) [Pointwise sampling] Suppose $f \in L^p$ with $Qf \in L^p$, f continuous a.e. and with $x = a_j^{-1} b z_j(k)$ being a partial Lebesgue point for f, for each $j \ge J, k \in \mathbb{Z}^d$. (For example, if $d = 1$ then it suffices to suppose $f \in L^p \cap BV(\mathbb{R})$ and f is either left or right continuous.)

Then parts (a)–(d) of Proposition 10 hold, with $j(n) = J + n$ in part (b).

Remarks on Corollary 11.

1. The pointwise sampling in part (II)(b) of the corollary yields (when $b = I$ and $z_j(k) = k$)

$$f(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=J+1}^{J+N} \left(\sum_{k\in\mathbb{Z}^d} f(a_j^{-1}k)\psi(a_j x - k) \right) \qquad \text{in } L^p, \quad 1 \le p < \infty. \tag{56}$$

2. Pointwise convergence can be included in parts (I) and (II). Under the hypotheses of Corollary 11(II) in dimension $d = 1$ (noting $\psi \in BV \subset L^\infty$ in one dimension) and with $1 \le p < \infty$, if ψ has a radially decreasing L^1 majorant and the a_j expand nicely, then Proposition 10(e) tells us that parts (a) and (b) of Proposition 10 hold also with pointwise convergence, using $j(n) = J + n$ in part (b). For example (56) holds pointwise a.e., under these conditions.

3. Dyadic dilations ($a_j = 2^j I$) certainly grow exponentially.

Proof of Corollary 11. In dimension $d = 1$ the bounded variation of ψ guarantees $Q\psi \in L^1$ and $P|\psi| \in L^\infty$, by Lemmas 24 and 23. Thus when $1 \le p < \infty$, we have $P|\psi| \in L^\infty$ in all dimensions, hence $\psi \in L^\infty$; this together with $\psi \in L^1$ ensures $\psi \in L^p$. When $p = \infty$ we have $Q\psi \in L^1$ in all dimensions, hence $\psi \in L^\infty$ by Lemma 23.

Let $\varepsilon = 1$.

Part (I). The hypotheses of Proposition 9 are satisfied, and so all we need justify is that the averaging in part (b) can be taken over the sequence $j(n) = J + n$. That is, we want to show (46) holds in the following simple form when $1 \le p < \infty$:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=J+1}^{J+N} (P\psi)(a_j x) = 1 \qquad \text{in } L^p_{loc} \text{ and pointwise a.e.} \tag{57}$$

We know $P\psi \in L^\infty$, and the bounded variation of ψ implies $P\psi \in BV_{loc}$ by Lemma 20. Further $P\psi$ has mean value $\int_{\mathbb{R}^d} \psi\, dx = 1$ by (6). Since the dilations grow exponentially, the last paragraph of Lemma 5 applies to $g(x) = (P\psi)(x) - 1 \in BV_{loc} \cap L^\infty$ and shows that (57) holds.

Part (II). The hypotheses of Proposition 10 are satisfied, and so the averaging we established in (57) completes the proof in part (II), except for the parenthetical statement. For that, suppose we are in dimension $d = 1$ and have $1 \leq p \leq \infty$ with $f \in L^p \cap BV(\mathbb{R})$ being either left or right continuous. Then $Qf \in L^p(\mathbb{R})$ by Lemma 24 and f is continuous except on a countable set, because f has bounded variation. And every point $x \in \mathbb{R}$ is a partial Lebesgue point for f (by the one-sided continuity of f). The hypotheses of part (II) are now satisfied, as we needed to show. $\square$

The second corollary considers ψ whose Fourier transform has values in ℓ^1 on the lattice, and integer dilation matrices that need not grow exponentially. Again the main conclusion is the validity of averaging over all dilation scales $j > J$, when ψ has nonconstant periodization.

For simplicity we restrict to $p < \infty$.

Corollary 12. *Assume $\psi \in L^1 \cap L^\infty$ with $\int_{\mathbb{R}^d} \psi \, dx = 1$ and*

$$\sum_{\ell \in \mathbb{Z}^d} |\widehat{\psi}(\ell)| < \infty. \tag{58}$$

Assume $P(|\psi|) \in L^\infty$ (this is unnecessary if $\psi \geq 0$).

Suppose the dilation matrices a_j (for $j \geq J \in \mathbb{Z}$) are invertible, expanding, have integer entries, and $a_{j_1} - a_{j_2}$ is invertible whenever $j_1 > j_2 \geq J$. Take $b = I$ and $1 \leq p < \infty$.

(I) [Average sampling] Let $\delta \in \{0, 1\}$ and suppose $\phi \in L^q, P(|\phi|^\delta) \in L^\infty$ with $\int_{\mathbb{R}^d} \phi \, dx = 1$, and let $f \in L^p$.

Then parts (a)–(d) of Proposition 9 hold, with $j(n) = J + n$ in part (b).

(II) [Pointwise sampling] Suppose $f \in L^p$ with $Qf \in L^p$, f continuous a.e. and with $x = a_j^{-1} b z_j(k)$ being a partial Lebesgue point for f, for each $j \geq J, k \in \mathbb{Z}^d$. (For example, if $d = 1$ then it suffices to suppose $f \in L^p \cap BV(\mathbb{R})$ and f is either left or right continuous.)

Then parts (a)–(d) of Proposition 10 hold, with $j(n) = J + n$ in part (b).

The translation matrix is $b = I$ throughout the corollary and its conclusions.

Remarks on Corollary 12.

1. If ψ is a Schwartz function with $\int_{\mathbb{R}^d} \psi \, dx = 1$ then all the hypotheses on ψ are satisfied.
2. If ψ has a radially decreasing L^1 majorant then the hypothesis $P|\psi| \in L^\infty$ is satisfied, by Lemma 19.
3. In all dimensions, the dyadic dilations $a_j = 2^j I$ satisfy the requirements in the corollary, for $j \geq 0$. In dimension $d = 1$, the same is true if the a_j are a strictly increasing sequence of positive integers.
4. The condition $\sum_{\ell \in \mathbb{Z}^d} |\widehat{\psi}(\ell)| < \infty$ in (58) is equivalent to absolute convergence of the Fourier series of $P\psi$ (see the proof below). Sufficient conditions for this absolute convergence are well known: for example in dimension $d = 1$ it is true if $P\psi$ is Hölder continuous with Hölder exponent bigger than 1/2; see [29, Theorem I.6.3].

Hence if $\psi \in C_c(\mathbb{R})$ is Hölder continuous with exponent bigger than 1/2, and $\int_{\mathbb{R}} \psi \, dx = 1$, then ψ satisfies all the hypotheses of the corollary for $d = 1$.

Proof of Corollary 12. The function $P\psi$ is locally bounded and $\mathbb{Z}^d$-periodic, with Fourier series $\sum_{\ell \in \mathbb{Z}^d} \widehat{\psi}(\ell) e^{2\pi i \ell x}$. This series converges absolutely to a continuous function, in view of assumption (58), and hence equals $P\psi$ pointwise a.e.

(In particular if $\psi \geq 0$ then $P|\psi| = P\psi \leq \sum_{\ell \in \mathbb{Z}^d} |\widehat{\psi}(\ell)| < \infty$ so that the assumption $P|\psi| \in L^\infty$ follows from (58).)

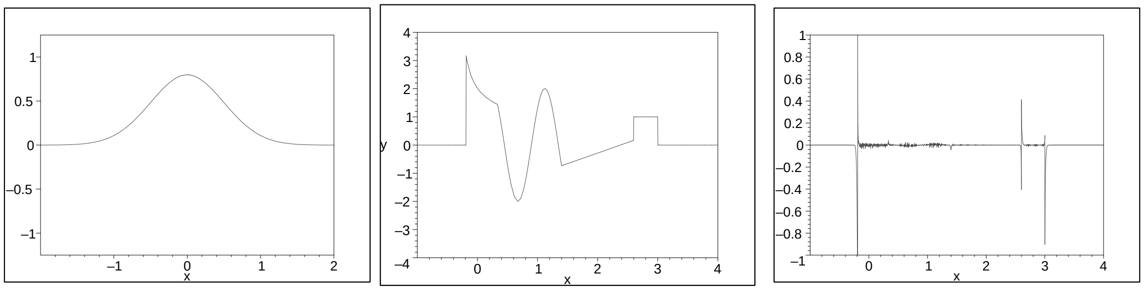

FIGURE 1. Left: $\psi(x) = \sqrt{2/\pi}e^{-2x^2}$ = Gaussian. Middle: $f(x)$. Right: Error $= f(x)$ minus its $J = 4, N = 4$ approximation from (1).

Let $\varepsilon = 1$. To prove Part (I) (or Part (II)) of the corollary, observe the hypotheses of case (i) of Proposition 9 (or 10) are satisfied. Thus once more all we need justify is the averaging in (57).

Since $\widehat{\psi}(0) = \int_{\mathbb{R}^d} \psi \, dx = 1$, we have

$$\frac{1}{N} \sum_{j=J+1}^{J+N} (P\psi)(a_j x) - 1 = \sum_{\ell \neq 0} \widehat{\psi}(\ell) \sigma_{N,\ell}(x) \tag{59}$$

where $\sigma_{N,\ell}$ is an arithmetic mean of exponentials:

$$\sigma_{N,\ell}(x) = \frac{1}{N} \sum_{j=J+1}^{J+N} e^{2\pi i \ell a_j x}, \qquad N \in \mathbb{N}, \quad \ell \neq 0.$$

The exponential functions $x \mapsto e^{2\pi i \ell a_j x}$, with $\ell \neq 0$ fixed, are mutually orthogonal in $L^2(\mathcal{C})$ for $j \geq J$ because the a_j have integer entries and $\ell(a_{j_1} - a_{j_2}) \neq 0$ when $j_1 > j_2 \geq J$. Hence $\sigma_{N,\ell}(x) \to 0$ as $N \to \infty$ for almost every $x \in \mathcal{C}$, by the strong law of large numbers for bounded, orthogonal random variables (*e.g.* Lemma 29).

Consequently $\|\sigma_{N,\ell}\|_{L^p(\mathcal{C})} \to 0$ as $N \to \infty$ by dominated convergence (with domination provided by $\|\sigma_{N,\ell}\|_\infty \leq 1$), whenever $1 \leq p < \infty$. Therefore

$$\Big\| \sum_{\ell \neq 0} \widehat{\psi}(\ell) \sigma_{N,\ell} \Big\|_{L^p(\mathcal{C})} \leq \sum_{\ell \neq 0} |\widehat{\psi}(\ell)| \, \|\sigma_{N,\ell}\|_{L^p(\mathcal{C})} \to 0 \qquad \text{as } N \to \infty.$$

That is, the $L^p(\mathcal{C})$ norm of (59) tends to zero. (And (59) also tends to zero pointwise a.e. in $\mathcal{C}$ by a similar argument.) This proves the desired L^p_{loc} and pointwise convergence in (57), because $(P\psi)(a_j x)$ is $\mathbb{Z}^d$-periodic (recalling a_j has integer entries). $\square$

Gaussian Example. The normalized Gaussian $\psi(x) = e^{-|x|^2/\tau}/(\pi\tau)^{d/2}$ satisfies the hypotheses of Corollary 12.

Figure 1 illustrates the dilation-averaged pointwise sampling method from Corollary 12 (II)(b) for the Gaussian ψ with $\tau = 1/2$. The lefthand graph shows ψ, the middle graph shows a function f, and the righthand graph shows the error between f and its $J = 4, N = 4$ approximation from (56), for $d = 1$ and $a_j = 2^j$.

Take $b = I$. Note the Gaussian has periodization

$$(P\psi)(x) = 1 + \sum_{\ell \in \mathbb{Z}^d \setminus \{0\}} e^{-|\ell|^2 \pi^2 \tau} \cos(2\pi \ell x)$$

by the Poisson summation formula. With $d = 1$ and $\tau = 1/2$ we deduce $\|P\psi - 1\|_\infty < 0.015$. Thus while the periodization of the Gaussian is not constant, it is within 1.5% of being constant and so one can obtain a reasonable approximation to f even without averaging over the dilations in the pointwise sampling formula (56). (Though of course the averaging over dilations is crucial if one wants actual convergence in the limit.)

As far as Figure 1 is concerned, the errors would get only slightly worse if instead of averaging the $j = 5, 6, 7, 8$ terms we used the $j = 8$ term on its own.

Approximate approximations of this kind were investigated in detail for the Gaussian by Maz'ya and Schmidt [33]. Their broader work is surveyed in [40], with L^p-results in particular in [32]. They consider Sobolev spaces as well. We emphasize that approximate approximations possess *inescapable* saturation errors due to the periodization of ψ not being constant, that is due to $\{\psi(x - k) : k \in \mathbb{Z}^d\}$ forming only an *approximate* partition of unity.

Maz'ya and Schmidt also require certain higher moments of their ψ to vanish, which we do not require in this paper.

7. Sampling in Sobolev space

7.1. **Strang–Fix condition implies constant periodization.** We start with a lemma to explain Theorem 3's hypotheses on the zeros of $\widehat{\psi}$. Recall $X(x) = x$ is the identity function, and $\chi_r = 1 + |X|^r$ when $r \geq 0$.

Lemma 13. *Take a nonnegative integer m and suppose $\psi \in W^{m,1}$ with $\chi_n \psi^{(\rho)} \in L^1$ for some multiindex ρ of order $|\rho| \leq m$ and some integer $n \geq 0$. Then $\widehat{\psi} \in C^n(\mathbb{R}^d \setminus \{0\})$, and if $\rho = 0$ then $\widehat{\psi} \in C^n(\mathbb{R}^d)$.*

Now let $0 \leq \tilde{n} \leq n$ and assume $(D^\tau \widehat{\psi})(\ell b^{-1}) = 0$ for all $|\tau| \leq \tilde{n}$ and all row vectors $\ell \in \mathbb{Z}^d \setminus \{0\}$. Then for every $|\tau| \leq \tilde{n}$ the periodization of $(-X)^\tau \psi^{(\rho)}$ is constant with

$$P((-X)^\tau \psi^{(\rho)})(x) = \begin{cases} \frac{\tau!}{(\tau-\rho)!} \int_{\mathbb{R}^d} (-y)^{\tau-\rho} \psi(y)\, dy & \text{if } \tau \geq \rho, \\ 0 & \text{otherwise,} \end{cases} \qquad a.e.$$

Proof of Lemma 13. Suppose $|\tau| \leq n$. Then $(-2\pi i X)^\tau \psi^{(\rho)}$ is integrable by the assumption $\chi_n \psi^{(\rho)} \in L^1$. So we can differentiate the transform $\widehat{\psi^{(\rho)}}(\xi) = \int_{\mathbb{R}^d} \psi^{(\rho)}(x) e^{-2\pi i \xi x}\, dx$ through the integral τ times, obtaining $\widehat{\psi^{(\rho)}} \in C^n(\mathbb{R}^d)$ since τ was arbitrary. But $\widehat{\psi^{(\rho)}}(\xi) = (2\pi i \xi)^\rho \widehat{\psi}(\xi)$, and so $\widehat{\psi}(\xi)$ has n continuous derivatives away from the origin, at least.

The periodization $x \mapsto P((2\pi i(-X))^\tau \psi^{(\rho)})(bx)$ is $\mathbb{Z}^d$-periodic and is locally integrable by Lemma 18. Its ℓ^{th} Fourier coefficient is

$$\begin{aligned}
&\int_{\mathcal{C}} P((2\pi i(-X))^\tau \psi^{(\rho)})(bx) e^{-2\pi i \ell x}\, dx \\
&= \int_{b\mathcal{C}} \sum_{k\in\mathbb{Z}^d} (2\pi i(-x+bk))^\tau \psi^{(\rho)}(x-bk) e^{-2\pi i \ell b^{-1} x}\, dx \qquad \text{by } x\mapsto b^{-1}x \text{ and definition of } P \\
&= \int_{\mathbb{R}^d} (-2\pi i x)^\tau \psi^{(\rho)}(x) e^{-2\pi i \ell b^{-1} x}\, dx \qquad \text{by } x \mapsto x+bk \qquad (60) \\
&= D_\xi^\tau \int_{\mathbb{R}^d} \psi^{(\rho)}(x) e^{-2\pi i \xi x}\, dx \Big|_{\xi=\ell b^{-1}} \\
&= D_\xi^\tau (2\pi i \xi)^\rho \widehat{\psi}(\xi) \Big|_{\xi = \ell b^{-1}}
\end{aligned}$$

by parts. This last expression is zero when $|\tau| \le \tilde{n}$ and $\ell \in \mathbb{Z}^d \setminus \{0\}$, by hypothesis on the zeros of $\widehat{\psi}$. Thus all the Fourier coefficients of $P((2\pi i(-X))^\tau \psi^{(\rho)})$ vanish, except possibly the zeroth one, and so $P((2\pi i(-X))^\tau \psi^{(\rho)})$ is a constant function.

This constant value is given by the $\ell = 0$ Fourier coefficient, which by (60) equals

$$\int_{\mathbb{R}^d} (-2\pi i x)^\tau \psi^{(\rho)}(x)\, dx = \begin{cases} (2\pi i)^{|\tau|} \frac{\tau!}{(\tau-\rho)!} \int_{\mathbb{R}^d} (-x)^{\tau-\rho} \psi(x)\, dx & \text{if } \tau \ge \rho, \\ 0 & \text{otherwise,} \end{cases}$$

after integrating by parts ρ times. □

7.2. **Average sampling.** Now we develop sampling formulas in Sobolev space. Recall $\frac{1}{p} + \frac{1}{q} = 1$, and $|f|^0 = \mathbb{1}_{\{f\neq 0\}}$, and that $p(\varepsilon) = \varepsilon + (1-\varepsilon)p$ and $q(\delta) = \delta + (1-\delta)q$.

Proposition 14. *Take $m \in \mathbb{N}$ and $\varepsilon, \delta \in \{0,1\}$. Assume one of the following conditions holds:*

(i)′ *$1 \le p < \infty, \psi \in W^{m,p}$ and $\chi_{|\mu|}\psi^{(\mu)} \in L^p, (p-1)P(|\chi_{|\mu|}\psi^{(\mu)}|^\varepsilon) \in L^\infty$ for all $|\mu| \le m$, and ϕ is measurable with $\chi_m \phi \in L^{q(\delta)}$ and $|\chi_m \phi|^\delta \in L^1$, and $f \in C_c^m$;*

(ii) *$1 \le p < \infty, \psi \in W^{m,p}$ and $\chi_{|\mu|}\psi^{(\mu)} \in L^p, (p-1)P(|\chi_{|\mu|}\psi^{(\mu)}|^\varepsilon) \in L^\infty, Q(|\chi_{|\mu|}\psi^{(\mu)}|^{p(\varepsilon)}) \in L^1$ for all $|\mu| \le m$, and $\phi \in L^q$ with ϕ having compact support, and $f \in W^{m,p}$;*

(iii) *$p = \infty, \psi \in W^{m,\infty}$ and $Q(\chi_{|\mu|}\psi^{(\mu)}) \in L^1$ for all $|\mu| \le m$, and $\phi \in L^\infty$ with compact support, and $f \in W^{m,\infty}$.*

Assume $\int_{\mathbb{R}^d} \psi\, dx = 1$ and $\int_{\mathbb{R}^d} \phi\, dx = 1$. Suppose

$$(D^\mu \widehat{\psi})(\ell b^{-1}) = 0 \qquad (61)$$

for all $|\mu| < m$ and all row vectors $\ell \in \mathbb{Z}^d \setminus \{0\}$. Assume the dilations a_j expand nicely. Then (a)–(e) hold:

(a) [Strang–Fix sampling] If (61) holds also whenever $|\mu| = m$, then

$$f = \lim_{j\to\infty} f_j \qquad \text{in } W^{m,p} \qquad (62)$$

provided when $p = \infty$ we also assume $f \in UC^m$ (meaning $D^\rho f$ is uniformly continuous for each $|\rho| \le m$). The function f_j was defined in (44).

(b) [Dilation-averaged sampling] Suppose (i)′ or (ii) holds, and if (ii) holds and $\varepsilon = 0$ then further suppose $f \in W^{m,\infty}$ with compact support. Let $J \in \mathbb{Z}$.

Then a strictly increasing integer sequence $\{j(n)\}_{n=1}^{\infty}$ exists (independent of f) such that $j(1) \geq J$ and

$$f = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_{j(n)} \qquad \text{in } W^{m,p}. \tag{63}$$

(c) [Stability] If either (ii) or (iii) holds, and $\varepsilon = 1$, then $\|f_j\|_{W^{m,p}} \leq C(\psi, \phi, p, m)\|f\|_{W^{m,p}}$ for all $j \in \mathbb{Z}$.

(d) [Spanning] If (i)′ or (ii) holds, or if (iii) holds and $f \in L^1$, then $f_j \in W^{m,p}$-span$\{\psi_{j,k} : k \in \mathbb{Z}^d\}$.

(e) [Pointwise convergence] Instead of (i)′, (ii) or (iii), assume $1 \leq p < \infty, \psi \in W^{m,p}$ and $\chi_{|\mu|}\psi^{(\mu)} \in L^1 \cap L^\infty$ for all $|\mu| \leq m$, and that $\phi \in L^q$ has compact support. Suppose either ψ has compact support and $f \in W^{m,p}$, or else $\chi_{|\mu|}\psi^{(\mu)}$ has a radially decreasing L^1-majorant for each $|\mu| \leq m$ and $f \in W^{m,\infty}$.

Then parts (a) and (b) above hold (just ignoring the first sentence of part (b)) with the $W^{m,p}$-convergence in limits (62) and (63) replaced by pointwise convergence a.e., and with pointwise convergence holding also in the analogous limits for the derivatives of order $\leq m$.

Remarks on Proposition 14.

1. If $\psi \in W^{m,p}$ has compact support, then case (i)′ in the proposition holds with $\varepsilon = 0$. If ψ is a Schwartz function then cases (i)′ and (ii) hold with $\varepsilon = 1$.

2. The requirements on the sequence $j(n)$ in part (b) reduce, when $m = 0$, back to the L^p requirement (46), as one can see putting $m = 0$ into (92) in the proof below. (The other requirement (91) in the proof becomes vacuous when $m = 0$.)

3. One can average over all dilations in part (b), meaning $j(n) = J+n$, under the following additional hypotheses: if the dilations a_j are isotropic ($a_j = \lambda_j I$) and grow exponentially ($\lambda_{j+1} \geq \gamma\lambda_j$ for all $j \in \mathbb{Z}$, for some constant $\gamma > 1$) and $(-X)^\tau\psi^{(\rho)} \in L^1 \cap BV$ and $P((-X)^\tau\psi^{(\rho)}) \in L^\infty$ for each $|\rho| = |\tau| = m$.

We prove this claim after proving Proposition 14.

Incidentally, in one dimension we need not assume $P((-X)^\tau\psi^{(\rho)}) \in L^\infty$ here because it follows from the periodicity of $P((-X)^\tau\psi^{(\rho)}) \in BV_{loc}(\mathbb{R})$ (noting that bounded variation implies boundedness, in one dimension).

4. Our proof does not work for dilation matrices such as $\begin{pmatrix} 3^j & 0 \\ 0 & 2^j \end{pmatrix}$ that are expanding without expanding nicely.

5. Case (i)′ is used to prove our Sobolev spanning result, Theorem 3. Notice case (ii) assumes more about ψ than case (i)′ does, but case (ii) has the advantage of applying to all $f \in W^{m,p}$, not just to $f \in C_c^m$, and also it yields a stability estimate in Proposition 14(c).

Sobolev sampling literature relevant to Proposition 14. As discussed in "Remarks on the Sobolev spanning literature", after Theorem 3, the best previous affine approximation results for $W^{m,p}$ are of Strang–Fix type under the assumptions that ψ satisfies the Strang–Fix condition to order m, that ψ has compact support, that the dilations are expanding and isotropic ($a_j = \lambda_j I$), and that $p = 2$ or $p = \infty$. Following are some further details.

The approximation formulas of Babuška [6, Theorem 4.1] and Strang–Fix [44, Theorem I] for $p = 2$ are not explicit, for they approximate f using sampled values of $\hat{f}$ rather than of f itself. When $p = \infty$, Strang and Fix [44, Theorem III] did use sampled pointwise values of $f \in W^{m+1,\infty}$ to approximate f in the $W^{m,\infty}$ norm (cf. Proposition 16 below).

Di Guglielmo [23, Théorème 2′] essentially proved Proposition 14(a)(c) for $p \geq 2$, but only for ψ having the special convolution form discussed in Section 3.2 (which is much stronger than the Strang–Fix condition). Di Guglielmo also assumed that $\psi \in W^{m,\infty}$ and $\phi \in L^\infty$ have compact support, that $f \in W^{m+1,p}$, and that the dilation matrices a_j are diagonal.

Our Proposition 14 improves on this literature because it treats all $1 \leq p \leq \infty$, it assumes the Strang–Fix condition only to order $m - 1$ (in part (b)), and it does not assume ψ has compact support. Further, the average sampling formulas we prove are for arbitrary $f \in W^{m,p}$, with convergence both in the Sobolev norm and pointwise, for all nicely expanding dilation matrices.

Proof of Proposition 14. Our first task is to show $f_j \in W^{m,p}$.

Fix a multiindex μ of order

$$r := |\mu| \leq m.$$

If we *formally* take the derivative through the sum over k in the definition of f_j, in (44), we find

$$(D^\mu f_j)(x) = |\det b| \sum_{k \in \mathbb{Z}^d} \left(\int_{\mathbb{R}^d} f(a_j^{-1} y) \phi(y - bk)\, dy \right) \sum_{\rho : |\rho| = r} c^\mu_\rho(a_j) \psi^{(\rho)}(a_j x - bk) \tag{64}$$

where the chain rule coefficients $c^\mu_\rho(a_j)$ depend only on the matrix a_j, in fact with

$$c^\mu_\rho(a_j) = D^\mu \left(\frac{(a_j x)^\rho}{\rho!} \right), \qquad |\rho| = |\mu|, \tag{65}$$

as one sees by applying the chain rule to the righthand side of (65). (In the special case of isotropic dilation matrices $a_j = \lambda_j I$, the coefficient $c^\mu_\rho(a_j)$ equals $\lambda_j^{|\mu|}$ if $\rho = \mu$, and equals 0 otherwise.)

Each entry of a_j is bounded by the matrix norm $\|a_j\|$ and so one deduces

$$|c^\mu_\rho(a_j)| \leq (d\|a_j\|)^r, \qquad |\rho| = |\mu| = r, \tag{66}$$

from (65) and induction on the multiindex μ. We will use this bound later.

To make the formal derivation of (64) rigorous, let $h(x, y) = f(x + y)$ and notice the righthand side of equation (64) equals $\sum_\rho c^\mu_\rho(a_j) I_j[\psi^{(\rho)}, \phi]h$, which belongs to L^p by Lemma 6 (noting in case (iii) that $P|\psi^{(\rho)}| \in L^\infty$ by Lemma 23 because $Q(\psi^{(\rho)}) \in L^1$ is assumed in case (iii)). Lemma 6 proves the sum over k in (64) converges pointwise absolutely a.e. to an L^p function. Then it is straightforward to show $D^\mu f_j$ exists weakly and is given by (64). Hence $f_j \in W^{m,p}$.

Part (d). In fact $f_j \in W^{m,p}$-span$\{\psi_{j,k} : k \in \mathbb{Z}^d\}$ if (i)′ or (ii) holds, or if (iii) holds and $f \in L^1$, because then the sum over k in (64) converges unconditionally in L^p by Lemma 6 (using case (iv) of Lemma 6 to handle case (iii) of this proposition).

Parts (a)(b). To prove the sampling formulas in parts (a) and (b), we will first show

$$D^\mu f = \lim_{j \to \infty} D^\mu f_j \qquad \text{in } L^p \qquad \text{if } |\mu| < m. \tag{67}$$

Then to complete the sampling formula in (a) we will show that if hypothesis (61) holds for all multiindices of order $\leq m$ (not just $< m$), then

$$D^\mu f = \lim_{j\to\infty} D^\mu f_j \qquad \text{in } L^p \quad \text{if } |\mu| = m. \tag{68}$$

To complete the sampling formula in (b) we will show (under the additional hypotheses of part (b), in particular assuming $1 \leq p < \infty$) that

$$D^\mu f = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} D^\mu f_{j(n)} \qquad \text{in } L^p \quad \text{if } |\mu| = m, \tag{69}$$

for a suitably chosen sequence $j(n)$ (independent of f).

In all of this work we will assume $f \in UC^m$ when $p = \infty$.

Our first step is to add and subtract an appropriate Taylor polynomial inside the formula (64) for $D^\mu f_j$. Specifically, we will show for almost every x that

$$(D^\mu f_j)(x) =$$

$$|\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} \left[f(a_j^{-1}y) - \sum_{|\sigma|\leq r} \frac{f^{(\sigma)}(x)}{\sigma!}(a_j^{-1}y - x)^\sigma \right] \phi(y-bk)\,dy \right) \sum_{|\rho|=r} c_\rho^\mu(a_j)\psi^{(\rho)}(a_j x - bk) \tag{70}$$

$$+ \sum_{|\sigma|=|\tau|\leq r} f^{(\sigma)}(x) \sum_{|\rho|=r} \frac{c_\rho^\mu(a_j)c_\sigma^\tau(a_j^{-1})}{\tau!} \sum_{\upsilon\leq\tau} \binom{\tau}{\upsilon} \int_{\mathbb{R}^d} y^{\tau-\upsilon}\phi(y)\,dy \cdot P((-X)^\upsilon \psi^{(\rho)})(a_j x) \tag{71}$$

where $\binom{\tau}{\upsilon}$ is a binomial coefficient with $\binom{\tau}{\tau} = 1$ (see below).

Note the integral $\int_{\mathbb{R}^d} y^{\tau-\upsilon}\phi(y)\,dy$ in (71) makes sense since the hypotheses of this proposition ensure $\chi_m\phi \in L^1$, by arguing like at the beginning of the proof of Lemma 7. And the periodization $P((-X)^\upsilon\psi^{(\rho)})$ makes sense because the hypotheses similarly ensure $\chi_{|\rho|}\psi^{(\rho)} \in L^1$ hence $(-X)^\upsilon\psi^{(\rho)} \in L^1$. In particular $\psi^{(\rho)} \in L^1$ and so $\psi \in W^{m,1}$.

To justify (70) and (71) we start with an identity that follows from (65):

$$\frac{(ax)^\sigma}{\sigma!} = \sum_{\tau:|\tau|=|\sigma|} c_\sigma^\tau(a)\frac{x^\tau}{\tau!}$$

whenever a is a $d\times d$ matrix. Applying this identity to $a_j^{-1}(y - a_j x)$ instead of to ax yields that

$$\begin{aligned} \frac{(a_j^{-1}y - x)^\sigma}{\sigma!} &= \sum_{\tau:|\tau|=|\sigma|} c_\sigma^\tau(a_j^{-1})\frac{(y-a_jx)^\tau}{\tau!} \\ &= \sum_{\tau:|\tau|=|\sigma|} \frac{c_\sigma^\tau(a_j^{-1})}{\tau!} \sum_{\upsilon:\upsilon\leq\tau} \binom{\tau}{\upsilon}(y-bk)^{\tau-\upsilon}(bk-a_jx)^\upsilon \end{aligned} \tag{72}$$

by binomial expansions. Now substitute this expansion (72) into (70), leading to cancellation with all the terms in (71), and thereby reducing us back to the known formula (64) for $D^\mu f_j$ as we wanted.

Remainder terms (70). Now that we have decomposed $D^\mu f_j$ into (70) and (71), we proceed to show the "remainder" terms (70) vanish in L^p in the limit $j \to \infty$. Indeed we take absolute values and aim to show

$$\lim_{j\to\infty} |\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} h_r(x, a_j^{-1}y - x)|a_j^{-1}y - x|^r |\phi(y - bk)|\, dy \right) |c_\rho^\mu(a_j)\psi^{(\rho)}(a_j x - bk)| = 0 \tag{73}$$

in L^p, where $|\rho| = r$ and

$$h_r(x,y) = \begin{cases} \left| f(x+y) - \sum_{|\sigma|\le r} \frac{1}{\sigma!} f^{(\sigma)}(x) y^\sigma \right| \Big/ |y|^r & \text{when } y \neq 0, \\ 0 & \text{when } y = 0. \end{cases} \tag{74}$$

Taylor's formula with integral remainder allows us to rewrite $h_r(x,y)$ as

$$\begin{aligned} h_r(x,y) &= \left| \int_{[0,1]} \sum_{|\sigma|=r} \frac{1}{\sigma!} [f^{(\sigma)}(x+ty) - f^{(\sigma)}(x)] y^\sigma \, d\omega_r(t) \right| \Big/ |y|^r \\ &\le \int_{[0,1]} \sum_{|\sigma|=r} |f^{(\sigma)}(x+ty) - f^{(\sigma)}(x)| \, d\omega_r(t) \\ &=: H_r(x,y) \end{aligned} \tag{75}$$

for almost every $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$, where ω_r is the probability measure on $[0,1]$ defined by

$$d\omega_r(t) = \begin{cases} r(1-t)^{r-1}\, dt & \text{if } r > 0, \\ d\delta_1(t) & \text{if } r = 0. \end{cases} \tag{76}$$

Since $h_r \le H_r$ it suffices to replace h_r in the desired estimate (73) with H_r.

We can further simplify (73) by noting that

$$\begin{aligned} |a_j^{-1}y - x|^r |c_\rho^\mu(a_j)| &\le \|a_j^{-1}\|^r |y - a_j x|^r (d\|a_j\|)^r && \text{by (66)} \\ &\le C(d,r)|y - a_j x|^r && \text{by (5) since } a_j \text{ expands nicely} \qquad (77) \\ &\le C(d,r)(|y - bk|^r + |bk - a_j x|^r) && \text{by the triangle inequality} \\ &\le C(d,r)(1 + |y - bk|^r)(1 + |bk - a_j x|^r). && \qquad (78) \end{aligned}$$

After putting (75) and (78) into (73) we see it's enough to prove

$$\lim_{j\to\infty} I_j[|\chi_r \psi^{(\rho)}|, \phi_r] H_r = 0 \qquad \text{in } L^p \tag{79}$$

where $\phi_r = |\chi_r \phi|$.

Our hypotheses on ϕ guarantee in case (i)$'$ that $\phi_r \in L^{q(\delta)}$ and $|\phi_r|^\delta \in L^1$, and in case (ii) that $\phi_r \in L^q$ with compact support, and in case (iii) that $\phi_r \in L^\infty$ with compact support.

Hence in case (i)$'$, the desired limit (79) follows from Lemma 8, because H_r has the form required of H^* in that lemma and $f^{(\sigma)} \in C_c$.

In cases (ii) and (iii) (the latter of which implies $\chi_r \psi^{(\rho)} \in L^\infty$ by Lemma 23), we see that (79) follows from Lemma 7(a) provided we show $H_r \in L^{(p,\infty)}$ and $H_r(\cdot, y) \to H_r(\cdot, 0) = 0$ in

L^p as $y \to 0$. But (75) implies

$$\|H_r(\cdot, y)\|_p \leq \int_{[0,1]} \sum_{|\sigma|=r} \|f^{(\sigma)}(\cdot + ty) - f^{(\sigma)}\|_p \, d\omega_r(t) \leq 2 \sum_{|\sigma|=r} \|f^{(\sigma)}\|_p < \infty \quad \text{for all } y \tag{80}$$

$$\to 0 \qquad \text{as } y \to 0 \tag{81}$$

where this final convergence is straightforward in case (ii) (because translation is continuous in the L^p-norm for $p < \infty$) and in case (iii) is justified by the uniform continuity of $f^{(\sigma)}$ (recall $f \in UC^m$ when $p = \infty$, in part (a)). Hence $H_r \in L^{(p,\infty)}$ and $H_r(\cdot, y) \to H_r(\cdot, 0) = 0$ in L^p as $j \to \infty$. This completes our proof that the remainder terms (70) vanish in L^p in the limit.

Main terms (71). Next we examine the main terms (71), that is, the non-remainder terms. Since $|\upsilon| \leq |\tau| = |\sigma| \leq r = |\rho|$, if either $\upsilon < \tau$ or $|\sigma| < r$ then $0 \leq |\upsilon| < |\rho|$ and so

$$P((-X)^\upsilon \psi^{(\rho)}) = 0 \quad \text{a.e.} \tag{82}$$

by Lemma 13 with $n = |\rho|$ and $\tilde{n} = |\rho| - 1$. In Lemma 13 we have used hypothesis (61) on the zeros of $\widehat{\psi}$.

Most terms in (71) vanish by (82), and the ones that are left have $|\tau| = |\sigma| = r$ and $\upsilon = \tau$, so that $\binom{\tau}{\upsilon} = \binom{\tau}{\tau} = 1$ and $\int_{\mathbb{R}^d} y^{\tau-\upsilon} \phi(y) \, dy = \int_{\mathbb{R}^d} \phi(y) \, dy = 1$. Thus (71) simplifies to

$$\sum_{|\sigma|=|\tau|=r} f^{(\sigma)}(x) \sum_{|\rho|=r} \frac{c^\mu_\rho(a_j) c^\tau_\sigma(a_j^{-1})}{\tau!} P((-X)^\tau \psi^{(\rho)})(a_j x). \tag{83}$$

We split (83) into the cases $\rho \neq \tau$ and $\rho = \tau$ to obtain

$$\sum_{|\sigma|=r} f^{(\sigma)}(x) \left\{ \sum_{|\rho|=|\tau|=r, \rho\neq\tau} c^\mu_\rho(a_j) c^\tau_\sigma(a_j^{-1}) \frac{P((-X)^\tau \psi^{(\rho)})(a_j x)}{\tau!} \right. \tag{84}$$

$$\left. + \sum_{|\rho|=r} c^\mu_\rho(a_j) c^\rho_\sigma(a_j^{-1}) \left(\frac{P((-X)^\rho \psi^{(\rho)})(a_j x)}{\rho!} - 1 \right) \right\} \tag{85}$$

$$+(D^\mu f)(x) \tag{86}$$

where we have used also that

$$\sum_{|\rho|=r} c^\mu_\rho(a_j) c^\rho_\sigma(a_j^{-1}) = c^\mu_\sigma(I) = \delta^\mu_\sigma \tag{87}$$

(this last identity is justified by evaluating $\delta^\mu_\sigma = D^\mu (a_j^{-1} a_j x)^\sigma / \sigma!$ with two applications of the chain rule).

Proof of limits (67) *and* (68). For proving the first limit (67) we suppose $r = |\mu| < m$. Then $|\tau| = |\rho| = r < m$ and so

$$P((-X)^\tau \psi^{(\rho)})(x) = \begin{cases} \rho! & \text{if } \tau = \rho \\ 0 & \text{otherwise} \end{cases} \tag{88}$$

for almost every x, by Lemma 13 with $n = \tilde{n} = r$ (and recalling $\int_{\mathbb{R}^d} \psi(y)\,dy = 1$). Hence (84) and (85) vanish and so (71) reduces to just $(D^\mu f)(x)$, meaning (67) follows immediately from our remainder estimate (the vanishing of (70) as $j \to \infty$).

To prove the next desired limit (68), let $r = |\mu| = m$ and suppose hypothesis (61) holds for all multiindices of order $\leq m$ (not just $< m$). Then (88) holds again whenever $|\tau| = |\rho| = r = m$, by Lemma 13 with $n = \tilde{n} = |\rho|$. Hence (68) follows from our remainder estimate.

Proof of limit (69). To prove the third desired limit (69), we suppose as in part (b) that (i)′ or (ii) holds (so that $1 \leq p < \infty$) and that if (ii) holds and $\varepsilon = 0$ then $f \in W^{m,\infty}$ with compact support. Consider $r = |\mu| = m$, so that $|\tau| = |\sigma| = |\rho| = r = m$.

Define the function

$$g_{\tau;\rho} = \begin{cases} P((-X)^\tau \psi^{(\rho)})/\tau! & \text{if } \tau \neq \rho, \\ P((-X)^\rho \psi^{(\rho)})/\rho! - 1 & \text{if } \tau = \rho, \end{cases}$$

so that $g_{\tau;\rho}$ is $b\mathbb{Z}^d$-periodic. The main terms (71) can be written as

$$\sum_{|\sigma|=m} f^{(\sigma)}(x) \sum_{|\rho|=|\tau|=m} c^\mu_\rho(a_j) c^\tau_\sigma(a_j^{-1}) g_{\tau;\rho}(a_j x) + (D^\mu f)(x) \tag{89}$$

when $r = |\mu| = m$, just by putting the definition of $g_{\tau;\rho}$ into (84)–(86).

Note the coefficients in (89) are bounded, with

$$|c^\mu_\rho(a_j) c^\tau_\sigma(a_j^{-1})| \leq C \qquad \text{for all } j \in \mathbb{Z} \tag{90}$$

by (66) and (5), since a_j expands nicely. By passing to a subsequence $j = j(n)$ we can assume the coefficient sequence converges:

$$c^\mu_\rho(a_{j(n)}) c^\tau_\sigma(a_{j(n)}^{-1}) \to c^{\mu,\tau}_{\rho,\sigma} \qquad \text{as } n \to \infty \tag{91}$$

for some real constants $c^{\mu,\tau}_{\rho,\sigma}$, for each $|\mu| = |\rho| = |\sigma| = |\tau| = m$. We can suppose $j(1) \geq J$.

Next we show $g_{\tau;\rho} \in L^p_{loc}$. If $p = 1$ then $g_{\tau;\rho} \in L^1_{loc}$ by Lemma 18, since we already know $(-X)^\tau \psi^{(\rho)} \in L^1$. If $1 < p < \infty$ then $P|\chi_m \psi^{(\rho)}|^\varepsilon \in L^\infty$ is assumed in cases (i)′ and (ii), remembering $|\rho| = m$. When $\varepsilon = 1$ this implies $P((-X)^\tau \psi^{(\rho)}) \in L^\infty \subset L^p_{loc}$. And when $\varepsilon = 0$ it means $\psi^{(\rho)}$ has the "finite intersection" property $P\mathbb{1}_{\{\psi^{(\rho)} \neq 0\}} \in L^\infty$, which together with $(-X)^\tau \psi^{(\rho)} \in L^p$ implies $P((-X)^\tau \psi^{(\rho)}) \in L^p_{loc}$.

Each function $g_{\tau;\rho}$ has mean value zero, because

$$|b\mathcal{C}|^{-1} \int_{b\mathcal{C}} P((-X)^\tau \psi^{(\rho)})\,dx = \int_{\mathbb{R}^d} (-X)^\tau \psi^{(\rho)}\,dx = \begin{cases} \rho! & \text{if } \tau = \rho \\ 0 & \text{otherwise} \end{cases}$$

by parts, using $|\tau| = |\rho| = m$.

Therefore Lemma 5 applies to each of the $g_{\tau;\rho}$, and repeated application of that lemma gives a subsequence of $j(n)$ (which we continue to call $j(n)$) such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g_{\tau;\rho}(a_{j(n)} x) = 0 \qquad \text{in } L^p_{loc} \text{ and pointwise a.e.} \tag{92}$$

for all $|\tau| = |\rho| = m$. Then

$$\frac{1}{N}\sum_{n=1}^{N} c_\rho^\mu(a_{j(n)})c_\sigma^\tau(a_{j(n)}^{-1})g_{\tau;\rho}(a_{j(n)}x) = \frac{1}{N}\sum_{n=1}^{N}[c_\rho^\mu(a_{j(n)})c_\sigma^\tau(a_{j(n)}^{-1}) - c_{\rho,\sigma}^{\mu,\tau}]g_{\tau;\rho}(a_{j(n)}x)$$
$$+ c_{\rho,\sigma}^{\mu,\tau}\frac{1}{N}\sum_{n=1}^{N} g_{\tau;\rho}(a_{j(n)}x)$$
$$\to 0 \qquad \text{in } L^p_{loc} \text{ as } N \to \infty, \tag{93}$$

for each $|\mu| = |\rho| = |\sigma| = |\tau| = m$, by combining (91) and (92) and using that for each ball E, the norm $\|g_{\tau;\rho}(a_j x)\|_{L^p(E)}$ is bounded for all large j, by (128) in Lemma 25.

The L^p_{loc} convergence in (93) implies that if $f^{(\sigma)}$ is bounded and has compact support then

$$\frac{1}{N}\sum_{n=1}^{N}[\text{formula (89) with } j = j(n)] \to D^\mu f \qquad \text{in } L^p. \tag{94}$$

Certainly this applies in case (i)′, where $f \in C_c^m$, and in case (ii) when $\varepsilon = 0$ because then part (b) assumes $f \in W^{m,\infty}$ has compact support. Thus in these cases, (94) tells us that the main terms (71) tend to $D^\mu f$ in L^p after averaging over $j = j(1), \ldots, j(N)$. Together with the remainder (70) vanishing as $j \to \infty$, this implies L^p convergence in (69).

It remains to handle case (ii) when $\varepsilon = 1$. In that case $Q(|\chi_m \psi^{(\rho)}|) \in L^1$ and so $P(|\chi_m \psi^{(\rho)}|) \in L^\infty$ by Lemma 23, hence $g_{\tau;\rho} \in L^\infty$. Then (91) and (92) imply that (93) holds pointwise a.e. Dominated convergence now proves (94) with convergence in L^p, since $f^{(\sigma)} \in L^p$ and the $g_{\tau;\rho}$ are bounded. Again this implies L^p convergence in (69).

Part (c). Assume (ii) or (iii) holds, and $\varepsilon = 1$. Let $0 \le r = |\mu| \le m$. We commence by proving stability of the remainder term:

$$\begin{aligned}
&\|\text{formula (70)}\|_p \\
&\le \sum_{|\rho|=r} C(d,r)\|I_j[|\chi_r \psi^{(\rho)}|, \phi_r]H_r\|_p && \text{by the reduction of (70) to (79)} \\
&\le C(\psi,\phi,p,d,r)\cdot 2\sum_{|\sigma|=r}\|f^{(\sigma)}\|_p && \text{by Lemma 6 and (80).}
\end{aligned}$$

(Lemma 6 does *not* state a j-independent norm estimate in case (i)′.) Then we prove stability of the main terms:

$$\|\text{formula (71)}\|_p \le C\sum_{|\sigma|=|\rho|=|\tau|=r}\|f^{(\sigma)}\|_p\frac{\|P((-X)^\tau\psi^{(\rho)})\|_\infty}{\tau!} \qquad \text{by (83) and (90).}$$

Note $\|P((-X)^\tau\psi^{(\rho)})\|_\infty$ is finite as follows. Cases (ii) and (iii) with $\varepsilon = 1$ ensure $Q(|\chi_r\psi^{(\rho)}|) \in L^1$ hence $P|\chi_r\psi^{(\rho)}| \in L^\infty$ by Lemma 23, therefore $P((-X)^\tau\psi^{(\rho)}) \in L^\infty$.

Combining the above two stability estimates and summing over $|\mu| = r$ gives the seminorm stability $|f_j|_{W^{r,p}} \le C(\psi,\phi,p,r)|f|_{W^{r,p}}$, then summing over $r = 0,\ldots,m$ gives the norm stability $\|f_j\|_{W^{m,p}} \le C(\psi,\phi,p,m)\|f\|_{W^{m,p}}$.

Part (e). To obtain *pointwise* convergence in (67)–(69) we modify the above proof. The main alterations are as follows.

We show $f_j \in W^{m,p}$ or $W^{m,\infty}$. Recall the righthand side of (64) equals $\sum_\rho c^\mu_\rho(a_j) I_j[\psi^{(\rho)}, \phi]h$ (where $|\rho| = |\mu| = r \le m$ and $h(x,y) = f(x+y)$). We first need to show this quantity is well defined.

Our hypotheses in part (e) ensure $\chi_r \psi^{(\rho)} \in L^\infty$ has a bounded radially decreasing L^1-majorant. Hence $P|\chi_r \psi^{(\rho)}| \in L^\infty$ by Lemma 19, and $Q(\chi_r \psi^{(\rho)})$ has a bounded radially decreasing L^1-majorant by Lemma 21. Also note $\phi_r = |\chi_r \phi| \in L^q$ since $\phi \in L^q$ has compact support.

The hypotheses in part (e) also imply $f \in W^{m,p}$ or $W^{m,\infty}$, so that $h \in L^{(p,\infty)}$ or $L^{(\infty,\infty)}$. If $h \in L^{(p,\infty)}$ then $I_j[\psi^{(\rho)}, \phi]h$ is covered by case (ii) in Lemma 6 with $\varepsilon = 1$, while if $h \in L^{(\infty,\infty)}$ then case (iii) in Lemma 6 is satisfied with $\varepsilon = 1$. Thus in any event, Lemma 6 guarantees that the series defining $I_j[\psi^{(\rho)}, \phi]h$ converges pointwise absolutely a.e. to a function in either L^p or L^∞.

Now it is straightforward to show $D^\mu f_j$ exists weakly and is given by (64), so that $f_j \in W^{m,p}$ or $W^{m,\infty}$.

The next alteration concerns the remainder terms (70). Instead of using Lemma 7(a) to get the L^p convergence to zero of the remainder estimate (79), we use Lemma 7(d) to get pointwise convergence to zero. Note that if $f \in W^{m,p}$ then $H_r \in L^{(p,\infty)}$ and if $f \in W^{m,\infty}$ then $H_r \in L^{(\infty,\infty)}$, by (80), and so formula (38) in Lemma 7(d) can be applied directly to prove the remainder estimate (79) pointwise a.e.

The main terms (71) are well defined because the hypotheses in part (e) ensure $P(|\chi_r \psi^{(\rho)}|) \in L^\infty$, as remarked above. This shows $g_{\tau;\rho} \in L^\infty \subset L^p_{loc}$, so that from (91) and (92) we deduce (93) holds pointwise a.e., hence (94) holds pointwise a.e. as desired. □

Proof of Remark 3 after Proposition 14. Because the dilations are isotropic, the chain rule coefficients in (65) simplify to $c^\mu_\rho(a_j) = \lambda_j^{|\mu|} \delta^\mu_\rho$, so that the requirement (91) on the sequence $j(n)$ holds as an identity with $c^{\mu,\tau}_{\rho,\sigma} = \delta^\mu_\rho \delta^\tau_\sigma$.

Further, $P((-X)^\tau \psi^{(\rho)}) \in BV_{loc}$ by Lemma 20, and so the functions $g_{\tau;\rho}$ belong to $BV_{loc} \cap L^\infty$ when $|\rho| = |\tau| = m$. Hence the last paragraph of Lemma 5 guarantees that $j(n) = J + n$ has the required averaging property (92). □

Next we extend Proposition 14 to handle $\psi = \psi_0 + \psi_1$ where ψ_0 and ψ_1 each satisfies the regularity and periodization hypotheses in Proposition 14 but the Strang–Fix type hypothesis is satisfied only by their sum $\psi = \psi_0 + \psi_1$.

To keep matters simple, we consider just case (i)′ with $\delta = 1$ and state only the sampling and spanning conclusions from Proposition 14(b)(d).

Corollary 15. *Take $m \in \mathbb{N}, 1 \le p < \infty$, and assume for $\varepsilon = 0, 1$ that*

- (i)′ *$\psi_\varepsilon \in W^{m,p}$ and $\chi_{|\mu|} \psi_\varepsilon^{(\mu)} \in L^p, (p-1)P(|\chi_{|\mu|} \psi_\varepsilon^{(\mu)}|^\varepsilon) \in L^\infty$ for all $|\mu| \le m$, and ϕ is measurable with $\chi_m \phi \in L^1$, and $f \in C^m_c$.*

Write $\psi = \psi_0 + \psi_1$ and assume $\int_{\mathbb{R}^d} \psi \, dx = \int_{\mathbb{R}^d} \phi \, dx = 1$. Suppose $(D^\mu \widehat{\psi})(\ell b^{-1}) = 0$ for all $|\mu| < m$ and all row vectors $\ell \in \mathbb{Z}^d \setminus \{0\}$. Assume the dilations a_j expand nicely, and let $J \in \mathbb{Z}$.

Then a strictly increasing integer sequence $\{j(n)\}_{n=1}^\infty$ exists (independent of f) such that $j(1) \ge J$ and $f = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N f_{j(n)}$ in $W^{m,p}$, where $f_j \in W^{m,p}$-span$\{\psi_{j,k} : k \in \mathbb{Z}^d\}$ was defined in (44).

The corollary is used in proving Theorem 3, the Sobolev spanning result.

To satisfy (i)′ in the corollary, one could for example take $\psi_0 \in W^{m,p}$ to have compact support and $\psi_1 \in W^{m,\infty}$ to decay quickly at infinity along with all its derivatives.

Proof of Corollary 15. Since ψ_0 and ψ_1 both satisfy case (i)′ of Proposition 14 (with $\varepsilon = 0$ and $\varepsilon = 1$ respectively, and $\delta = 1$), only two significant changes need be made to the proof of Proposition 14(b)(d).

When considering the remainder terms, substitute $|\psi^{(\rho)}| \leq |\psi_0^{(\rho)}| + |\psi_1^{(\rho)}|$ into (79) and proceed to estimate the two terms separately (with $\varepsilon = 0$ and $\varepsilon = 1$ respectively).

When considering the main terms, split ψ into $\psi_0 + \psi_1$ when showing $g_{\tau;\rho} \in L^p_{loc}$, so that the $\varepsilon = 0$ and $\varepsilon = 1$ portions can be justified separately. □

7.3. **Pointwise sampling.** Now we develop an analogue of Proposition 14 that uses *pointwise* sampling for C^m-smooth Sobolev functions. (The C^m-smoothness is convenient, but could be weakened like in Proposition 10.) The sampling will be uniform: $z_j(k) = k$.

Proposition 16. *Take $m \in \mathbb{N}$ and $\varepsilon \in \{0, 1\}$. Assume one of the following conditions holds:*

(i)′ $1 \leq p < \infty, \psi \in W^{m,p}$ *and* $\chi_{|\mu|}\psi^{(\mu)} \in L^p, (p-1)P(|\chi_{|\mu|}\psi^{(\mu)}|^\varepsilon) \in L^\infty$ *for all* $|\mu| \leq m$, *and* $f \in C_c^m$;

(ii) $1 \leq p < \infty, \psi \in W^{m,p}$ *and* $\chi_{|\mu|}\psi^{(\mu)} \in L^p, (p-1)P(|\chi_{|\mu|}\psi^{(\mu)}|^\varepsilon) \in L^\infty, Q(|\chi_{|\mu|}\psi^{(\mu)}|^{p(\varepsilon)}) \in L^1$ *for all* $|\mu| \leq m$, *and* $f \in W^{m,p} \cap C^m$ *with* $Q(f^{(\mu)}) \in L^p$ *for all* $|\mu| \leq m$;

(iii) $p = \infty, \psi \in W^{m,\infty}$ *and* $Q(\chi_{|\mu|}\psi^{(\mu)}) \in L^1$ *for all* $|\mu| \leq m$, *and* $f \in W^{m,\infty} \cap C^m$.

Assume $\int_{\mathbb{R}^d} \psi \, dx = 1$ *and*

$$(D^\mu \widehat{\psi})(\ell b^{-1}) = 0 \tag{95}$$

for all $|\mu| < m$ *and all row vectors* $\ell \in \mathbb{Z}^d \setminus \{0\}$. *Assume the dilations* a_j *expand nicely. Then (a)–(e) hold:*

(a) [Strang–Fix type sampling] If (95) *holds also whenever* $|\mu| = m$, *then*

$$f = \lim_{j \to \infty} f_j^\bullet \qquad \text{in } W^{m,p} \tag{96}$$

provided when $p = \infty$ *we also assume* $f \in UC^m$. *Here* $f_j^\bullet(x) = |\det b| \sum_{k \in \mathbb{Z}^d} f(a_j^{-1}bk)\psi_{j,k}(x)$ *is as defined in* (49), *for the uniform sampling points* $z_j(k) = k$.

(b) [Dilation-averaged sampling] Suppose (i)′ or (ii) holds, and if (ii) holds and $\varepsilon = 0$ *then further suppose* f *has compact support. Let* $J \in \mathbb{Z}$.

Then a strictly increasing integer sequence $\{j(n)\}_{n=1}^\infty$ *exists (independent of* f*) such that* $j(1) \geq J$ *and*

$$f = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^N f_{j(n)}^\bullet \qquad \text{in } W^{m,p}. \tag{97}$$

(c) [Stability] If either (ii) or (iii) holds, and $\varepsilon = 1$, *then for all* $j \geq J$ *we have* $\|f_j^\bullet\|_{W^{m,p}} \leq C(\psi, p, m, J) \sum_{|\mu| \leq m} \|Q(f^{(\mu)})\|_p$.

(d) [Spanning] If (i)′ or (ii) holds, or if (iii) holds and $Qf \in L^1$, *then* $f_j^\bullet \in W^{m,p}$*-span*$\{\psi_{j,k} : k \in \mathbb{Z}^d\}$.

(e) [Pointwise convergence] Instead of (i)′, (ii) or (iii), assume $1 \leq p < \infty, \psi \in W^{m,p}, f \in W^{m,p} \cap C^m$, *and that for all* $|\mu| \leq m$ *we have* $\chi_{|\mu|}\psi^{(\mu)} \in L^1 \cap L^\infty$ *and* $\chi_{|\mu|}\psi^{(\mu)}$ *has a radially decreasing* L^1*-majorant, and* $Q(f^{(\mu)}) \in L^p$.

Then parts (a) and (b) above hold (just ignoring the first sentence of part (b)) with the $W^{m,p}$-convergence in limits (96) *and* (97) *replaced by pointwise convergence a.e., and with pointwise convergence holding also in the analogous limits for the derivatives of order $\leq m$.*

Remarks on Proposition 16.

1. $f \in W^{m,\infty}$ in case (ii), because the assumption $Q(f^{(\mu)}) \in L^p$ implies $f^{(\mu)} \in L^\infty$ by Lemma 23.

2. That assumption $Q(f^{(\mu)}) \in L^p$ certainly holds true in dimension $d = 1$ if $f^{(\mu)}$ has bounded variation, just by Lemma 24.

Proof of Proposition 16. Our initial task is to show $f_j^\bullet \in W^{m,p}$. Fix a multiindex μ with $r := |\mu| \leq m$. Like in Proposition 14, formally differentiating the definition (49) of $f_j^\bullet$ yields that

$$(D^\mu f_j^\bullet)(x) = |\det b| \sum_{k \in \mathbb{Z}^d} f(a_j^{-1}bk) \sum_{\rho:|\rho|=r} c_\rho^\mu(a_j)\psi^{(\rho)}(a_j x - bk) \tag{98}$$

by the chain rule. The righthand side of this equation is exactly $\sum_\rho c_\rho^\mu(a_j) f_j^\bullet[\psi^{(\rho)}]$, where the temporary notation $f_j^\bullet[\psi^{(\rho)}]$ denotes the function obtained by replacing ψ with $\psi^{(\rho)}$ in the definition (49) of $f_j^\bullet$ (still with uniform sampling). Now to show rigorously that $f_j^\bullet$ is weakly differentiable with derivative given by (98), it is enough (like in the proof of Proposition 14) to observe that the series defining $f_j^\bullet[\psi^{(\rho)}]$ converges absolutely a.e. to an L^p function, which it does by Proposition 10(d). Note that if ψ satisfies case (i)′ or (ii) here then $\psi^{(\rho)}$ satisfies case (i) in Proposition 10.

Hence $f_j^\bullet \in W^{m,p}$.

<u>Part (d).</u> In fact $f_j^\bullet \in W^{m,p}\text{-span}\{\psi_{j,k} : k \in \mathbb{Z}^d\}$ if (i)′ or (ii) holds, or if (iii) holds and $Qf \in L^1$, because then the sum over k in (98) converges unconditionally in L^p by Proposition 10(d) applied to $\psi^{(\rho)}$.

<u>Parts (a)(b).</u> We will first show

$$D^\mu f = \lim_{j \to \infty} D^\mu f_j^\bullet \qquad \text{in } L^p \quad \text{if } |\mu| < m. \tag{99}$$

Then to complete the sampling formula in (a) we will show that if hypothesis (95) holds for all multiindices of order $\leq m$ (not just $< m$), then

$$D^\mu f = \lim_{j \to \infty} D^\mu f_j^\bullet \qquad \text{in } L^p \quad \text{if } |\mu| = m. \tag{100}$$

And to complete the sampling formula in (b) we will show (under the additional hypotheses of part (b), in particular assuming $1 \leq p < \infty$) that

$$D^\mu f = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} D^\mu f_{j(n)}^\bullet \qquad \text{in } L^p \quad \text{if } |\mu| = m, \tag{101}$$

for a suitably chosen sequence $j(n)$ (independent of f).

In all of this work we assume $f \in UC^m$ when $p = \infty$.

To begin with, we calculate that for almost every x,

$$(D^\mu f_j^\bullet)(x) = \text{formula (83)} +$$
$$|\det b| \sum_{k\in\mathbb{Z}^d} \left[f(a_j^{-1}bk) - \sum_{|\sigma|\le r} \frac{f^{(\sigma)}(x)}{\sigma!}(a_j^{-1}bk - x)^\sigma \right] \sum_{|\rho|=r} c_\rho^\mu(a_j)\psi^{(\rho)}(a_j x - bk) \tag{102}$$

by arguing like for (70) and (71) and simplifying (71) to (83). (The uniformity of the pointwise sampling is used in these calculations.)

The L^p convergence of (83) that is desired for proving (99)–(101) has been established already in Proposition 14, during the proof of (67)–(69). (Fortunately (83) involves only ψ, f and a_j, and the assumptions on these quantities in Proposition 16 are at least as strong as the corresponding assumptions in Proposition 14.)

Thus to prove (99)–(101) we have only to show the "remainder" terms (102) vanish in the limit as $j \to \infty$. After taking absolute values, we would like

$$\lim_{j\to\infty} |\det b| \sum_{k\in\mathbb{Z}^d} h_r(x, a_j^{-1}bk - x)|a_j^{-1}bk - x|^r |c_\rho^\mu(a_j)\psi^{(\rho)}(a_j x - bk)| = 0$$

in L^p for each $|\rho| = r$, where the function h_r was defined in (74).

Our first step is to apply estimate (77) with $y = bk$ and to recall $h_r \le H_r$ by the Taylor remainder estimate (75). Together, these show it suffices to prove

$$\lim_{j\to\infty} |\det b| \sum_{k\in\mathbb{Z}^d} H_r(x, a_j^{-1}bk - x)\, |(\chi_r\psi^{(\rho)})(a_j x - bk)| = 0 \tag{103}$$

for each $|\rho| = r$, where $H_r(x,y) = \sum_{|\sigma|=r} \int_{[0,1]} |f^{(\sigma)}(x+ty) - f^{(\sigma)}(x)|\, d\omega_r(t)$.

Let $\phi = \mathbb{1}_{b\mathcal{C}}/|b\mathcal{C}|$ and subtract the quantity $I_j[|\chi_r\psi^{(\rho)}|, \phi]H_r$ from (103). This quantity tends to zero in L^p as $j \to \infty$, as follows. In case (i)$'$ just use by Lemma 8, noting H_r has the form required of H^* in that lemma and $f^{(\sigma)} \in C_c$. In cases (ii) and (iii) (the latter of which implies $\chi_r\psi^{(\rho)} \in L^\infty$ by Lemma 23), instead apply Lemma 7(a), observing $H_r(\cdot, y) \to H_r(\cdot, 0) = 0$ in L^p as $y \to 0$ (when $p = \infty$ this convergence uses uniform continuity of $f^{(\sigma)}$).

After performing the subtraction of $I_j[|\chi_r\psi^{(\rho)}|, \phi]H_r$ from (103) and then taking absolute values, we see it would be enough to prove

$$\lim_{j\to\infty} |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d}\int_{[0,1]} \left| f^{(\sigma)}(x + t(a_j^{-1}bk - x)) - f^{(\sigma)}(x + t(a_j^{-1}y - x))\right| d\omega_r(t)\, \phi(y - bk)\, dy$$
$$\cdot\, |(\chi_r\psi^{(\rho)})(a_j x - bk)| = 0.$$

But $\phi(y - bk) \ne 0$ implies $y \in b(k + \mathcal{C})$ and so $|a_j^{-1}bk - a_j^{-1}y| \le \|a_j^{-1}b\|\sqrt{d}$, so that the last limit would follow from

$$\lim_{j\to\infty} |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d}\int_{[0,1]} (S_{a_j^{-1}b} f^{(\sigma)})(x + t(a_j^{-1}y - x))\, d\omega_r(t)\, \phi(y - bk)\, dy\, |(\chi_r\psi^{(\rho)})(a_j x - bk)| = 0,$$

where the operator $S_{a_j^{-1}b}$ was defined in the proof of Proposition 10 (note $\sqrt{d} \le \Delta$). Thus our goal is now to prove

$$\lim_{j\to\infty} I_j[|\chi_r\psi^{(\rho)}|, \phi]T_j = 0 \qquad \text{in } L^p \tag{104}$$

where $T_j(x,y) = \sum_{|\sigma|=r} \int_{[0,1]} (S_{a_j^{-1}b} f^{(\sigma)})(x+ty)\, d\omega_r(t)$.

We have

$$\|I_j[|\chi_r \psi^{(\rho)}|, \phi]T_j\|_p \leq C(\psi,p,r,\varepsilon)\|T_j\|_{(p,\infty)} \leq C(\psi,p,r,\varepsilon) \sum_{|\sigma|=r} \|S_{a_j^{-1}b} f^{(\sigma)}\|_p \tag{105}$$

by first the stability estimate in Lemma 6 (with $\delta = 1$) and then Minkowski's integral inequality. When $p=\infty$, $\|S_{a_j^{-1}b} f^{(\sigma)}\|_\infty \to 0$ as $j \to \infty$ by the assumed uniform continuity of $f^{(\sigma)}$ and because $\|a_j^{-1}b\| \to 0$. So suppose $p < \infty$. Then $S_{a_j^{-1}b} f^{(\sigma)} \to 0$ pointwise by the continuity of $f^{(\sigma)}$. Hence $S_{a_j^{-1}b} f^{(\sigma)} \to 0$ in L^p by dominated convergence (with the dominating function constructed from (52), noting that $\mathcal{K}(a_j^{-1}b) = \{0\}$ for all large j by the paragraph after (52)). Therefore $I_j[|\chi_r \psi^{(\rho)}|, \phi]T_j \to 0$ in L^p by (105), as desired.

Part (c). Assume $\varepsilon = 1$ and either (ii) or (iii) holds. For proving $W^{m,p}$-stability of $f_j^\bullet$, we return to the decomposition of $D^\mu f_j^\bullet$ in (102) and (83).

A stability estimate for the main term (83) was essentially established in the proof of Proposition 14(c), namely that $\|\text{formula (83)}\|_p \leq C(\psi,p,m)\|f\|_{W^{m,p}}$.

To get L^p-stability of the remainder terms (102), it suffices to show (in view of our proof above) that

$$\big\|I_j[|\chi_r \psi^{(\rho)}|, \phi]H_r\big\|_p \leq C(\psi,p,r) \sum_{|\sigma|=r} \|f^{(\sigma)}\|_p, \tag{106}$$

$$\big\|I_j[|\chi_r \psi^{(\rho)}|, \phi]T_j\big\|_p \leq C(\psi,p,r,J) \sum_{|\sigma|=r} \|Q(f^{(\sigma)})\|_p, \tag{107}$$

for each $|\rho| = r, j \geq J$. The first inequality follows from cases (ii) and (iii) of Lemma 6 together with the Minkowski integral inequality estimate that $\|H_r\|_{(p,\infty)} \leq \sum_{|\sigma|=r} 2\|f^{(\sigma)}\|_p$, and the second inequality follows from (105) and the fact that $\|S_{a_j^{-1}b} f^{(\sigma)}\|_p \leq C(J)\|Qf^{(\sigma)}\|_p$ for all $j \geq J$ (the final inequality being justified by (52) and the paragraph after it).

Part (e). Observe $f \in W^{m,\infty}$, since the hypothesis $Q(f^{(\mu)}) \in L^p$ implies $f^{(\mu)} \in L^\infty$ by Lemma 23. Thus the hypotheses of Proposition 14(e) are satisfied.

The desired pointwise convergence of the terms in (83) has been proved already in the proof of Proposition 14(e). Thus from (102) and the work following it, we need only show that (103) holds pointwise.

Note $H_r \in L^{(\infty,\infty)}$ and $H_r(x,0) = 0$ and $\chi_r \psi^{(\rho)} \in L^1 \cap L^\infty$. Therefore by applying estimate (38) in Lemma 7(d) we obtain that $I_j[|\chi_r \psi^{(\rho)}|, \phi]H_r \to 0$ pointwise a.e.

By subtracting this quantity $I_j[|\chi_r \psi^{(\rho)}|, \phi]H_r$ like in parts (a) and (b) above, we see it now suffices to prove pointwise convergence a.e. in (104). Write $d(\ell) = \max_{j \geq \ell} \|a_j^{-1}b\|$ for the maximal stretching of the matrices $a_j^{-1}b, j \geq \ell$. Defining $U_\ell(x,y) = \int_{[0,1]} (S_{d(\ell)I} f^{(\sigma)})(x + ty)\, d\omega_r(t)$, we deduce $T_j \leq U_\ell$ for all $j \geq \ell$, and hence for each fixed ℓ,

$$\begin{aligned}\limsup_{j\to\infty} (I_j[|\chi_r \psi^{(\rho)}|, \phi]T_j)(x) &\leq \limsup_{j\to\infty} (I_j[|\chi_r \psi^{(\rho)}|, \phi]U_\ell)(x) \\ &\leq (\text{const})U_\ell(x,0) = (\text{const})(S_{d(\ell)I} f^{(\sigma)})(x)\end{aligned}$$

pointwise a.e. by estimate (38) in Lemma 7(d) (noting that $U_\ell \in L^{(\infty,\infty)}$ because $\|S_{d(\ell)I} f^{(\sigma)}\|_\infty \leq 2\|f^{(\sigma)}\|_\infty$). But $(S_{d(\ell)I} f^{(\sigma)})(x) \to 0$ as $\ell \to \infty$ for all x, by the continuity of $f^{(\sigma)}$ and since $d(\ell) \to 0$. This proves pointwise convergence a.e. in (104), as needed for part (e). □

7.4. **Rate of approximation.** The Sobolev sampling in the preceding two propositions can be adapted to yield explicit rates of approximation. But we must use more sophisticated coefficients.

Take $m \in \mathbb{N}$ and consider functions ϕ and ψ as in the next proposition, with $\widehat{\phi}(0) \neq 0$ and $\widehat{\psi}(0) \neq 0$, and with $\widehat{\phi}, \widehat{\psi} \in C^m$ since $\chi_m \phi, \chi_m \psi \in L^1$ will be assumed.

We claim there exists a finite set of lattice points $K \subset \mathbb{Z}^d$ and coefficients $\alpha_k, \beta_k \in \mathbb{C}$ such that the linear combinations

$$\Phi(x) = \sum_{k \in K} \alpha_k \phi(x - bk) \qquad \text{and} \qquad \Psi(x) = \sum_{k \in K} \beta_k \psi(x + bk)$$

satisfy the moment conditions

$$\int_{\mathbb{R}^d} x^\mu \Phi(x)\, dx = \begin{cases} 1 & \text{if } \mu = 0 \\ 0 & \text{if } 0 < |\mu| \leq m \end{cases}, \quad \int_{\mathbb{R}^d} (-x)^\mu \Psi(x)\, dx = \begin{cases} 1 & \text{if } \mu = 0 \\ 0 & \text{if } 0 < |\mu| \leq m-1 \end{cases}. \tag{108}$$

To justify this claim for Ψ we adapt the reasoning on [44, p. 833] as follows. Let $K = \{k \in \mathbb{Z}^d : |k_1| + \cdots + |k_d| \leq m-1\}$ and write $B(\xi) = \sum_{k \in K} \beta_k e^{2\pi i \xi k}$ for the trigonometric polynomial with coefficients β_k to be determined. After checking that

$$\int_{\mathbb{R}^d} (-x)^\mu \Psi(x)\, dx = (2\pi i)^{-|\mu|} D^\mu \left(B(\xi b) \widehat{\psi}(\xi) \right) \Big|_{\xi=0}$$

we see the task in (108) is to choose $B(\xi)$ such that the derivatives of $B(\xi b)$ agree up to order $m-1$ at $\xi = 0$ with the derivatives of $\widehat{\psi}(\xi)^{-1}$. In other words the derivatives of $B(\xi)$ should agree with those of $\widehat{\psi}(\xi b^{-1})^{-1}$ up to order $m-1$, at $\xi = 0$. This is true if we take

$$B(\xi) = \sum_{|\mu| \leq m-1} D_\theta^\mu \left(\widehat{\psi}(\theta b^{-1})^{-1} \right) \Big|_{\theta=0} \; p_\mu(\xi)$$

where $\theta \in \mathbb{R}^d$ is regarded as a row vector and $p_0(\xi) \equiv 1$ and where for $0 < |\mu| \leq m-1$ we write $p_\mu(\xi)$ for the unique polynomial of degree $m-1$ jointly in $e^{2\pi i \xi_1}, \ldots, e^{2\pi i \xi_d}$ such that

$$(D^\sigma p_\mu)(0) = \begin{cases} 1 & \text{if } \sigma = \mu \\ 0 & \text{otherwise} \end{cases} \qquad \text{for all } |\sigma| \leq m-1.$$

This $B(\xi)$ has the desired form $\sum_{k \in K} \beta_k e^{2\pi i \xi k}$, and so the coefficients β_k are determined.

Argue similarly to construct Φ, except using m instead of $m-1$ throughout the argument so that we can handle moments of Φ up to $|\mu| \leq m$, in (108).

We can now prove sampling of $f \in W^{m,p}$ with a precise rate of approximation in the $W^{r,p}$-norm, for $0 \leq r \leq m$. Recall $\chi_m(x) = 1 + |x|^m$ and that $|f|_{W^{r,p}} = \left(\sum_{|\mu|=r} \|D^\mu f\|_p^p \right)^{1/p}$ is the Sobolev seminorm.

Proposition 17. *Take $m \in \mathbb{N}$. Assume one of the following conditions holds:*

(ii) $1 \leq p < \infty, \psi \in W^{m,p}$ *and* $Q(\chi_m \psi^{(\mu)}) \in L^1$ *for all* $|\mu| \leq m$, *and* $\phi \in L^q$ *with compact support;*

(iii) $p=\infty, \psi \in W^{m,\infty}$ and $Q(\chi_m \psi^{(\mu)}) \in L^1$ for all $|\mu| \leq m$, and $\phi \in L^1$ with $\chi_m \phi \in L^1$. Assume $\int_{\mathbb{R}^d} \psi \, dx \neq 0, \int_{\mathbb{R}^d} \phi \, dx \neq 0$, and that Φ and Ψ satisfy the moment conditions (108).

Suppose $(D^\mu \widehat{\psi})(\ell b^{-1}) = 0$ for all $|\mu| < m$ and all row vectors $\ell \in \mathbb{Z}^d \setminus \{0\}$.

(a) [Average sampling] If $f \in W^{m,p}$ then for each $r = 0, 1, \ldots, m$,

$$\begin{aligned} |f - F_j|_{W^{r,p}} &\leq C(\psi, \phi, p, m) |f|_{W^{m,p}} \|a_j^{-1}\|^m \|a_j\|^r \qquad \text{for all } j \in \mathbb{Z} \\ &= O(\|a_j^{-1}\|^{m-r}) \text{ if the dilations } a_j \text{ expand nicely,} \end{aligned} \tag{109}$$

where F_j is defined by average sampling:

$$\begin{aligned} F_j(x) &= |\det b| \sum_{k \in \mathbb{Z}^d} \left(\int_{\mathbb{R}^d} f(a_j^{-1} y) \Phi(y - bk) \, dy \right) \Psi_{j,k}(x) \\ &= |\det b| \sum_{k \in \mathbb{Z}^d} \left(\sum_{k_1, k_2 \in K} \alpha_{k_1} \beta_{k_2} \int_{\mathbb{R}^d} f(a_j^{-1} y) \phi(y - bk - bk_1 - bk_2) \, dy \right) \psi_{j,k}(x). \end{aligned}$$

(b) [Pointwise uniform sampling] Suppose $f \in W^{m,p} \cap C^m$. When $1 \leq p < \infty$ further suppose $Q(f^{(\mu)}) \in L^p$ for all $|\mu| \leq m$. Then for each $r = 0, 1, \ldots, m$,

$$\begin{aligned} |f - F_j^\bullet|_{W^{r,p}} &\leq C(\psi, p, m, J) \left(\sum_{|\mu| = m} \|Q(f^{(\mu)})\|_p^p \right)^{1/p} \|a_j^{-1}\|^m \|a_j\|^r \qquad \text{for all } j \geq J \\ &= O(\|a_j^{-1}\|^{m-r}) \text{ if the dilations } a_j \text{ expand nicely,} \end{aligned} \tag{110}$$

where $F_j^\bullet$ is defined by pointwise uniform sampling:

$$\begin{aligned} F_j^\bullet(x) &= |\det b| \sum_{k \in \mathbb{Z}^d} f(a_j^{-1} bk) \Psi_{j,k}(x) \\ &= |\det b| \sum_{k \in \mathbb{Z}^d} \left(\sum_{k_2 \in K} \beta_{k_2} f(a_j^{-1} b(k + k_2)) \right) \psi_{j,k}(x). \end{aligned}$$

Remarks on Proposition 17.

1. Note $\int_{\mathbb{R}^d} \Phi \, dx = \int_{\mathbb{R}^d} \Psi \, dx = 1$ by taking $\mu = 0$ in the moment condition (108). Hence Propositions 14 and 16 apply to Ψ and Φ and can give further information about F_j and $F_j^\bullet$ such as pointwise convergence to f.

2. Proposition 17 considers only cases (ii) and (iii) (with $\varepsilon = 1$) and omits case (i)$'$, because stability estimates underpin the proof.

3. The dilation averaging technique used in the other sampling results in this paper does *not* help here for obtaining rates of L^p approximation. The problem is that the averaged, rescaled periodization $\frac{1}{N} \sum_{n=1}^N (P\psi)(a_{j(n)} x)$ will generally *fail* to converge uniformly to its mean value, in particular if $(P\psi)(0) \neq$ (mean value) and $P\psi$ is continuous, and this failure destroys any hope of a convergence estimate in terms of $\|f\|_p$, in (40) and in similar formulas such as (111)–(112).

4. The approximation rate proved in Proposition 17(a) implies when the a_j expand nicely that $W^{m,p}$ lies in the $W^{m-1,p}$-span of $A_J(\psi)$. Since $W^{m,p}$ is dense in $W^{m-1,p}$, we conclude $A_J(\psi)$ spans $W^{m-1,p}$. In fact a better result holds: $A_J(\psi)$ spans $W^{m,p}$ by Proposition 14(b).

This illustrates the "gain of one order" achieved by dilation averaging, in this paper.

Sobolev sampling rate literature relevant to Proposition 17. The best previous approximation rate results are in the papers described in "Remarks on the L^p spanning literature", after Theorem 1. Those papers all restrict themselves to $r = 0$ (that is, approximating a $W^{m,p}$ function in the L^p norm) except for Babuška [6, Theorem 4.1] (for $p = 2$), Strang–Fix [44, Theorem I,III] (for $p = 2, \infty$) and di Guglielmo [23, Théorème 6] (for $p = 2$), who all approximate a $W^{m,p}$ function in the $W^{r,p}$ norm at rate $O(\|a_j^{-1}\|^{m-r})$ for $r = 0, 1, \ldots, m$, for isotropic dilations $a_j = \lambda_j I$. Di Guglielmo further obtained a $o(1)$ approximation rate when $r = m$ (like we do in Proposition 14(a)), although only for ψ having the special convolution form described in Section 3.2.

Proposition 17(a) improves on this literature because it treats all $1 \leq p \leq \infty$ and all $r = 0, 1, \ldots, m$, with ψ not required to have compact support and the dilation matrices required only to be nicely expanding. The pointwise sampling rate result in Proposition 17(b) also seems to be new. (For $p = \infty$ with isotropic dilations, see [44, Theorem III].)

Incidentally, Strang and Fix [44, Theorem I] proved a converse saying the condition $(D^\mu \widehat{\psi})(\ell b^{-1}) = 0$ for $|\mu| < m, \ell \in \mathbb{Z}^d \setminus \{0\}$, is *necessary* for approximating an arbitrary $f \in W^{m,2}$ in the $W^{r,2}$ norm at rate $O(\|a_j^{-1}\|^{m-r})$ in a "controlled" fashion by functions of the form $\sum_{k \in \mathbb{Z}^d} c_{j,k} \psi_{j,k}$.

Proof of Proposition 17. Fix a multiindex μ with $r := |\mu| \leq m$.

Part (a). We decompose $D^\mu F_j$ into

$$(D^\mu F_j)(x) =$$

$$|\det b| \sum_{k \in \mathbb{Z}^d} \left(\int_{\mathbb{R}^d} \left[f(a_j^{-1} y) - \sum_{|\sigma| \leq m} \frac{f^{(\sigma)}(x)}{\sigma!} (a_j^{-1} y - x)^\sigma \right] \Phi(y - bk)\, dy \right) \sum_{|\rho| = r} c_\rho^\mu(a_j) \Psi^{(\rho)}(a_j x - bk) \tag{111}$$

$$+ \sum_{|\sigma| = |\tau| \leq m} f^{(\sigma)}(x) \sum_{|\rho| = r} \frac{c_\rho^\mu(a_j) c_\sigma^\tau(a_j^{-1})}{\tau!} \cdot P((-X)^\tau \Psi^{(\rho)})(a_j x), \tag{112}$$

which is analogous to (70) and (71) except here we sum over $|\sigma| \leq m$ instead of $|\sigma| \leq r$, and we use the moment conditions (108) on Φ to evaluate all the moments occurring in (71).

The periodization $P((-X)^\tau \Psi^{(\rho)})$ belongs to L^∞ by Lemma 23, because $Q(\chi_m \Psi^{(\rho)}) \in L^1$.

We first show the "remainder" term (111) satisfies the estimate:

$$\|\text{formula (111)}\|_p \leq C(\psi, \phi, p, m) |f|_{W^{m,p}} \|a_j^{-1}\|^m \|a_j\|^r \qquad \text{for all } j \in \mathbb{Z}. \tag{113}$$

Now, the expression in (111) is bounded by the sum over ρ of the expression

$$|\det b| \sum_{k \in \mathbb{Z}^d} \left(\int_{\mathbb{R}^d} h_m(x, a_j^{-1} y - x) |a_j^{-1} y - x|^m |\Phi(y - bk)|\, dy \right) |c_\rho^\mu(a_j) \Psi^{(\rho)}(a_j x - bk)| \tag{114}$$

where h_m is defined by taking "$r = m$" in (74). And (114) is bounded by $C(d, m)\|a_j^{-1}\|^m \|a_j\|^r$ times $I_j[|\chi_m \Psi^{(\rho)}|, \Phi_m] h_m$ where $\Phi_m = |\chi_m \Phi|$, because

$$|a_j^{-1} y - x|^m |c_\rho^\mu(a_j)| \leq C(d, m)(1 + |y - bk|^m)(1 + |bk - a_j x|^m) \|a_j^{-1}\|^m \|a_j\|^r$$

by adapting the derivation of (78). Hence (113) would follow if we could show

$$\|I_j[|\chi_m \Psi^{(\rho)}|, \Phi_m] h_m\|_p \leq C(\psi, \phi, p, m) |f|_{W^{m,p}} \qquad \text{for all } j \in \mathbb{Z}. \tag{115}$$

Note that if $p < \infty$ then ϕ has compact support by case (ii) and so Φ has compact support as well, hence $\Phi_m \in L^q$. If $p = \infty$ then $\chi_m \phi \in L^1$ by case (iii) and so $\Phi_m \in L^1$.

Clearly (115) follows from cases (ii) and (iii) of Lemma 6 (with $\varepsilon = 1$), in view of the following observations. The hypothesis $Q(\chi_m \psi^{(\rho)}) \in L^1$ implies $\chi_m \Psi^{(\rho)} \in L^1 \cap L^\infty \subset L^p$ by Lemmas 22 and 23, and also $P(|\chi_m \Psi^{(\rho)}|) \in L^\infty$ by Lemma 23. Lastly note $h_m \in L^{(p,\infty)}$ with $\|h_m\|_{(p,\infty)} \leq 2|f|_{W^{m,p}}$ by (75) and (80) (with r changed to m). This justifies (115), hence the estimate (113) on the remainder.

Next we simplify the main terms (112). First note that $(D^\mu \widehat{\Psi})(\ell b^{-1}) = 0$ for all $|\mu| \leq m-1$ and all row vectors $\ell \in \mathbb{Z}^d \setminus \{0\}$, since the same is assumed for ψ. Hence Lemma 13 applied to Ψ, with $n = m$ and $\tilde{n} = m-1$, gives when $|\tau| \leq m-1$ that

$$\begin{aligned} P((-X)^\tau \Psi^{(\rho)}) &= \begin{cases} \frac{\tau!}{(\tau-\rho)!} \int_{\mathbb{R}^d} (-y)^{\tau-\rho} \Psi(y)\, dy & \text{if } \tau \geq \rho \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \tau! & \text{if } \tau = \rho \\ 0 & \text{otherwise} \end{cases} \qquad \text{by the moment condition (108) on } \Psi. \end{aligned} \tag{116}$$

Thus the only terms in (112) that can make a nonzero contribution are those with either $|\tau| = m$ or else $|\tau| \leq m-1$ and $\tau = \rho$. Hence (112) can be rewritten as

$$f^{(\mu)}(x) \quad (\text{ if } |\mu| \leq m-1\,) \tag{117}$$

$$+ \sum_{|\sigma|=|\tau|=m} f^{(\sigma)}(x) \sum_{|\rho|=r} \frac{c^\mu_\rho(a_j) c^\tau_\sigma(a_j^{-1})}{\tau!} \cdot P((-X)^\tau \Psi^{(\rho)})(a_j x), \tag{118}$$

where in (117) we have also used the identity (87) on the chain rule coefficients.

Therefore

$$D^\mu F_j - D^\mu f = (111) + (112) - D^\mu f = (111) + (118) \qquad \text{if } r = |\mu| \leq m-1. \tag{119}$$

If on the other hand $r = |\mu| = m$, then (117) is eliminated and (118) is the same as (83) evaluated with $r = m$, so that

$$\begin{aligned} D^\mu F_j - D^\mu f = (111) + (112) - D^\mu f &= (111) + (83)_{r=m} - D^\mu f \\ &= (111) + (84)_{r=m} + (85)_{r=m} \qquad \text{if } |\mu| = m. \end{aligned} \tag{120}$$

When estimating the righthand sides of (119) and (120), we can ignore the remainder term (111) because we have already proved a suitable L^p estimate on it, in (113).

Consider $r = |\mu| = |\rho| \leq |\sigma| = |\tau| = m$. From the bound $|c^\mu_\rho(a_j)| \leq (d\|a_j\|)^r$ in (66) and the fact that $P((-X)^\tau \Psi^{(\rho)}) \in L^\infty$, we deduce

$$\|\text{formula (118)}\|_p \leq C(\psi, m) |f|_{W^{m,p}} \|a_j\|^r \|a_j^{-1}\|^m.$$

This gives the desired estimate on $\|D^\mu F_j - D^\mu f\|_p$ on the righthand side of (119). And when $|\mu| = m$ we similarly find

$$\|(84)_{r=m} + (85)_{r=m}\|_p \leq C(\psi, m) |f|_{W^{m,p}} \|a_j\|^m \|a_j^{-1}\|^m,$$

giving the desired estimate on $\|D^\mu F_j - D^\mu f\|_p$ on the righthand side of (120).

Lastly, if the a_j expand nicely then $\|a_j\|^r \leq C\|a_j^{-1}\|^{-r}$ by (5), so that $\|a_j^{-1}\|^m \|a_j\|^r \leq C\|a_j^{-1}\|^{m-r}$.

Part (b). Similar to part (a) we decompose $D^\mu F_j^\bullet$ into

$$(D^\mu F_j^\bullet)(x) = \text{formula (112)} +$$
$$|\det b| \sum_{k\in\mathbb{Z}^d} \left[f(a_j^{-1}bk) - \sum_{|\sigma|\le m} \frac{f^{(\sigma)}(x)}{\sigma!}(a_j^{-1}bk - x)^\sigma \right] \sum_{|\rho|=r} c_\rho^\mu(a_j)\Psi^{(\rho)}(a_j x - bk) \tag{121}$$

(formally, just put $\phi = \delta_0$ and $\Phi = \phi$ into the proof of part (a), so that the moment conditions (108) on Φ are automatically satisfied).

The term (112) was discussed already in part (a), and so to prove (110) it suffices to show the "remainder" term (121) satisfies the estimate:

$$\|\text{formula (121)}\|_p \le C(\psi, p, m, J) \left(\sum_{|\sigma|=m} \|Qf^{(\sigma)}\|_p^p \right)^{1/p} \|a_j^{-1}\|^m \|a_j\|^r \qquad \text{for all } j \ge J,$$

which is analogous to the estimate (113) on the remainder (111) proved in (a).

By modifying the treatment of the remainder term in part (a) down to estimate (115), then further estimating h_m with H_m as in (75), we reduce the goal to showing

$$\left\| |\det b| \sum_{k\in\mathbb{Z}^d} H_m(x, a_j^{-1}bk - x)\, |(\chi_m \Psi^{(\rho)})(a_j x - bk)| \right\|_p \le C(\Psi, p, m, J) \left(\sum_{|\tau|=m} \|Qf^{(\tau)}\|_p^p \right)^{1/p} \tag{122}$$

for all $j \ge J$, where $H_m(x,y) = \sum_{|\sigma|=m} \int_{[0,1]} |f^{(\sigma)}(x+ty) - f^{(\sigma)}(x)|\, d\omega_r(t)$. Next we subtract and add the quantity $I_j[|\chi_m \Psi^{(\rho)}|, \phi] H_m$ inside the L^p norm on the left of (122), where $\phi = \mathbb{1}_{b\mathcal{C}}/|b\mathcal{C}|$. By reasoning like we did leading up to (104) (using the continuity of $f^{(\sigma)}$), we find (122) will follow if we can verify

$$\left\| I_j[|\chi_m \Psi^{(\rho)}|, \phi] H_m \right\|_p \le C(\Psi, p, m) \left(\sum_{|\sigma|=m} \|f^{(\sigma)}\|_p^p \right)^{1/p}, \tag{123}$$

$$\left\| I_j[|\chi_m \Psi^{(\rho)}|, \phi] T_j \right\|_p \le C(\Psi, p, m, J) \left(\sum_{|\sigma|=m} \|Qf^{(\sigma)}\|_p^p \right)^{1/p}, \tag{124}$$

where $T_j(x,y) = \sum_{|\sigma|=m} \int_{[0,1]} (S_{a_j^{-1}b} f^{(\sigma)})(x+ty)\, d\omega_r(t)$. But inequalities (123)–(124) are essentially the same as (106)–(107) except with $r = m$, and so they are proved already by the paragraph after (106)–(107). □

8. Acknowledgments

We thank Robert Jerrard for pointing us to Mazur's theorem and discussing the class BV. Special thanks to Rick Beatson, Allan Pinkus and Shayne Waldron for references to approximation theory literature, and to Hans Feichtinger, Say Song Goh and Wenchang Sun for references to sampling theory literature . Gilbert Strang contributed some helpful historical perspective, and P. Wojtaszczyk some references. Laugesen thanks the Department of Mathematics and Statistics at the University of Canterbury for hosting him during the latter part of this work.

Appendix A. **Basic properties of periodizations**

Recall the definition of periodization from Section 3.1:

$$(Pf)(x) = |\det b| \sum_{k \in \mathbb{Z}^d} f(x - bk).$$

Clearly if $f : \mathbb{R}^d \to [0, \infty]$ is measurable then $Pf : \mathbb{R}^d \to [0, \infty]$ is well defined and measurable. We also have:

Lemma 18. *If $f \in L^1$ then the series for $(Pf)(x)$ converges absolutely for almost every x, and Pf is $b\mathbb{Z}^d$-periodic and locally integrable.*

Proof. Integrating over $b\mathcal{C}$, where $\mathcal{C} = [0, 1)^d$ is the unit cube, we find

$$\int_{b\mathcal{C}} \sum_{k \in \mathbb{Z}^d} |f(x - bk)| \, dx = \int_{\mathbb{R}^d} |f(x)| \, dx < \infty$$

because $f \in L^1$. Hence the series for $(Pf)(x)$ converges absolutely for almost every x, and Pf is locally integrable. The $b\mathbb{Z}^d$-periodicity of Pf is clear. □

We say f has a *radially decreasing L^1-majorant* if $|f| \leq \eta$ a.e., for some radial function $\eta(|x|) \in L^1$ such that $\eta(|x|)$ decreases as a function of $|x|$.

Lemma 19. *If $f \in L^\infty$ has a radially decreasing L^1-majorant then $Pf \in L^\infty$.*

Proof. Here $f \in L^1$ and so $Pf \in L^1_{loc}$ is well defined a.e.

Write $\mathcal{K}$ for the finite collection of lattice points $k \in \mathbb{Z}^d$ such that $|bk| < 3 \operatorname{diam}(b\mathcal{C})$. Then for all $k \in \mathbb{Z}^d \setminus \mathcal{K}, x \in b\mathcal{C}$ and $y \in b(k + \mathcal{C})$, we have

$$|x - bk| \geq |bk| - \operatorname{diam}(b\mathcal{C}) \geq (|bk| + \operatorname{diam}(b\mathcal{C}))/2 \geq |y|/2,$$

and so $|f(x - bk)| \leq \eta(|x - bk|) \leq \eta(|y|/2)$. Hence for almost every $x \in b\mathcal{C}$,

$$\begin{aligned} |(Pf)(x)| &\leq |\det b|(\#\mathcal{K})\|f\|_\infty + \sum_{k \notin \mathcal{K}} \frac{|\det b|}{|b(k + \mathcal{C})|} \int_{b(k+\mathcal{C})} \eta(|y|/2) \, dy \\ &\leq |\det b|(\#\mathcal{K})\|f\|_\infty + \frac{1}{|\mathcal{C}|} \int_{\mathbb{R}^d} \eta(|y|/2) \, dy < \infty. \end{aligned}$$

□

Recall that $BV = BV(\mathbb{R}^d)$ denotes the class of functions with bounded variation in $\mathbb{R}^d$. The next lemma says the periodization of a BV function is locally in BV, which we use in remarks after Proposition 9.

Lemma 20. $P : L^1 \cap BV \to BV_{loc}$.

Proof. Take $f \in L^1 \cap BV$, so that $Pf \in L^1_{loc}$ by Lemma 18. Let $\mathcal{O} \subset \mathbb{R}^d$ be bounded and open. We must show $Pf \in BV(\mathcal{O})$. We can suppose f is real valued, by considering the real and imaginary parts separately.

For all vector fields $V \in C_c^1(\mathcal{O};\mathbb{R}^d)$ with $|V| \leq 1$ we have

$$\begin{aligned}
\int_{\mathcal{O}} (Pf)(x)(\nabla \cdot V)(x)\,dx &= |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathcal{O}} f(x-bk)(\nabla \cdot V)(x)\,dx \\
&\leq |\det b| \sum_{k\in\mathbb{Z}^d} \|\nabla f\|(\mathcal{O}-bk) \qquad \text{by [18, p. 170] since } f \in BV, \\
&\leq |\det b|\,(\#\mathcal{K})\,\|\nabla f\|(\mathbb{R}^d) \\
&< \infty
\end{aligned}$$

where $\mathcal{K}$ is any finite collection of lattice points such that $\mathcal{O} \subset \cup_{k\in\mathcal{K}} b(k+\mathcal{C})$. Our estimates are independent of V, and so Pf satisfies the requirements of belonging to $BV(\mathcal{O})$, as defined in [18, Chapter 5]. □

In one dimension, Lemma 20 can also be proved using the "classical" definition of $BV(\mathbb{R})$ from [21, §3.5].

Appendix B. The operators Q and S

Throughout this appendix, we take f to be a measurable function on $\mathbb{R}^d$ that is finite a.e. Define a local supremum operator by

$$(Qf)(x) = \text{ess. sup}_{|y-x|<\sqrt{d}} |f(y)| = \|f\|_{L^\infty(B(x,\sqrt{d}))},$$

where the choice of radius $\sqrt{d}$ will turn out to be convenient but not essential. Then $0 \leq (Qf)(x) \leq \infty$, and the function Qf is measurable because it is lower semicontinuous.

Incidentally the norm equivalence

$$\|Qf\|_p \approx \|f\|_{W(L^p)}, \qquad 1 \leq p \leq \infty,$$

is not difficult to show, where $W(L^p)$ is the Wiener amalgam space considered by Feichtinger and others (see e.g. [1], [2]).

Next define a "modulus of continuity" operator

$$(Sf)(x) = \text{ess. sup}_{|y-x|<\sqrt{d}} |f(x) - f(y)|,$$

which has a well defined value in $[0, +\infty]$ wherever $f(x)$ is finite. To prove Sf is measurable, note that

$$(Sf)(x) = \lim_{p\to\infty} \|f(x) - f(\cdot)\|_{L^p(B(x,\sqrt{d}))}$$

where

$$\|f(x) - f(\cdot)\|_{L^p(B(x,\sqrt{d}))} = \left(\int_{\mathbb{R}^d} \mathbb{1}_{B(0,\sqrt{d})}(y-x)|f(x)-f(y)|^p\,dy \right)^{1/p}$$

is a measurable function of x by the Fubini–Tonelli theorem.

The first lemma says that if f is bounded and has a radially decreasing L^1 majorant (as defined in Appendix A) then $Qf \in L^1$.

Lemma 21. *If $f \in L^\infty$ has a radially decreasing L^1-majorant then Qf has a bounded and radially decreasing L^1-majorant, and in particular $Qf \in L^1$.*

Proof. Write $\eta(|x|)$ for the radially decreasing L^1-majorant of f. We might as well take η to be bounded with $\|f\|_\infty \le \eta(0) < \infty$. Then $(Qf)(x)$ is majorized by

$$(Q\eta)(x) \le \begin{cases} \eta(0) & \text{if } |x| < \sqrt{d}, \\ \eta(|x| - \sqrt{d}) & \text{if } |x| \ge \sqrt{d}, \end{cases}$$

which is bounded, radially decreasing and belongs to L^1. □

Relations between P, Q and S. First we derive pointwise relations for Q and S.

Lemma 22. *The following inequalities hold pointwise a.e.:*

$$\begin{aligned} |f| &\le Qf, \\ 0 \le Qf &\le |f| + Sf, \\ 0 \le Sf &\le |f| + Qf. \end{aligned}$$

And if E is a bounded set in $\mathbb{R}^d$ then

$$|f(x)| \le (\widetilde{Q}f)(y) := \sum_{k:|k|<\operatorname{diam}(E)+\sqrt{d}} (Qf)(y+k) \tag{125}$$

for almost every $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ with $x - y \in E$.

Proof of Lemma 22. Consider the set $F = \{x \in \mathbb{R}^d : (Qf)(x) < \infty\}$, and the larger open set $G = \cup_{x\in F} B(x, \sqrt{d})$ on which f is locally bounded hence locally integrable. The Lebesgue differentiation theorem implies that at almost every $x \in G$,

$$|f(x)| \le \limsup_{\varepsilon\to 0} \frac{1}{|B(x,\varepsilon)|} \int_{B(x,\varepsilon)} |f(y)|\, dy \le (Qf)(x)$$

as we wanted. And if $x \notin G$ then $x \notin F$, so that $(Qf)(x) = \infty \ge |f(x)|$.

The inequality $Qf \le |f| + Sf$ follows just from the triangle inequality $|f(y)| \le |f(x)| + |f(x) - f(y)|$. Similarly for $Sf \le |f| + Qf$.

Now suppose E is a bounded set in $\mathbb{R}^d$, and $y \in \mathbb{R}^d$. Let $x \in G$ be a Lebesgue point for f such that $x - y \in E$. Choose $k \in \mathbb{Z}^d$ with $x - y \in k + \mathcal{C}$ so that $x \in B(y + k, \sqrt{d})$ and $|k| < \operatorname{diam}(E) + \sqrt{d}$. Then the Lebesgue differentiation theorem implies $|f(x)| \le (Qf)(y+k)$. On the other hand, if $x \notin G$ and $x-y \in E$ then choosing k as before shows that $(Qf)(y+k) = \infty$ (otherwise $y + k \in F$, which implies $x \in G$). Either way, we have proved (125). □

Now we prove norm relations for P and Q.

Lemma 23.

$$\|f\|_\infty \le \|Qf\|_p \qquad \textit{for all } 1 \le p \le \infty,$$
$$\|Pf\|_\infty \le \|P|f|\|_\infty \le C\|Qf\|_1.$$

Proof of Lemma 23. Obviously $\|f\|_\infty \le \|Qf\|_\infty$ by Lemma 22. So suppose $1 \le p < \infty$. For each $y \in \mathbb{R}^d$,

$$\begin{aligned} \|f\|_\infty^p &\le \sup_{k\in\mathbb{Z}^d} \|f\|^p_{L^\infty(B(y+k,\sqrt{d}))} \\ &\le \sum_{k\in\mathbb{Z}^d} \|f\|^p_{L^\infty(B(y+k,\sqrt{d}))} \\ &= \sum_{k\in\mathbb{Z}^d} (Qf)(y+k)^p, \end{aligned}$$

and so integrating over $y \in \mathcal{C}$ gives $\|f\|_\infty^p \le \|Qf\|_p^p$, which is the first inequality in the lemma.

Next, $|f(x-bk)| \le (\widetilde{Q}f)(x-bk-y)$ for all $k \in \mathbb{Z}^d$ and almost every $(x,y) \in \mathbb{R}^d \times b\mathcal{C}$, by applying Lemma 22 with $E = b\mathcal{C}$. For such x and y values, the definition of P implies

$$|\det b|^{-1} |(P|f|)(x)| \le \sum_{k\in\mathbb{Z}^d} (\widetilde{Q}f)(x-bk-y).$$

Integrating over $y \in b\mathcal{C}$ yields that for almost every x,

$$\begin{aligned} |(P|f|)(x)| &\le \int_{b\mathcal{C}} \sum_{k\in\mathbb{Z}^d} (\widetilde{Q}f)(x-bk-y)\, dy \\ &= \int_{\mathbb{R}^d} (\widetilde{Q}f)(x-y)\, dy = \|\widetilde{Q}f\|_1 \le C(b\mathcal{C}) \|Qf\|_1 \end{aligned}$$

by definition of $\widetilde{Q}$ in (125). $\square$

Counterexample. If $Qf \in L^1$ then $P|f| \in L^\infty$ by Lemma 23. The converse implication is true for compactly supported f (since $P|f| \in L^\infty$ implies $f \in L^\infty$, which implies $Qf \in L^\infty$, and Qf has compact support because f does). But the converse is false for non-compactly supported f as demonstrated by the example

$$f(x) = \sum_{\ell=0}^{\infty} \mathbb{1}_{\ell+(2^{-\ell-1},2^{-\ell}]}(x) \tag{126}$$

in one dimension with $b = 1$. In this example $P|f| \equiv 1 \in L^\infty$ but $Qf = \mathbb{1}_{(-1/2,\infty)} \notin L^1$.

Bounded variation. We next examine the effect of Q on functions of bounded variation, as needed for remarks after Propositions 10, 14 and 16.

Lemma 24. *In dimension $d = 1$, we have $Q : BV \cap L^p(\mathbb{R}) \to L^s(\mathbb{R})$ for all $1 \le p \le s \le \infty$, and hence $Q : W^{1,1}(\mathbb{R}) \to L^s(\mathbb{R})$ for all $1 \le s \le \infty$.*

Proof. Suppose $f \in BV \cap L^p(\mathbb{R})$. For proving $Qf \in L^s(\mathbb{R})$ we might as well suppose f is real valued, because $Qf \le Q(\operatorname{Re} f) + Q(\operatorname{Im} f)$.

For every two points x, y of approximate continuity of f with $|x-y| < 1$ we have

$$|f(x) - f(y)| \le \operatorname{ess} V_{x-1}^{x+1} f = \|f'\|(x-1, x+1)$$

by [18, §5.10], where $\operatorname{ess} V$ denotes the essential variation and $\|f'\|$ is a positive Radon measure with $\|f'\|(\mathbb{R}) < \infty$ because $f \in BV$. Since the set of points of approximate continuity has full measure [18, p. 47], we conclude

$$(Qf)(x) = \text{ess. sup}_{|y-x|<1}|f(y)| \le |f(x)| + \|f'\|(x-1,x+1) = |f(x)| + \int_{\mathbb{R}} \mathbb{1}_E(x,z)\, d\|f'\|(z)$$

for almost every x, where $E = \{(x,z) \in \mathbb{R}^2 : x-1 < z < x+1\}$ is a diagonal strip. Minkowski's integral inequality now yields that

$$\begin{aligned}\|Qf\|_s &\le \|f\|_s + \int_{\mathbb{R}} \|\mathbb{1}_E(\cdot,z)\|_s\, d\|f'\|(z) \\ &= \|f\|_s + 2^{1/s}\|f'\|(\mathbb{R}) < \infty\end{aligned}$$

so that $Qf \in L^s$. (Here we use $f \in L^s$, which is valid since $f \in L^p, p \le s$, and $f \in BV(\mathbb{R}) \subset L^\infty(\mathbb{R})$).

The last statement of the Lemma is clear since $W^{1,1} \subset BV \cap L^1$, in one dimension. □

Appendix C. **A Riemann–Lebesgue result**

A periodic measure applied to a large set should yield just the measure of each individual period cell times the number of period cells in the large set. This is the content of our first lemma.

Lemma 25. *Suppose μ is a $b\mathbb{Z}^d$-periodic complex Borel measure. If E is a ball then*

$$\frac{\mu(a_jE)}{|a_jE|} \to \frac{\mu(b\mathcal{C})}{|b\mathcal{C}|} \qquad \text{as } j \to \infty. \tag{127}$$

Further, if $J \in \mathbb{Z}$ is fixed then a constant $C(E,J) > 0$ exists such that

$$\frac{|\mu(a_jE)|}{|a_jE|} \le C(E,J)|\mu|(b\mathcal{C}) \qquad \text{for all } j \ge J. \tag{128}$$

Proof of Lemma 25. Write $N_1(j)$ for the number of lattice points $k \in \mathbb{Z}^d$ such that the set $b(k+\mathcal{C})$ is contained entirely within a_jE, and $N_2(j)$ for the (larger) number of lattice points for which $b(k+\mathcal{C})$ intersects a_jE. Then

$$N_1(j)|b\mathcal{C}| \le |a_jE| = |\det a_j||E| \le N_2(j)|b\mathcal{C}|.$$

In the reverse direction,

$$\begin{aligned}N_2(j)|b\mathcal{C}| &\le |\{x \in \mathbb{R}^d : \text{dist}(x, a_jE) < \text{diam}(b\mathcal{C})\}| \\ &\le |\det a_j||\{x \in \mathbb{R}^d : \text{dist}(x,E) < \|a_j^{-1}\|\,\text{diam}(b\mathcal{C})\}|\end{aligned} \tag{129}$$

and similarly

$$\begin{aligned}N_1(j)|b\mathcal{C}| &\ge |\{x \in a_jE : \text{dist}(x, \partial(a_jE)) > \text{diam}(b\mathcal{C})\}| \\ &\ge |\det a_j||\{x \in E : \text{dist}(x,\partial E) > \|a_j^{-1}\|\,\text{diam}(b\mathcal{C})\}|.\end{aligned}$$

Hence $N_1(j)|b\mathcal{C}| \sim |\det a_j||E| \sim N_2(j)|b\mathcal{C}|$ as $j \to \infty$, because $\|a_j^{-1}\| \to 0$ (the a_j are expanding).

The periodicity of μ ensures

$$|\mu(a_jE) - N_1(j)\mu(b\mathcal{C})| \le (N_2(j) - N_1(j))|\mu|(b\mathcal{C}),$$

and so dividing through by $|a_jE| = |\det a_j||E|$ and letting $j \to \infty$ completes the proof of (127). Finally, if $j \geq J \in \mathbb{Z}$ then

$$|\mu(a_jE)| \leq N_2(j)|\mu|(b\mathcal{C}) \leq C|a_jE|\,|\mu|(b\mathcal{C})$$

by (129) for some positive constant $C = C(E, J)$, since $\|a_j^{-1}\|$ is bounded for $j \geq J$. □

Next we prove a Riemann–Lebesgue lemma, as needed in the proofs of Lemma 5 and Lemma 7.

Lemma 26. *Let $1 \leq p \leq \infty$. Suppose $g \in L^p_{loc}$ and $h \in L^q$, and when $1 \leq p < \infty$ suppose h has compact support. If g is $b\mathbb{Z}^d$-periodic with mean value zero, then*

$$\int_{\mathbb{R}^d} g(a_jx)h(x)\,dx \to 0 \qquad \text{as } j \to \infty. \tag{130}$$

(The compact support assumption on h ensures the integrals make sense; the assumption is unnecessary when $p = \infty$ because then $g \in L^\infty_{loc}$ is globally bounded by the periodicity.)

Proof of Lemma 26. We need only consider $b = I$ being the identity, because then the general case follows by considering $g(bx), h(bx)$ and $b^{-1}a_jb$ instead of g, h and a_j. (Note the matrices $b^{-1}a_jb$ are expanding if and only if the a_j are expanding.)

First we reduce to h being bounded with compact support. This is immediate when $p = 1, q = \infty$. Suppose $1 < p < \infty$. Let $\delta > 0$ and choose E to be a ball containing the support of h. Choose $\tilde{h} \in L^\infty$ with support in E and with $\|h - \tilde{h}\|_q < \delta$ (possible since $q < \infty$). Then

$$\left| \int_{\mathbb{R}^d} g(a_jx)h(x)\,dx - \int_{\mathbb{R}^d} g(a_jx)\tilde{h}(x)\,dx \right| \tag{131}$$

$$\leq \left(|\det a_j|^{-1} \int_{a_jE} |g(x)|^p\,dx \right)^{1/p} \|h - \tilde{h}\|_q \qquad \text{by Hölder and } x \mapsto a_j^{-1}x,$$

$$\to \left(|E||\mathcal{C}|^{-1} \int_{\mathcal{C}} |g(x)|^p\,dx \right)^{1/p} \|h - \tilde{h}\|_q \qquad \text{as } j \to \infty, \text{ by Lemma 25 with } b = I, \tag{132}$$

$$\leq \left(|E||\mathcal{C}|^{-1}\right)^{1/p} \|g\|_{L^p(\mathcal{C})} \cdot \delta.$$

Since δ is arbitrary, we see it is enough to prove the lemma for g and $\tilde{h}$. Now suppose $p = \infty$, so that $g \in L^\infty$. Choose $\tilde{h} \in L^\infty$ with compact support and $\|h - \tilde{h}\|_1 < \delta$, then simply estimate (131) by $\|h - \tilde{h}\|_1\|g\|_\infty \leq \delta\|g\|_\infty$.

Thus for all p we can suppose h is bounded with support contained in some ball E. It further suffices to prove the lemma for a dense subclass of periodic $g \in L^p_{loc}$ with mean value zero, because if $\tilde{g} \in L^p_{loc}$ is $\mathbb{Z}^d$-periodic then

$$\limsup_{j\to\infty} \left| \int_{\mathbb{R}^d} g(a_jx)h(x)\,dx - \int_{\mathbb{R}^d} \tilde{g}(a_jx)h(x)\,dx \right| \leq \left(|E||\mathcal{C}|^{-1}\right)^{1/p} \|g - \tilde{g}\|_{L^p(\mathcal{C})}\|h\|_q$$

by arguing as for (132). In particular, then, it is enough to consider the subclass of periodic $g \in L^\infty_{loc}$ with mean value zero.

Hence it suffices to prove the $p = q = 2$ case of the lemma, for $g \in L^2_{loc}$ periodic with mean value zero and $h \in L^2$ with compact support. And then (as we have seen) it suffices to consider a dense subclass of periodic $\tilde{g} \in L^2_{loc}$ with mean value zero. We choose $\tilde{g}(x) = \sum_{|\ell| \le L} \hat{g}(\ell) e^{2\pi i \ell x}$ to be a partial sum of the Fourier series of g, noting that this partial sum converges to g in L^2_{loc} as $L \to \infty$.

But for $\tilde{g}$ and h we have

$$\int_{\mathbb{R}^d} \tilde{g}(a_j x) h(x)\, dx = \sum_{|\ell| \le L} \hat{g}(\ell) \hat{h}(-\ell a_j) \to 0 \qquad \text{as } j \to \infty$$

by the usual Riemann–Lebesgue lemma [42, Theorem I.1.2], using that $\hat{g}(0) = 0$ (since g has mean value zero) and that $|\ell a_j| \to \infty$ for each $\ell \neq 0$ by the expanding property of the a_j. This proves the lemma. □

Appendix D. Weak convergence implies norm and pointwise convergence of arithmetic means

Here we state a result of Banach–Saks and Szlenk, and develop a pointwise analogue. Then we deduce a "local" version of the theorem as needed for Lemma 5.

Throughout this appendix (X, μ) is a measure space and L^p means $L^p(X)$ (whereas in the rest of the paper L^p means $L^p(\mathbb{R}^d)$).

Theorem 27. *Let $1 \le p < \infty$ and $g_1, g_2, g_3, \ldots \in L^p$. If $p = 1$ then assume $\mu(X) < \infty$. Assume $g_j \rightharpoonup 0$ weakly in L^p as $j \to \infty$.*

Then an increasing integer sequence $0 < j_1(1) < j_1(2) < j_1(3) < \cdots$ exists such that for every subsequence j_2 of j_1, we have $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g_{j_2(n)} = 0$ in L^p and pointwise μ-a.e.

Proof. First consider $1 < p < \infty$. Banach and Saks [7, Théorème I] proved the desired L^p convergence in the theorem. (Riesz and Nagy [36, p. 80] then deduced a short proof for $p = 2$.) Banach and Saks were working on the unit interval with Lebesgue measure, but their proof holds verbatim in a general measure space. And while they stated the theorem without subsequences in the conclusion (meaning they took $j_2 = j_1$), it is a simple matter to strengthen the inductive algorithm in their proof to ensure

$$\|g_{j_2(1)} + \cdots + g_{j_2(n)}\|_p \le C(n + n^{p-1} + 1)^{1/p} \le C n^{\max(1/p, 1/q)}$$

for all $n \in \mathbb{N}$ and every subsequence j_2 of j_1, where $C = C(p) > 0$; compare with [7, p. 55] in the proof of Banach and Saks. This last estimate implies

$$\|(g_{j_2(1)} + \cdots + g_{j_2(n)})/n\|_p \le C/n^{\min(1/p, 1/q)}, \tag{133}$$

which gives the desired L^p convergence to zero, as $n \to \infty$.

For pointwise convergence, note $\|g_j\|_p \le C$ for all j (since weak convergence implies norm boundedness) and recall estimate (133), then just apply Lemma 30 below to obtain the pointwise convergence $(g_{j_2(1)} + \cdots + g_{j_2(n)})/n \to 0$ μ-a.e.

Now consider $p = 1$. Szlenk [45] proved the L^1 convergence in the theorem, again working on the unit interval with Lebesgue measure but with a proof that actually holds in a general *finite* measure space. For pointwise convergence, take the sequence j_1 provided by Szlenk (for which the arithmetic means converge to zero in L^1) and then use a theorem of Komlós [30, Theorem 1a] to pass to a subsequence (which we also call j_1) for which the arithmetic mean over each subsequence j_2 is pointwise convergent μ-a.e. Clearly this pointwise limit must equal zero μ-a.e. □

We will need a *local* version of Theorem 27.

Corollary 28. *Let $1 \le p < \infty$ and suppose $B_1, B_2, B_3, \ldots$ is a sequence of measurable sets in X. If $p = 1$ then assume $\mu(B_t) < \infty$ for all $t \in \mathbb{N}$.*

Suppose $g_1, g_2, g_3, \ldots$ are measurable functions with $g_j \in L^p(B_t)$ for all $j, t \in \mathbb{N}$ and with $g_j \rightharpoonup 0$ weakly in $L^p(B_t)$ as $j \to \infty$, for each t.

Then an increasing integer sequence $0 < j_1(1) < j_1(2) < j_1(3) < \cdots$ exists such that for every subsequence j_2 of j_1 and for every $t \in \mathbb{N}$, we have $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} g_{j_2(n)} = 0$ in $L^p(B_t)$ and pointwise μ-a.e. in B_t.

Proof. First construct a sequence that works on B_1, by using Theorem 27, and then pass to a subsequence that works on B_2, again by Theorem 27, and so on. Then take the diagonal sequence. We leave details to the reader. □

Remarks on Theorem 27 and Corollary 28.

1. Mazur's theorem [39, Theorem 3.13] is better known than the ones of Banach–Saks and Szlenk. Mazur's theorem applies to all normed linear spaces, not just to L^p, but has two disadvantages in our view.

First, its proof is non-explicit (relying on the Hahn–Banach theorem), whereas the result of Banach and Saks is proved by an explicit recursive algorithm. Second, Mazur's theorem yields convergence in norm for some unknown *convex combinations* of the weakly convergent sequence, rather than for the simple *arithmetic means* of that sequence as considered by Banach–Saks and Szlenk.

2. The Banach–Saks result (that weak convergence of a sequence implies norm convergence of the arithmetic means of some subsequence) has been given a new proof by Wojtaszczyk [49, p. 101], who also proved pointwise convergence [49, p. 102]. For references to earlier literature on pointwise convergence, see [49, p. 106]. Instead of using that literature, we obtain pointwise convergence in the proof of Theorem 27 by working directly with Lemma 30 below, because that lemma connects up nicely with the inductive construction by Banach and Saks.

3. Incidentally, the Banach–Saks theorem under the hypothesis of just norm boundedness (rather than weak convergence) has been extended by Kakutani [28] from L^p $(1 < p < \infty)$ to all uniformly convex spaces.

In proving the last paragraph of Lemma 5, we use the following version of the L^2-Strong Law of Large Numbers. The usual orthogonality hypothesis on the random variables is replaced here by a decay bound (134) on their inner products.

Lemma 29. *Let $h_1, h_2, h_3, \ldots \in L^2$ and suppose*

$$\mathrm{Re} \int_X h_m \overline{h_n}\, d\mu \le \beta(m-n) \qquad \text{for all } m, n \in \mathbb{N}, \tag{134}$$

for some function $\beta : \mathbb{Z} \to [0,\infty)$ satisfying $\sum_{\ell\in\mathbb{Z}} \beta(\ell) < \infty$. (In particular (134) holds if the h_n are orthonormal.) Then

$$\lim_{n\to\infty} \frac{h_1 + \cdots + h_n}{n} = 0 \quad \text{both in } L^2 \text{ and } \mu\text{-a.e.}$$

Proof of Lemma 29.

$$\begin{aligned}\|h_1+\cdots+h_N\|_2^2 &= \operatorname{Re}\sum_{m=1}^{N}\sum_{n=1}^{N}\int_X h_m\overline{h_n}\,d\mu \\ &\leq N\sum_{\ell\in\mathbb{Z}}\beta(\ell) \qquad \text{by (134), where } \ell=m-n, \\ &= CN. \end{aligned} \tag{135}$$

Hence $\|(h_1+\cdots+h_N)/N\|_2^2 \leq C/N \to 0$ as $N\to\infty$, giving the desired L^2-convergence.

The almost everywhere convergence now follows from Lemma 30 with $p=2$ and $r=1/2$, using that $\|h_n\|_2 \leq \beta(0)$ for all n by taking $m=n$ in (134). □

We use the next lemma to prove pointwise convergence in Theorem 27 (when $1<p<\infty$) and in Lemma 29.

Lemma 30. *Let $1<p\leq\infty$ and suppose $h_1,h_2,h_3,\ldots\in L^p$ with*

$$\|h_n\|_p \leq C, \tag{136}$$

$$\|\frac{h_1+\cdots+h_n}{n}\|_p \leq \frac{C}{n^r}, \tag{137}$$

for all n, for some constants $C,r>0$. Then

$$\frac{h_1+\cdots+h_n}{n} \to 0 \quad \textit{in } L^p \textit{ and } \mu\textit{-a.e.}$$

Proof of Lemma 30. Write $\sigma_n = h_1+\cdots+h_n$. The L^p convergence $\sigma_n/n\to 0$ is immediate from assumption (137). Hence $\sigma_n/n\to 0$ pointwise μ-a.e. when $n\to\infty$ through some subsequence of n-values. The goal is to prove this pointwise convergence as $n\to\infty$ through *all* n values.

When $p=\infty$, pointwise convergence follows from norm convergence. So suppose $1<p<\infty$.

Fix a positive integer $s>1/pr$. Define the set

$$X_m(\varepsilon) = \{x\in X : |\sigma_n(x)/n| \geq 2\varepsilon \text{ for some } n \text{ with } m^s\leq n<(m+1)^s\}$$

whenever $\varepsilon>0$ and $m\in\mathbb{N}$. We have

$$\begin{aligned}\mu(X_m(\varepsilon)) &\leq \mu(\{|\sigma_n|\geq 2m^s\varepsilon \text{ for some } n \text{ with } m^s\leq n<(m+1)^s\}) \\ &\leq \mu(\{|\sigma_{m^s}|\geq m^s\varepsilon\}) + \mu(\{\sum_{\ell:m^s<\ell<(m+1)^s}|h_\ell|\geq m^s\varepsilon\}) \\ &\qquad \text{since } |\sigma_n| \leq |\sigma_{m^s}| + \sum_{\ell:m^s<\ell<(m+1)^s}|h_\ell| \\ &\leq \frac{\|\sigma_{m^s}\|_p^p}{(m^s\varepsilon)^p} + \frac{\|\sum_{m^s<\ell<(m+1)^s}|h_\ell|\|_p^p}{(m^s\varepsilon)^p} \qquad \text{by Chebyshev's inequality} \\ &\leq C\varepsilon^{-p}\left(\frac{(m^s)^{(1-r)p}}{m^{sp}} + \frac{([(m+1)^s-m^s]C)^p}{m^{sp}}\right) \qquad \text{by (137) and (136)} \\ &\leq C\varepsilon^{-p}(m^{-prs}+m^{-p}) \end{aligned} \tag{138}$$

since $(m+1)^s - m^s \le C m^{s-1}$. Summing the estimate (138) over m gives $\sum_{m=1}^{\infty} \mu(X_m(\varepsilon)) < \infty$ because $prs > 1$ by choice of s while $p > 1$ by hypothesis. Therefore $\sum_{m=1}^{\infty} \mathbb{1}_{X_m(\varepsilon)} \in L^1$, and so $\sum_{m=1}^{\infty} \mathbb{1}_{X_m(\varepsilon)}(x)$ is finite μ-a.e. Thus

$$\limsup_{m\to\infty} \max_{n: m^s \le n < (m+1)^s} |\sigma_n(x)/n| \le 2\varepsilon \qquad \mu\text{-a.e.}$$

Letting $\varepsilon \to 0$ (through a countable set) now implies $\lim_{n\to\infty} |\sigma_n(x)/n| = 0$ μ-a.e. $\square$

Remark. We will not need the following fact, but the norm boundedness hypothesis (136) in Lemma 30 can be weakened to just

$$\sum_{n=1}^{\infty} \frac{\|h_n\|_p^p}{n^{1+\nu}} < \infty \qquad \text{for some } \nu \in (0, p(p-1)r). \tag{139}$$

Proof. Take $s = \max(1, (p-1)/\nu)$ in the proof above, noting that $s > 1/pr$ because $\nu < p(p-1)r$. Since this new number s need not be an integer, we replace σ_{m^s} in the proof of Lemma 30 with $\sigma_{\lfloor m^s \rfloor}$. And instead of estimating each $\|h_\ell\|_p$ with a constant C by (136), we now use:

$$\begin{aligned} m^{-sp} \left(\sum_{m^s < \ell < (m+1)^s} \|h_\ell\|_p \right)^p &\le \sum_{m^s < \ell < (m+1)^s} \|h_\ell\|_p^p \frac{((m+1)^s - m^s + 1)^{p-1}}{m^{sp}} \quad \text{by Hölder on the sum} \\ &\le C \sum_{m^s < \ell < (m+1)^s} \frac{\|h_\ell\|_p^p}{m^{s+p-1}} \qquad \text{since } (m+1)^s - m^s + 1 \le C m^{s-1} \\ &\le C \sum_{m^s < \ell < (m+1)^s} \frac{\|h_\ell\|_p^p}{\ell^{1+(p-1)/s}} = C \sum_{m^s < \ell < (m+1)^s} \frac{\|h_\ell\|_p^p}{\ell^{1+\nu}}. \end{aligned}$$

Summing this last quantity over m again gives a finite result, by the new assumption (139), which means we can complete the proof of Lemma 30 like before.

References

[1] A. Aldroubi and H.G. Feichtinger. *Exact iterative reconstruction algorithm for multivariate irregularly sampled functions in spline-like spaces: The L^p-theory.* Proc. AMS 126:2677–2686, 1998.

[2] A. Aldroubi and K. Gröchenig. *Nonuniform sampling and reconstruction in shift-invariant spaces.* SIAM Review 43:585–620, 2001.

[3] A. Aldroubi, Q. Sun and W.-S. Tang. *Non-uniform average sampling and reconstruction in multiply generated shift-invariant spaces.* Constructive Approx 20:173–189, 2004.

[4] A. Atzmon and A. Olevskiĭ. *Completeness of integer translates in function spaces on $\mathbb{R}$.* J Approx Theory 87:291–327, 1996.

[5] T. Aubin. Nonlinear analysis on manifolds. Springer–Verlag, New York, 1982.

[6] I. Babuška. *Approximation by hill functions.* Comment Math Univ Carolinae 11:787–811, 1970.

[7] S. Banach and S. Saks. *Sur la convergence forte dans les champs L^p.* Studia Math 2:51–57, 1930.

[8] J.J. Benedetto. *Irregular sampling and frames.* In Wavelets: A tutorial in theory and applications, pp 445–507. Ed. C.K. Chui, Academic Press, San Diego, 1992.

[9] C. de Boor, R. DeVore and A. Ron. *Approximation from shift-invariant subspaces of $L_2(\mathbb{R}^d)$.* Trans AMS 341:787–806, 1994.

[10] C. de Boor and R.-Q. Jia. *Controlled approximation and a characterization of the local approximation order.* Proc AMS 95:547–553, 1985.

[11] J. Bruna, A. Olevskiĭ and A. Ulanovskii. *Completeness in $L^1(\mathbb{R})$ of discrete translates.* Preprint, 2003.

[12] H.-Q. Bui and R. S. Laugesen. *Spanning and sampling in $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$.* Preprint, 2004. `www.math.uiuc.edu/~laugesen/publications.html`

[13] P.L. Butzer and R.L. Stens. *Sampling theory for not necessarily band-limited functions: A historical overview.* SIAM Review 34:40–53, 1992.

[14] P.L. Butzer, G. Schmeisser and R.L. Stens. *An introduction to sampling analysis.* In Non uniform sampling: Theory and practice, pp. 17–121. Ed. F. Marvasti, Kluver/Plenum, New York, 2001.

[15] A. Calogero. *A characterization of wavelets on general lattices.* J Geom Anal 10:597–622, 2000.

[16] W. Dahmen and C. A. Micchelli. *On the approximation order from certain multivariate spline spaces.* J Austral Math Soc Ser B 26:233–246, 1984.

[17] I. Daubechies. Ten lectures on wavelets. CBMS–NSF Regional Conference Series in Applied Mathematics, No. 61. Society for Industrial and Applied Mathematics, 1992.

[18] L. C. Evans and R. F Gariepy. Measure theory and fine properties of functions. CRC Press, Boca Raton, Florida, 1992.

[19] V. I. Filippov and P. Oswald. *Representation in L_p by series of translates and dilates of one function.* J Approx Th 82:15–29, 1995.

[20] G. Fix and G. Strang. *Fourier analysis of the finite element method in Ritz–Galerkin theory.* Stud Appl Math 48:265–273, 1969.

[21] G. B. Folland. Real analysis. Modern techniques and their applications. Second edition. John Wiley & Sons, New York, 1999.

[22] D. Gilbarg and N. S. Trudinger. Elliptic partial differential equations of second order. Springer–Verlag, New York, 1983.

[23] F. di Guglielmo. *Construction d'approximations des espaces de Sobolev sur des réseaux en simplexes.* Calcolo 6:279–331, 1969.

[24] E. J. Halton and W. A. Light. *On local and controlled approximation order.* J Approx Theory 72:268–277, 1993.

[25] J. E. Gilbert, Y. S. Han, J. A. Hogan, J. D. Lakey, D. Weiland and G. Weiss. *Smooth molecular decompositions of functions and singular integral operators.* Mem Amer Math Soc 156 (no. 742):1–74, 2002.

[26] E. Hernández and G. Weiss. A First Course on Wavelets. CRC Press, Boca Raton, Florida, 1996.

[27] R.-Q. Jia and J. Lei. *Approximation by multiinteger translates of functions having global support.* J Approx Theory 72:2–23, 1993.

[28] S. Kakutani. *Weak convergence in uniformly convex spaces.* Tohoku Math J 45:188–193, 1938.

[29] Y. Katznelson. An introduction to harmonic analysis. John Wiley & Sons, New York, 1968.

[30] J. Komlós. *A generalization of a problem of Steinhaus.* Acta Math Acad Sci Hungar 18:217–229, 1967.

[31] W. A. Light and E. W. Cheney. *Quasi-interpolation with translates of a function having noncompact support.* Constr Approx 8:35–48, 1992.

[32] V. Maz'ya and G. Schmidt. *"Approximate approximations" and the cubature of potentials.* Atti Accad Naz Lincei Cl Sci Fis Mat Natur Rend Lincei (9) Mat Appl 6:161–184, 1995.

[33] V. Maz'ya and G. Schmidt. *On approximate approximations using Gaussian kernels.* IMA J Numer Anal 16:13–29, 1996.

[34] Y. Meyer. Wavelets and operators. Cambridge University Press, Cambridge, 1992.

[35] S. G. Mikhlin. Approximation on a rectangular grid. Translated by R. S. Anderssen and T. O. Shaposhnikova. Sijthoff & Noordhoff, The Netherlands, 1979.

[36] F. Riesz and B. Sz.-Nagy. Functional analysis. Second edition. Translated by L. F. Boron. Ungar, New York, 1955.

[37] A. Ron. *Approximation orders of and approximation maps from local principal shift-invariant spaces.* J Approx Theory 81:38–65, 1995.

[38] J. M. Rosenblatt and K. L. Shuman. *Cyclic functions in $L^p(\mathbb{R}), 1 \leq p < \infty$.* J Fourier Anal Appl 9:181–192, 2003.

[39] W. Rudin. Functional Analysis. Second edition. McGraw-Hill, New York, 1991.

[40] G. Schmidt. *Approximate approximations and their applications.* In: The Maz'ya anniversary collection, Vol. 1 (Rostock, 1998), pp. 111–136. Oper Theory Adv Appl 109. Birkhäuser, Basel, 1999.

[41] I. J. Schoenberg. *Contributions to the problem of approximation of equidistant data by analytic functions. Parts A, B.* Quart Appl Math 4:45–99, 112–141, 1946.
[42] E. M. Stein and G. Weiss. Introduction to Fourier Analysis on Euclidean Spaces. Princeton University Press, Princeton, New Jersey, 1971.
[43] G. Strang. *The finite element method and approximation theory.* In: Numerical solution of partial differential equations, II (SYNSPADE 1970) (ed., B. Hubbard), pp. 547–583. Academic Press, New York, 1971.
[44] G. Strang and G. Fix. *A Fourier analysis of the finite element variational method.* In: Constructive aspects of functional analysis (ed. G. Geymonat), pp. 793–840. C.I.M.E., 1973.
[45] W. Szlenk. *Sur les suites faiblement convergentes dans l'espace L.* Studia Math 25:337–341, 1965.
[46] M. Unser. *Sampling – 50 years after Shannon.* Proc IEEE 88:569–587, 2000.
[47] J. Väisälä. Lectures on n-dimensional quasiconformal mappings. Springer–Verlag, New York, 1971.
[48] N. Wiener. *Tauberian theorems.* Ann Math 33:1–100, 1932.
[49] P. Wojtaszczyk. Banach spaces for analysts. Cambridge University Press, Cambridge, 1991.
[50] P. Wojtaszczyk. A mathematical introduction to wavelets. Cambridge University Press, Cambridge, 1997.
[51] W. P. Ziemer. Weakly differentiable functions. Springer–Verlag, New York, 1989.

Department of Mathematics, University of Canterbury, Christchurch 8020, New Zealand,
E-mail address: `Q.Bui@math.canterbury.ac.nz`

Department of Mathematics, University of Illinois, Urbana, IL 61801, U.S.A. ,
E-mail address: `Laugesen@math.uiuc.edu`