Sobolev spaces and approximation by affine spanning systems

H.-Q. Bui [1] and R.S. Laugesen [2]

[1] *Department of Mathematics & Statistics,*
University of Canterbury,
Private Bag 4800, Christchurch, New Zealand.

[2] *Department of Mathematics,*
University of Illinois,
Urbana, IL 61801, U.S.A.

Report Number: UCDMS2006/2 April 2006

Keywords: Completeness, quasi-interpolation, Strang-Fix, approximate identity, scale averaging.

SOBOLEV SPACES AND APPROXIMATION BY AFFINE SPANNING SYSTEMS

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

ABSTRACT. We develop conditions on a Sobolev function $\psi \in W^{m,p}(\mathbb{R}^d)$ such that if $\widehat{\psi}(0) = 1$ and ψ satisfies the Strang–Fix conditions to order $m-1$, then a scale averaged approximation formula holds for all $f \in W^{m,p}(\mathbb{R}^d)$:

$$f(x) = \lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} \sum_{k\in\mathbb{Z}^d} c_{j,k}\psi(a_j x - k) \qquad \text{in } W^{m,p}(\mathbb{R}^d).$$

The dilations $\{a_j\}$ are lacunary, for example $a_j = 2^j$, and the coefficients $c_{j,k}$ are explicit local averages of f, or even pointwise sampled values, when f has some smoothness.

For convergence just in $W^{m-1,p}(\mathbb{R}^d)$ the scale averaging is unnecessary and one has the simpler formula $f(x) = \lim_{j\to\infty} \sum_{k\in\mathbb{Z}^d} c_{j,k}\psi(a_j x - k)$. The Strang–Fix rates of approximation are recovered.

As a corollary of the scale averaged formula, we deduce new density or "spanning" criteria for the small scale affine system $\{\psi(a_j x - k) : j > 0, k \in \mathbb{Z}^d\}$ in $W^{m,p}(\mathbb{R}^d)$. We also span Sobolev space by derivatives and differences of affine systems, and we raise an open problem: does the Gaussian affine system span Sobolev space?

1. Introduction

We seek conditions on ψ under which every Sobolev function f can be approximated explicitly by linear combinations of the integer translates and small-scale dilates of ψ, that is by linear combinations of $\psi(a_j x - k)$ for $j > 0, k \in \mathbb{Z}^d$. The dilations a_j here are assumed to grow at least exponentially; for example $a_j = 2^j$. Our work on this *approximation problem* yields answers to the *spanning problem* of determining whether the $\psi(a_j x - k)$ span Sobolev space.

We illustrate our results now by stating them in one dimension, for the special case of Sobolev functions possessing one derivative.

Fix $1 \le p < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, take $\psi \in W^{1,p}$ and suppose $\phi \in L^q$ has compact support, for the remainder of this Introduction. Also assume

$$\widehat{\psi}(\ell) = 0 \qquad \text{for all integers } \ell \neq 0,$$

or equivalently that $\sum_{k\in\mathbb{Z}} \psi(x-k) \equiv \text{const}$.

Approximation results. Write

$$f_j(x) = \sum_{k\in\mathbb{Z}} \left(\int_{\mathbb{R}} f(a_j^{-1}y)\phi(y-k)\,dy \right) \psi(a_j x - k)$$

for the quasi-interpolant of f with analyzer ϕ and synthesizer ψ. Define a local supremum operator $Qf(x) = \|f\|_{L^\infty(x-1,x+1)}$. (See Sections 2 and 3 for more general definitions of Q and f_j.)

Theorem 1 proves scale averaged convergence for $f \in W^{1,p}$:

Date: September 16, 2005.

1991 *Mathematics Subject Classification.* Primary 41A25,42C40,46E35. Secondary 42B35,42C30.

Key words and phrases. Completeness, quasi-interpolation, Strang–Fix, approximate identity, scale averaging.

Laugesen was partially supported by N.S.F. Award DMS-0140481, a Maclaurin Fellowship from the New Zealand Institute of Mathematics and its Applications, and a Visiting Erskine Fellowship from the University of Canterbury.

$$f = \lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} f_j \quad \text{if } Q\psi, Q((1+|x|)\psi') \in L^1 \text{ and } \widehat{\psi}(0) = \widehat{\phi}(0) = 1.$$

(These Q-hypotheses say roughly that ψ and $(1+|x|)\psi'$ are bounded and decay integrably at infinity.)

Theorem 4 implies the same for the pointwise quasi-interpolant $f_j^\bullet(x) = \sum_{k\in\mathbb{Z}} f(a_j^{-1}k)\psi(a_j x - k)$, provided $f \in W^{1,p} \cap C^1$ with $Qf, Q(f') \in L^p$.

These two theorems give convergence rate $o(1)$ in the $W^{1,p}$ norm. For the L^p norm, the "Strang–Fix" rate of convergence $O(|a_j|^{-1})$ is obtained as expected, by *Theorem 6*:

> if $Q((1+|x|)\psi) \in L^1$ and $\widehat{\psi}(0) = \widehat{\phi}(0) = 1$ and $\widehat{\phi}'(0) = 0$, then $\|f - f_j\|_p \le C|f|_{W^{1,p}}|a_j|^{-1} = O(|a_j|^{-1})$ as $j \to \infty$, for each $f \in W^{1,p}$.

Spanning results. *Corollary 8* deduces that:

> if ψ' decays like $|x|^{-2-\varepsilon}$ at infinity, and $\widehat{\psi}(0) \neq 0$, then the small scale affine system $\{\psi(a_j x - k) : j > 0, k \in \mathbb{Z}\}$ spans $W^{1,p}$.

Finally, taking derivatives and differences of known spanning systems will generate yet more spanning systems, as *Proposition 9* and *Theorem 10* explain.

Outline of the paper. The standing assumptions on dilations and translations are established in Section 2, along with some definitions. Section 3 gives approximation formulas for $W^{m,p}(\mathbb{R}^d)$, with relevant literature summarized in Section 3.5. Spanning results are deduced in Section 4.

Spanning properties of the second difference of the Gaussian are determined in Section 5. Spanning properties of the second derivative of the Gaussian, a function known as the Mexican hat, remain mostly unknown. This open problem is related in Section 5 to a spanning conjecture for the Gaussian $\psi(x) = e^{-x^2/2}$ itself: does the small scale dyadic system $\{\psi(2^j x - k) : j > 0, k \in \mathbb{Z}\}$ span $W^{m,p}$?

The technical core of the paper is in Section 6, where discretized approximate identities are studied and scale averaging is introduced through formula (22). Then Theorems 1, 4 and 6 are proved in Sections 8, 9 and 11, after which appear the remaining proofs and an appendix about the Q-operator.

Remark. This paper builds on our L^p results in [5]. The Hardy space H^1 was treated in [7].

2. Definitions and notation

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

2. Let the *dilations* a_j for $j > 0$ be nonzero real numbers with $|a_j| \to \infty$ as $j \to \infty$. Define $a_{min} = \min_{j>0} |a_j|$.

Some of our results further assume the dilations *grow exponentially*, meaning $|a_{j+1}| \ge \gamma |a_j|$ for all $j > 0$, for some $\gamma > 1$ (so that the dilation sequence is lacunary).

3. Fix a *translation* matrix b, assumed to be an invertible $d \times d$ real matrix. Some of our constants and operators in this paper will depend implicitly on b and the dimension d.

4. Write $L^p = L^p(\mathbb{R}^d)$ for the class of complex valued functions with finite L^p-norm, and $W^{m,p} = W^{m,p}(\mathbb{R}^d)$ for the Sobolev functions with m derivatives in L^p. Given a multiindex μ of order $|\mu| = \mu_1 + \cdots + \mu_d$, we write $f^{(\mu)} = D^\mu f$ for the μ-th derivative of f.

5. Given $\psi \in L^p$ and $\phi \in L^q$, where by convention

$$\frac{1}{p} + \frac{1}{q} = 1,$$

we define

$$\psi_{j,k}(x) = |a_j|^{d/p}\psi(a_j x - bk), \qquad \phi_{j,k}(x) = |a_j|^{d/q}\phi(a_j x - bk), \qquad x \in \mathbb{R}^d,$$

for $j > 0, k \in \mathbb{Z}^d$. These rescalings satisfy $\|\psi_{j,k}\|_p = \|\psi\|_p$ and $\|\phi_{j,k}\|_q = \|\phi\|_q$.

6. The *periodization* of a function f is

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

If $f \in L^1$, then this series for Pf converges absolutely for almost every x, and Pf is locally integrable.

7. Define a *local supremum* operator

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

Appendix A explains some properties of Qf.

8. Write $\chi(x) = 1 + |x|$ and $X(x) = x$, for $x \in \mathbb{R}^d$.

9. Define the Fourier transform with 2π in the exponent: $\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x)e^{-2\pi i\xi x}\,dx$, for row vectors $\xi \in \mathbb{R}^d$.

10. A subset U of a topological vector space V 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

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

3. Approximation results

In this section we state our two main approximation theorems, for average sampling and pointwise sampling respectively, and then we extend them to give rates of approximation. At the end of the section we discuss related literature and prior results.

3.1. Approximation by average sampling. We define an *approximation to f at scale j* by

$$\begin{aligned} f_j(x) &= |\det b| \sum_{k\in\mathbb{Z}^d} \langle f, \overline{\phi}_{j,k}\rangle \psi_{j,k} \\ &= |\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} f(a_j^{-1}y)\phi(y - bk)\,dy \right) \psi(a_j x - bk), \qquad j > 0, \quad x \in \mathbb{R}^d, \end{aligned} \tag{1}$$

where f is the *signal*, ϕ is the *analyzer* and ψ is the *synthesizer*. To understand f_j, suppose ϕ is a delta function (like in Theorem 4 below); then with $b = I$ we get the quasi-interpolant $f_j(x) = \sum_{k\in\mathbb{Z}^d} f(a_j^{-1}k)\psi(a_j x - k)$.

Our first theorem finds conditions under which the f_j provide a good approximation to f.

Theorem 1. *Assume $\psi \in W^{m,p}$ for some $1 \le p < \infty, m \in \mathbb{N}$, and suppose one of the following conditions holds:*

(i) *$P|\chi^{|\mu|}\psi^{(\mu)}| \in L^p_{loc}$ for all $|\mu| \le m$, and $\chi^m\phi \in L^1$, and $f \in C_c^m$;*

(ii) *$Q(\chi^{|\mu|}\psi^{(\mu)}) \in L^1$ for all $|\mu| \le m$, and $\phi \in L^q$ with ϕ having compact support, and $f \in W^{m,p}$.*

Suppose

$$D^\mu\widehat{\psi}(\ell b^{-1}) = 0 \qquad \text{for all row vectors } \ell \in \mathbb{Z}^d \setminus \{0\}, \tag{2}$$

whenever $|\mu| < m$.

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

(a) [Strang–Fix approximation] If in addition (2) holds whenever $|\mu| = m$, then

$$f = \lim_{j\to\infty} f_j \qquad \text{in } W^{m,p}. \tag{3}$$

(b) [Scale-averaged approximation] If the dilations a_j grow exponentially, then

$$f = \lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} f_j \qquad \text{in } W^{m,p}.$$

(c) [Stability] If (ii) holds then $\|f_j\|_{W^{m,p}} \le C(\psi, \phi, m, p)\|f\|_{W^{m,p}}$ for all $j > 0$.
(d) [Span] $f_j \in W^{m,p}$-span$\{\psi_{j,k} : k \in \mathbb{Z}^d\}$ for each $j > 0$.

The proof is in Section 8.

Examples. A decay condition near infinity guarantees hypothesis (i) on ψ:

Lemma 2. *Let $\psi \in W^{m,p}$ for some $1 \le p < \infty, m \in \mathbb{N}$. If*

$$|\psi^{(\mu)}(x)| \le C|x|^{-d-m-\epsilon} \qquad \text{for each } |\mu| = m \text{ and almost every } x \text{ with } |x| > R, \tag{4}$$

for some constants $C, R, \epsilon > 0$, then $P|\chi^{|\mu|}\psi^{(\mu)}| \in L^p_{loc}$ for all $|\mu| \le m$.

Hypothesis (ii) holds if ψ and its derivatives are *bounded* and decay at infinity:

Lemma 3. *Let $\psi \in W^{m,\infty}$ for some $m \in \mathbb{N}$. If decay condition (4) holds, then $Q(\chi^{|\mu|}\psi^{(\mu)}) \in L^1$ for all $|\mu| \le m$.*

Lemma 2 and 3 are proved in Appendix A.

Notes on Theorem 1.

The L^p result corresponding to Theorem 1 is [5, Theorem 1]. The hypotheses there are roughly the same as case (i) with $m = 0$, except that f need not be continuous with compact support. Precisely, the L^p result assumes that $\psi \in L^p$ with $P|\psi| \in L^p_{loc}$, $\phi \in L^q$ with $P|\phi| \in L^\infty$, and $f \in L^p$. The reason our Sobolev result Theorem 1 can only handle $f \in C^m_c$, in case (i), boils down to our inability to prove a stability estimate in Lemma 11 case (i) for the general function $h(x, y)$.

Case (ii) assumes more on ψ than case (i) does (because $Q(\cdot) \in L^1$ implies $P|\cdot| \in L^\infty$ by [6, Lemma 23]). But case (ii) has the advantage of applying to all $f \in W^{m,p}$ and not just to $f \in C^m_c$. Also, case (ii) yields a stability estimate in Theorem 1(c).

We call condition (2) the *Strang–Fix condition* of order $m-1$, in view of the work of Strang and Fix in [13, 29, 30] (although historically, Schoenberg [28, Theorem 2] seems to have been the first to use the condition, in the context of polynomial interpolation and smoothing in one dimension). The Strang–Fix condition can be satisfied formally by putting

$$\psi = u * \cdots * u * \psi_0 \qquad \text{(with } m \text{ factors of } u) \tag{5}$$

where u has constant periodization $Pu = 1$ a.e. (meaning the integer translates of u form a partition of unity). Indeed $Pu = 1$ a.e. implies $\widehat{u}(0) = 1$ and $\widehat{u}(\ell b^{-1}) = 0$ for all $\ell \in \mathbb{Z}^d \setminus \{0\}$, by computing the Fourier coefficients of the $b\mathbb{Z}^d$-periodic function Pu, and thus the Strang–Fix condition (2) follows from the fact that $\widehat{\psi} = \widehat{u} \cdots \widehat{u}\,\widehat{\psi_0}$. For a different interpretation of the Strang–Fix condition, in terms of periodizations of moments of ψ, see Section 7.

Our methods for Theorem 1 extend to cover dilation *matrices* a_j that expand both exponentially ($\sup_{j>0} \|a_j a_{j+1}^{-1}\| < 1$) and nicely ($\|a_j^{-1}\|^d \le C|\det a_j^{-1}|$); see [6, §7]. But our method breaks down for dilations like $\begin{pmatrix} 3^j & 0 \\ 0 & 2^j \end{pmatrix}$ that do not expand nicely.

Relevant literature for Theorem 1 will be discussed in Section 3.5.

3.2. **Properties of f_j.** Observe that f_j discretizes a classical approximation to the identity:

$$\begin{aligned} f(x) &= \lim_{j\to\infty} (f * \psi_{a_j^{-1}})(x) \\ &= \lim_{j\to\infty} \int_{\mathbb{R}^d} f(z)|a_j|^d \psi(a_j(x-z))\,dz \\ &= \lim_{j\to\infty} \int_{\mathbb{R}^d} f(a_j^{-1}y)\psi(a_jx-y)\,dy \qquad \text{by } z=a_j^{-1}y \\ &\approx \lim_{j\to\infty} \sum_{k\in\mathbb{Z}^d} \left(\int_{k+C} f(a_j^{-1}y)\,dy \right) \psi(a_jx-k) \end{aligned} \tag{6}$$

by a Riemann sum approximation. This last line (6) is exactly $\lim_{j\to\infty} f_j$, with $\phi = \mathbb{1}_C$ and $b = I$. Caution is required in the Riemann sum approximation step, because we discretize with fixed step size 1. Theorem 1(a)(b) nonetheless shows the approximation (6) is exact in the $W^{m,p}$-norm as $j \to \infty$ provided either ψ satisfies Strang–Fix conditions to order m or else ψ satisfies them to order $m-1$ and the approximation formula is averaged over all dilation scales.

Second, we can express f_j in terms of an integral kernel as $f_j(x) = \int_{\mathbb{R}^d} K_j(x,y)f(y)\,dy$ where

$$K_j(x,y) = |a_j|^d K(a_jx, a_jy) \quad \text{and} \quad K(x,y) = |\det b| \sum_{k\in\mathbb{Z}^d} \psi(x-bk)\phi(y-bk).$$

The stability estimate in Theorem 1(c) says that $K_j : W^{m,p} \to W^{m,p}$ with a norm estimate that is independent of j, provided hypothesis (ii) holds.

3.3. **Approximation using pointwise sampling.** Now we develop an analogue of Theorem 1 that uses *pointwise* sampling. Write

$$f_j^\bullet(x) = |\det b| \sum_{k\in\mathbb{Z}^d} f(a_j^{-1}bk)\psi(a_jx-bk) \tag{7}$$

for the quasi-interpolant of f at scale j, sampled on the uniform grid $a_j^{-1}b\mathbb{Z}^d$. The "$\bullet$" notation refers to the pointwise nature of the sampling.

Theorem 4. *Assume $\psi \in W^{m,p}$ for some $1 \le p < \infty, m \in \mathbb{N}$, and that $Q(\chi^{|\mu|}\psi^{(\mu)}) \in L^1$ for all $|\mu| \le m$, and $f \in W^{m,p} \cap C^m$ with $Q(f^{(\mu)}) \in L^p$ for all $|\mu| \le m$. Suppose*

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

whenever $|\mu| < m$. Assume $\int_{\mathbb{R}^d} \psi\,dx = 1$. Then (a)–(d) hold:

(a) [Strang–Fix approximation] If in addition (8) holds whenever $|\mu| = m$, then

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

(b) [Scale-averaged approximation] If the dilations a_j grow exponentially, then

$$f = \lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} f_j^\bullet \qquad \text{in } W^{m,p}.$$

(c) [Stability] $\|f_j^\bullet\|_{W^{m,p}} \le C(\psi, m, p, a_{min}) \sum_{|\mu|\le m} \|Q(f^{(\mu)})\|_p$ for all $j > 0$, where $a_{min} = \min_{j>0} |a_j|$.

(d) [Span] $f_j^\bullet \in W^{m,p}$-span$\{\psi_{j,k} : k \in \mathbb{Z}^d\}$ for each $j > 0$.

See Section 9 for the proof.

The C^m-smoothness of f in the theorem is convenient, but it could be weakened like in the corresponding L^p result [5, Theorem 2].

For simplicity, Theorem 4 is stated only with hypothesis (ii) from Theorem 1, although it can be proved under hypothesis (i) also.

3.4. **Approximation rates.** The preceding two theorems can be adapted to give explicit rates of approximation of f_j to f. But we must first construct analyzers and synthesizers with suitably normalized moments.

Lemma 5. *Suppose $\phi, \psi \in L^1$ with $\chi^m \phi, \chi^{m-1}\psi \in L^1$ for some $m \in \mathbb{N}$. If $\int_{\mathbb{R}^d} \phi \, dx \neq 0$ and $\int_{\mathbb{R}^d} \psi \, dx \neq 0$, then there exists a finite set $K \subset \mathbb{Z}^d$ and coefficients $\alpha_k, \beta_k \in \mathbb{C}$ for $k \in K$ 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

$$\int_{\mathbb{R}^d} x^\mu \Phi(x)\, dx = \begin{cases} 1 & \text{if } \mu = 0, \\ 0 & \text{if } 0 < |\mu| \leq m, \end{cases} \qquad \int_{\mathbb{R}^d} (-x)^\mu \Psi(x)\, dx = \begin{cases} 1 & \text{if } \mu = 0, \\ 0 & \text{if } 0 < |\mu| \leq m-1. \end{cases} \tag{9}$$

The proof is in Section 10, along with examples of how to construct the linear combinations for Φ and Ψ.

Now we can determine the *rate* at which f_j approximates $f \in W^{m,p}$ in the $W^{r,p}$-norm, for $0 \leq r \leq m$. Recall $|f|_{W^{r,p}} = \left(\sum_{|\mu|=r} \|D^\mu f\|_p^p \right)^{1/p}$ is the Sobolev seminorm.

Theorem 6. *Assume $\psi \in W^{m-1,p}$ for some $1 \leq p < \infty, m \in \mathbb{N}$, with $Q(\chi^m \psi^{(\mu)}) \in L^1$ for all $|\mu| < m$, and take $\phi \in L^q$ with compact support. Suppose*

$$D^\mu \widehat{\psi}(\ell b^{-1}) = 0 \qquad \text{for all row vectors } \ell \in \mathbb{Z}^d \setminus \{0\},$$

whenever $|\mu| < m$.

Assume $\int_{\mathbb{R}^d} \psi \, dx \neq 0, \int_{\mathbb{R}^d} \phi \, dx \neq 0$, and that Ψ and Φ are as in Lemma 5.

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

$$|F_j - f|_{W^{r,p}} \leq C(\psi, \phi, m, p) |f|_{W^{m,p}} |a_j|^{r-m} = O(|a_j|^{r-m}) \qquad \text{for all } j > 0,$$

where F_j is defined by average sampling with analyzer Φ and synthesizer Ψ:

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

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

$$|F_j^\bullet - f|_{W^{r,p}} \leq C(\psi, m, p, a_{min}) \sum_{|\mu|=m} \|Q(f^{(\mu)})\|_p |a_j|^{r-m} = O(|a_j|^{r-m}) \qquad \text{for all } j > 0,$$

where $a_{min} = \min_{j>0} |a_j|$ and $F_j^\bullet$ is defined by uniform pointwise sampling with synthesizer Ψ:

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

The theorem is proved in Section 11.

Remarks on Theorem 6.

1. Theorems 1 and 4 can give further information on F_j and $F_j^\bullet$, such as stability estimates.

2. Theorem 6 does not consider case (i) of Theorem 1, because stability estimates underpin the proof and we only know stability in case (ii).

3. The scale averaging technique in Theorems 1(b) and 4(b) does *not* help obtain rates of approximation. The problem, when one digs into the proofs, is that the scale averaged periodization $\frac{1}{J}\sum_{j=1}^{J} P\psi(a_j x)$ will generally *fail* to converge uniformly to its mean value; in particular, convergence fails at $x = 0$ if $P\psi(0)$ is not equal to the mean value and $P\psi$ is continuous.

4. Theorem 6(a) implies that $F_j \to f$ in $W^{m-1,p}$, so that the $\psi_{j,k}$ span $W^{m,p}$ using the $W^{m-1,p}$-norm. In fact the $\psi_{j,k}$ span $W^{m,p}$ in its own norm, by Corollary 7. This illustrates the "gain of one order" provided by scale averaging, in our work.

3.5. **The Sobolev approximation literature and prior results.**

Overview. Our main contribution in Section 3 is the scale averaged approximation in Theorem 1(b), which is genuinely new.

The pointwise sampling results in Theorem 4 and Theorem 6(b) seem also to be new.

The average sampling results (big-O approximation rates) in Theorem 6(a) are essentially known.

Detailed discussion. We now give a more complete account of the literature, and our contributions.

The original approximation results in $W^{m,p}$ with $\int_{\mathbb{R}^d} \psi\,dx \neq 0$ all assume that $p = 2$ and ψ has compact support. See Babuška [4, Theorem 4.1] and Strang and Fix [30, Theorem I]. These approximation formulas are not explicit, in the sense that they use sampled values of $\hat{f}$, rather than of f, to construct an approximation to f by Fourier transform methods. These indirect Fourier methods are characteristic of the work of Strang and Fix and most of the papers inspired by them. By contrast, we work with explicit quasi-interpolants in this paper, namely the functions $f_j(x)$.

Di Guglielmo had earlier proved an explicit approximation result [16, Théorème 6] for $p = 2$, provided also ψ is a convolution like in (5) with u being the characteristic function of a unit 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 $i = 1, \dots, d\}$, and so $\widehat{\psi}$ vanishes on all these hyperplanes too, instead of just vanishing at the lattice points (where hyperplanes intersect) like in the work of Babuška, Strang and Fix.

For $p = 2$, these authors all prove big-O approximation rates that are analogous to our Theorem 6(a). That is, they show an arbitrary $f \in W^{m,2}$ can be approximated in the $W^{r,2}$ norm at rate $O(|a_j|^{r-m})$ as $j \to \infty$, for each $r = 0, 1, \dots, m-1$. The best possible result of this kind is due to Jetter and Zhou [19, Theorem 1], who completely characterized the functions ψ and ϕ for which these approximation rates can hold, when $p = 2$. See also Holtz and Ron [18, Theorems 7,9].

For all $1 \leq p < \infty$, Jia [20, Theorem 3.1] has proved analogous approximation rates under the assumption that ψ and ϕ have compact support. Thus Theorem 6(a) is known already in the compactly supported case. Jia's proof is different to ours, although both proofs avoid the Fourier transform and hence can treat $p \neq 2$ along with $p = 2$.

Theorem 6(a) improves on all these results in a technical sense (except for Jetter–Zhou and Holtz–Ron when $p = 2$), because the hypothesis $Q(\chi^m \psi^{(\mu)}) \in L^1$ can hold even when ψ does not have compact support.

Much more attention has been paid in the literature to the case $r = 0$ of Theorem 6(a) (approximation of Sobolev functions in the L^p-norm) than to the case $r > 0$ (approximation of Sobolev functions in Sobolev norms). See for example [22, §7] for all p, and the references in [18] for $p = 2$.

The approximation rates in Theorem 6(b), for pointwise sampling, seem to be new except when $p = 2$, which was considered by Jetter and Zhou [19, Theorem 5] provided $m > d/2$. When

$p = \infty$ (uniform approximation), Strang and Fix [30, Theorem III] did use pointwise sampling to approximate f in the $W^{r,\infty}$ norm, and other authors have since extended those results.

Turn now to Theorem 1. Part (a) was essentially proved by Di Guglielmo [16, Théorème 2′] for $p \geq 2$, under the strong "convolution" assumption on ψ mentioned above. Notice Theorem 1(a) only gives convergence at the rate $o(1)$, although to its credit this is accomplished without the vanishing moment assumption on the analyzer and synthesizer needed in Theorem 6.

Theorem 1(b) proves scale averaged convergence, which is new. We are aware of no precedents in the Strang–Fix tradition or in related approximation theory. Note that Theorem 1(b) gives convergence in the $W^{m,p}$-norm in a situation where Strang–Fix type results like Theorem 6 can only prove convergence in the $W^{m-1,p}$-norm (because the Strang–Fix condition is assumed only to order $m-1$). The convergence rate in Theorem 1 is merely $o(1)$, but that will later suffice to yield interesting spanning results, in Section 4.

Lastly, Theorem 4 is the analogue of Theorem 1 for approximation in $W^{m,p}$ by *pointwise* sampling. It seems not to have direct forbears in the literature.

Additional remarks. Strang and Fix [30, Theorem I] proved a converse saying that the Strang–Fix condition to order $m-1$ is *necessary* for approximating an arbitrary $f \in W^{m,2}$ in the $W^{r,2}$-norm (for $r = 0, 1, \ldots, m-1$) at rate $O(|a_j|^{r-m})$ in a "controlled" fashion by functions of the form $\sum_{k \in \mathbb{Z}^d} c_{j,k}\psi_{j,k}$. The point of Theorem 6(a) in this paper is to prove *sufficient* conditions under which the quasi-interpolant f_j achieves this best possible rate of approximation. We do not consider necessary conditions.

Mikhlin's monograph [26] develops Strang–Fix type approximation results using "primitive functions". Unfortunately the number of such generators must grow with m.

Maz'ya and Schmidt [23, 24, 27] developed a theory of *approximate approximations* that can be viewed as Strang–Fix theory without the full Strang–Fix conditions. Their approximations possess inescapable saturation errors and thus do not actually converge. Nonetheless, Maz'ya and Schmidt make a case that the saturation errors can be negligible in practical situations.

4. Spanning results — synthesizers and their derivatives and differences

First we deduce spanning results from our earlier approximation theorems.

Corollary 7. *Assume $\psi \in W^{m,p}$ for some $1 \leq p < \infty, m \in \mathbb{N}$, and suppose $P|\chi^{|\mu|}\psi^{(\mu)}| \in L^p_{loc}$ for all $|\mu| \leq m$. Assume*

$$D^\mu \widehat{\psi}(\ell b^{-1}) = 0 \qquad \text{for all row vectors } \ell \in \mathbb{Z}^d \setminus \{0\},$$

whenever $|\mu| < m$.

If $\int_{\mathbb{R}^d} \psi \, dx \neq 0$, then $\{\psi_{j,k} : j > 0, k \in \mathbb{Z}^d\}$ spans $W^{m,p}$.

Spanning means the finite linear combinations of the functions $\psi_{j,k}$ are dense in $W^{m,p}$.

The analogous L^p spanning result [5, Corollary 1] holds when $P|\psi| \in L^p_{loc}$.

Proof of Corollary 7. The dilations a_j can be taken to grow exponentially, by passing to a subsequence if necessary. And we can require $\int_{\mathbb{R}^d} \psi \, dx = 1$, since multiplying ψ by a nonzero constant does not affect the span of the $\psi_{j,k}$. Let ϕ be the characteristic function of a unit cube.

Then the $W^{m,p}$-span of the $\psi_{j,k}$ contains C_c^m, by Theorem 1(b)(d) case (i). By density of C_c^m, the $\psi_{j,k}$ therefore span all of $W^{m,p}$. □

Next we conclude that a simple decay condition near infinity suffices for the $\psi_{j,k}$ to span $W^{m,p}$, in conjunction with the Strang–Fix vanishing of the Fourier transform at the lattice points, to order $m-1$.

Corollary 8. *Assume $\psi \in W^{m,p}$ for some $1 \le p < \infty, m \in \mathbb{N}$, and that ψ decays according to*

$$|\psi^{(\mu)}(x)| \le C|x|^{-d-m-\epsilon} \qquad \text{for each } |\mu| = m \text{ and all large } |x|,$$

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

If $\int_{\mathbb{R}^d} \psi \, dx \neq 0$, then $\{\psi_{j,k} : j > 0, k \in \mathbb{Z}^d\}$ spans $W^{m,p}$.

To prove the corollary, just combine Corollary 7 with Lemma 2.

The analogous L^p result ($m = 0$) is in [5, Corollary 2]. We are not aware of any previous spanning results of this kind for Sobolev space.

Our next result spans by *derivatives* of a given spanning set.

Proposition 9. *Let $\mathcal{H} \subset W^{m,p}$ for some $1 < p < \infty, m \in \mathbb{N}$, and suppose $\mathcal{H}$ spans $W^{m,p}$.*

If ν is a multiindex of order $0 < |\nu| \le m$, then the collection $\{D^\nu h : h \in \mathcal{H}\}$ spans $W^{m-|\nu|,p}$.

The proposition is proved in Section 12. Clearly it fails for $p = 1$, since $\int_{\mathbb{R}^d} D^\nu h \, dx = 0$ always.

Example for Proposition 9. If ψ satisfies the hypotheses of Corollary 7 or 8, and $1 < p < \infty$, then the $(D^\nu \psi)_{j,k}$ span $W^{m-|\nu|,p}$ by Proposition 9. In particular, they span L^p when $|\nu| = m$.

Note the Fourier transform of our new affine generator $D^\nu \psi$ vanishes at *all* lattice points, with

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

whenever $\ell \in \mathbb{Z}^d \setminus \{0\}$ and $|\mu| < m$, and also whenever $\ell = 0$ and $\mu < \nu$.

Our final result shows that in most cases, the span of an affine system is not changed by taking *differences* of the generator. Our notation for first differences is

$$\Delta_{c,z}\psi(x) = \psi(x) - c\psi(x - z), \qquad c \in \mathbb{C}, \quad x, z \in \mathbb{R}^d.$$

When $c = 1$ we simply write $\Delta_z \psi(x) = \psi(x) - \psi(x - z)$.

Theorem 10. *Suppose $\psi \in W^{m,p}$ for some $1 \le p \le \infty, m \in \mathbb{N} \cup \{0\}$. Fix $j > 0$. Take $c \in \mathbb{C}$ and $\kappa \in \mathbb{Z}^d \setminus \{0\}$.*

If $1 < p < \infty$ or $|c| \neq 1$, then

$$W^{m,p}\text{-}span\{\psi_{j,k} : k \in \mathbb{Z}^d\} = W^{m,p}\text{-}span\{(\Delta_{c,b\kappa}\psi)_{j,k} : k \in \mathbb{Z}^d\}.$$

See Section 13 for the proof. Notice L^p-spaces are covered by the theorem (when $m = 0$).

Example for Theorem 10. Work in dimension $d = 1$ for simplicity. If $\psi \in L^\infty$ has compact support and $\int_{\mathbb{R}} \psi \, dx \neq 0$, then the small-scale affine system $\{\psi_{j,k} : j > 0, k \in \mathbb{Z}\}$ spans $L^p(\mathbb{R})$ for each $1 \le p < \infty$, by [5, Corollary 2]. Then Theorem 10 with $c = 1$ and $\kappa = 1$ implies that each $L^p(\mathbb{R}), 1 < p < \infty$, is also spanned by the small-scale affine systems generated by each of

$$\Delta\psi(x) = \psi(x) - \psi(x - b),$$
$$\Delta^2\psi(x) = \psi(x) - 2\psi(x - b) + \psi(x - 2b),$$

and so on.

For example the Haar wavelet $H = \mathbb{1}_{[0,1/2)} - \mathbb{1}_{[1/2,1)}$ can be written as a difference $H = \frac{1}{2}\Delta\psi$ of the function $\psi = 2\mathbb{1}_{[0,1/2)}$, provided $b = 1/2$, and so the oversampled, small-scale Haar system

$$\{H(2^{j+J}x - \frac{1}{2}k) : j > 0, k \in \mathbb{Z}\}$$

spans L^p for $1 < p < \infty$, for each $J \in \mathbb{N}$, by taking $a_j = 2^{j+J}$ above. Recall that the Haar system $\{H(2^j x - k) : j \in \mathbb{Z}, k \in \mathbb{Z}\}$ with no oversampling also spans L^p [17], though it needs all dilation scales $j \in \mathbb{Z}$ to do so.

Remark. The spanning by differences result in Theorem 10 is weaker (in the interesting case $c = 1$) than the spanning by derivatives result in Proposition 9. For suppose we want to span L^p. A difference of a function $\psi \in L^p$ will have Fourier transform vanishing on infinitely many hyperplanes (*e.g.* the unit difference $\psi(x) - \psi(x - e_1)$ in the x_1 direction has a factor of $1 - e^{-2\pi i \xi_1}$ in its Fourier transform, and this factor vanishes whenever $\xi_1 \in \mathbb{Z}$). If instead we started with $\psi \in W^{1,p}$ and then took a derivative such as $D_1\psi$, we would introduce zeros only on the single hyperplane $\xi_1 = 0$ through the origin in Fourier space; we would also need to impose a Strang–Fix condition $\widehat{\psi} = 0$ at the nonzero lattice points, to ensure that the $\psi_{j,k}$ span $W^{1,p}$ by our results (like Corollary 8) and hence that their derivatives span L^p. The upshot, though, is that when spanning by differences one needs $\widehat{\psi}$ to vanish on infinitely many hyperplanes, whereas when spanning by derivatives one only needs $\widehat{\psi}$ to vanish on one hyperplane and infinitely many lattice points.

Of course in dimension $d = 1$ the two approaches are equivalent, because hyperplanes reduce to points. And anyway, differences can be more convenient to use than derivatives.

Spanning by molecular and wavelet affine systems. The work of Gilbert *et al.* [15], and earlier Frazier and Jawerth [14], gives an affine spanning result in the homogeneous Sobolev space $\dot{W}^{m,p}, 1 < p < \infty$. In particular, the result [15, Theorem 1.5] proves a frame decomposition using the full affine system $\{\psi(a^j x - bk) : j \in \mathbb{Z}, k \in \mathbb{Z}^d\}$ provided ψ satisfies certain "molecular" decay and smoothness conditions. Hence the system spans $\dot{W}^{m,p}$. Strang–Fix conditions are not imposed. Unfortunately, [15, Theorem 1.5] holds only when the dilation step a is sufficiently close to 1 and the translation step b is sufficiently close to 0, depending on the synthesizer ψ. By contrast, in this paper our dilations and translations are independent of ψ.

In a different direction, orthonormal wavelet systems $\{\psi(2^j x - k) : j \in \mathbb{Z}, k \in \mathbb{Z}^d\}$ that satisfy some smoothness and decay conditions are known to provide unconditional bases for Sobolev space [17, p. 312], and hence span Sobolev space. See [21] for recent developments.

These molecular and wavelet results employ all the scales $j \in \mathbb{Z}$, and assume $\widehat{\psi}(0) = 0$. In contrast, this paper uses just the small scales $j > 0$ and assumes $\widehat{\psi}(0) \neq 0$. (The only generators with $\widehat{\psi}(0) = 0$ in this paper are those resulting from Proposition 9 when spanning by derivatives, and from Theorem 10 when spanning by differences.)

5. Open problems — the Gaussian and the Mexican hat

Our work in this paper on Sobolev space, and our earlier work on L^p in [5] and H^1 in [7], are motivated by Y. Meyer's unsolved "Mexican hat" problem. To describe it, consider now dyadic dilations $a_j = 2^j$ in dimension $d = 1$, and for simplicity take $b = 1$ throughout this section. Write $\theta(x) = (1 - x^2)e^{-x^2/2}$ for the Mexican hat function (whose graph resembles a sombrero).

Meyer [25, p. 137] asked: does the full Mexican hat system $\{\theta_{j,k} : j, k \in \mathbb{Z}\}$ span L^p for all $1 < p < \infty$? (It cannot span all of L^1 because the Mexican hat has integral zero.) The answer is Yes when $p = 2$, but the problem remains open for all other p-values. It is known that the Mexican hat system spans L^p provided the translations are sufficiently oversampled [8], or the translations and dilations are both sufficiently oversampled [15].

We propose a different approach. The Mexican hat is the second derivative of the Gaussian $-e^{-x^2/2}$, and so we wonder whether the Gaussian system spans Sobolev space.

Conjecture 1. *If $\psi(x) = e^{-x^2/2}$ then $\{\psi_{j,k} : j > 0, k \in \mathbb{Z}\}$ spans $W^{m,p}$ for each $1 \leq p < \infty, m \in \mathbb{N}$.*

If Conjecture 1 is true, then for all $m \in \mathbb{N}$, the mth derivative of the Gaussian ψ would generate a small scale system spanning $L^p, 1 < p < \infty$, by Proposition 9. In particular by taking $m = 2$, the small scale Mexican hat system $\{\theta_{j,k} : j > 0, k \in \mathbb{Z}\}$ would span L^p, answering Meyer's question.

Notice Conjecture 1 is true for $m = 0$ (the L^p case), by [5, Corollary 2] or earlier by [12] for dyadic dilations. Also note Corollary 8 fails to resolve the conjecture for $m > 0$, because the Fourier transform of the Gaussian vanishes nowhere and thus fails the Strang–Fix hypothesis $D^\mu \widehat{\psi}(\ell b^{-1}) = 0$ imposed in Corollary 8.

Conjecture 1 must be approached with caution, because not every reasonable ψ generates a system that spans Sobolev space. For example the tent function $\psi(x) = 2x$ for $x \in [0, 1/2]$ and $\psi(x) = 2 - 2x$ for $x \in [1/2, 1]$ does not generate a small scale dyadic spanning set for $W^{1,2}(\mathbb{R})$, because if it did then $\psi' = 2H$ would generate a small scale dyadic spanning set for $L^2(\mathbb{R})$ by Proposition 9, whereas spanning $L^2(\mathbb{R})$ requires the full dyadic Haar system (involving $j \in \mathbb{Z}$ and not just $j > 0$), by orthonormality.

We expect such counterexamples to be nongeneric, but they do show that small scales alone will not always suffice to span Sobolev space, or L^p.

The second difference of the Gaussian. Although we cannot so far resolve the Mexican hat spanning problem for the second *derivative* of the Gaussian, we can easily resolve the analogous problem for the second *difference* of the Gaussian. With $\psi(x) = e^{-x^2/2}$ being the Gaussian, write

$$\sigma(x) = \psi(x+1) - 2\psi(x) + \psi(x-1) = -\Delta_{-1}\Delta_1\psi(x)$$

for the symmetric second difference of the Gaussian with step size 1. As remarked above, the Gaussian system $\{\psi_{j,k} : j > 0, k \in \mathbb{Z}\}$ spans $L^p(\mathbb{R})$ for $1 \le p < \infty$, and so the second difference system $\{\sigma_{j,k} : j > 0, k \in \mathbb{Z}\}$ spans $L^p(\mathbb{R})$ for each $1 < p < \infty$, by two applications of Theorem 10 with $m = 0, c = 1$.

Figure 1 shows that the second difference σ of the Gaussian and the second derivative θ (the Mexican hat) behave very much the same way, in both time and frequency domains.

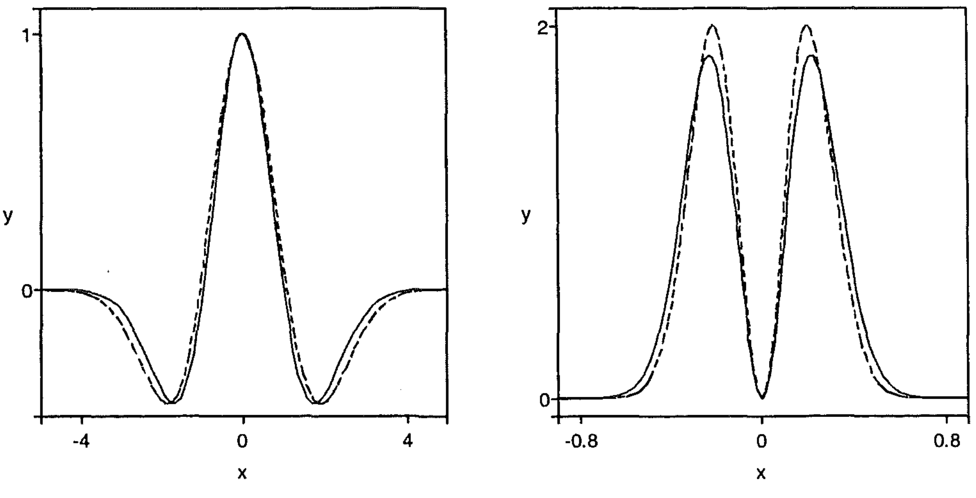

FIGURE 1. Left: The second derivative $\theta(x)$ and second difference $\sigma(x)$ of the Gaussian (solid and dashed curves, respectively), after normalization to 1 at $x = 0$. Right: Their Fourier transforms $\widehat{\theta}(\xi)$ and $\widehat{\sigma}(\xi)$.

Incidentally, the Mexican hat generates more than a spanning set for $L^2(\mathbb{R})$: it generates a dyadic frame by [11, p. 987] or [10, p. 264], meaning constants $0 < A \le B < \infty$ exist such that

$$A\|f\|_2^2 \le \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} |\langle f, \theta_{j,k}\rangle|^2 \le B\|f\|_2^2 \qquad \text{for all } f \in L^2(\mathbb{R}),$$

when $a_j = 2^j$. The second difference function $\sigma(x)$ also generates a dyadic frame: by numerically evaluating Casazza and Christensen's frame criterion (see [9, Theorem 2.5], [10, Theorem 11.2.3]) we have obtained the estimate $B/A \leq 1.088$ for the frame bounds, compared with $B/A \leq 1.095$ for the Mexican hat.

6. Discretized approximations to the identity

The basic approximation results of the paper are developed in this section. The key object is an operator $I_j[\psi, \phi]$ that acts on functions $h(x,y)$ by

$$(I_j[\psi,\phi]h)(x) = |\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} h(x, a_j^{-1}y - x)\phi(y - bk)\, dy \right) \psi(a_j x - bk), \qquad j > 0. \tag{10}$$

Lemma 11 specifies properties of the synthesizer ψ and analyzer ϕ under which I_j is well defined. We will 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.

For example if $h(x,y) = f(x+y)$ and $f \in L^p$ then $h \in L^{(p,\infty)}$ with $\|h\|_{(p,\infty)} = \|f\|_p$. This choice of h yields $I_j[\psi,\phi]h = f_j$, by comparing the definitions (1) and (10). Hence we call I_j a "discretized approximation to the identity" operator.

Lemma 11. *Assume $\psi \in L^p$ for some $1 \leq p < \infty$, and that one of the following conditions holds:*

(i) *$P|\psi| \in L^p_{loc}, \phi \in L^1$, 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) *$Q\psi \in L^1, \phi \in L^q$ with ϕ having compact support, and $h \in L^{(p,\infty)}$.*

Then the series (10) *defining $I_j[\psi,\phi]h$ converges pointwise absolutely a.e. to an L^p function. The series further converges unconditionally in L^p. And in case (ii) we obtain a stability estimate that is independent of j:*

$$\|I_j[\psi,\phi]h\|_p \leq C(p, \operatorname{spt}\phi)\|Q\psi\|_1\|\phi\|_q\|h\|_{(p,\infty)}. \tag{11}$$

Remarks on Lemma 11.

1. Case (i) assumes less about ψ than case (ii) does, but on the other hand it assumes a special form for h, and it does not yield a stability estimate.

2. The assumption $Q\psi \in L^1$ in case (ii) lets us bound the values of ψ at nearby points, so that we can estimate certain Riemann sums involving ψ with integrals involving $Q\psi$. See (17) below.

3. For $h(x,y) = f(x+y)$, Lemma 11 and also Lemma 12 below were proved in our L^p paper [5, Lemmas 1 and 2]. (The hypotheses there are stronger on the analyzer ϕ, but that matters little.) This special choice $h(x,y) = f(x+y)$ yields a stability estimate in both cases (i) and (ii).

See that paper [5] for an account of earlier literature with $h(x,y) = f(x+y)$, such as di Guglielmo [16, p. 288].

Proof of Lemma 11. 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, because in case (i) h is bounded and $\phi \in L^1$, and in case (ii) we see $y \mapsto h(x,y)$ belongs to L^p_{loc} for almost every x and $\phi \in L^q$ has compact support.

To start estimating I_j, notice

$$
\begin{aligned}
|(I_j[\psi,\phi]h)(x)|^p &\leq \left(|\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |h(x, a_j^{-1}y - x)||\phi(y-bk)|\,dy\,|\psi(a_jx - bk)| \right)^p \\
&\leq |\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} |h(x, a_j^{-1}y - x)||\phi(y-bk)|\,dy \right)^p |\psi(a_jx - bk)| \qquad (12) \\
&\qquad \cdot \left(|\det b| \sum_{k\in\mathbb{Z}^d} |\psi(a_jx - bk)| \right)^{p-1}
\end{aligned}
$$

by Hölder's inequality on the sum, when $p > 1$. (When $p = 1$ the last inequality is trivial.)

Case (i). By applying Hölder's inequality to the y-integral in (12) we find

$$
\begin{aligned}
&|(I_j[\psi,\phi]h)(x)|^p \\
&\leq |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |h(x, a_j^{-1}y - x)|^p |\phi(y-bk)|\,dy\,\|\phi\|_1^{p-1}\,|\psi(a_jx - bk)|(P|\psi|(a_jx))^{p-1} \qquad (13)
\end{aligned}
$$

After integrating (13) with respect to x and 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

$$
\|I_j[\psi,\phi]h\|_p^p \leq \int_{[0,1]} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} R_j(x,y,t)\,|\psi(x)|(P|\psi|(x))^{p-1}|\phi(y)|\,dxdyd\omega(t)\,\|\phi\|_1^{p-1} \qquad (14)
$$

where

$$
R_j(x,y,t) = |\det a_j^{-1}b| \sum_{k\in\mathbb{Z}^d} |f(a_j^{-1}(x+bk) + ta_j^{-1}(y-x))|^p. \qquad (15)
$$

We claim R_j is bounded, independently of x, y and t. For if we write $w = a_j^{-1}x + ta_j^{-1}(y-x)$ and $\mathcal{K} = \{k \in \mathbb{Z}^d : a_j^{-1}bk \in (\operatorname{spt} f) - w\}$, then

$$
\begin{aligned}
R_j(x,y,t) &\leq |\det a_j^{-1}b| \cdot \#\mathcal{K} \cdot \|f\|_\infty^p \\
&= \|f\|_\infty^p \cdot |\cup_{k\in\mathcal{K}}\, a_j^{-1}b(k+\mathcal{C})| \\
&\leq \|f\|_\infty^p \cdot |\{z \in \mathbb{R}^d : \operatorname{dist}(z, (\operatorname{spt} f) - w) \leq \operatorname{diam}(a_j^{-1}b\mathcal{C})\}| \\
&= \|f\|_\infty^p \cdot |\{z \in \mathbb{R}^d : \operatorname{dist}(z, \operatorname{spt} f) \leq \operatorname{diam}(a_j^{-1}b\mathcal{C})\}|, \qquad (16)
\end{aligned}
$$

which gives a bound on R_j that is uniform in x, y and t.

Also $\phi \in L^1$ by hypothesis in case (i), and $|\psi|(P|\psi|)^{p-1} \in L^1$ because

$$
\int_{\mathbb{R}^d} |\psi|(P|\psi|)^{p-1}\,dx = \int_{b\mathcal{C}} \sum_{k\in\mathbb{Z}^d} |\psi(x-bk)|(P|\psi|(x-bk))^{p-1}\,dx = |\det b|^{-1}\|P|\psi|\|_{L^p(b\mathcal{C})}^p < \infty.
$$

Therefore $I_j[\psi,\phi]h$ belongs to L^p by the estimate (14).

Case (ii). By using the compact support of ϕ in the y-integral in (12), and then applying Hölder's inequality, we find

$$
\begin{aligned}
&|(I_j[\psi,\phi]h)(x)|^p \\
&\le |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |h(x,a_j^{-1}y-x)|^p \mathbb{1}_{\operatorname{spt}\phi}(y-bk)\,dy\, \|\phi\|_q^p\, |\psi(a_jx-bk)| \|P|\psi|\|_\infty^{p-1} \\
&\le |\det b| \sum_{k\in\mathbb{Z}^d} \int_{\mathbb{R}^d} |h(x,a_j^{-1}y-x)|^p \mathbb{1}_E(y-bk) Q_E\psi(a_jx-y)\,dy \cdot \|P|\psi|\|_\infty^{p-1} \|\phi\|_q^p \qquad (17) \\
&\qquad\qquad \text{for almost every } x, \text{ by Lemma 17 with } f=\psi \text{ and } E=\operatorname{spt}\phi, \\
&\le \int_{\mathbb{R}^d} |h(x,-a_j^{-1}y)|^p Q_E\psi(y)\,dy \cdot \|P\mathbb{1}_E\|_\infty \|P|\psi|\|_\infty^{p-1}\|\phi\|_q^p \qquad \text{by } y\mapsto a_jx-y.
\end{aligned}
$$

Integrating with respect to x gives the norm estimate

$$
\begin{aligned}
\|I_j[\psi,\phi]h\|_p^p &\le \int_{\mathbb{R}^d} \|h(\cdot\,,-a_j^{-1}y)\|_p^p Q_E\psi(y)\,dy \cdot \|P\mathbb{1}_E\|_\infty \|P|\psi|\|_\infty^{p-1}\|\phi\|_q^p \qquad (18) \\
&\le C(E)\|h\|_{(p,\infty)}^p \|Q\psi\|_1 \cdot \|P|\psi|\|_\infty^{p-1}\|\phi\|_q^p,
\end{aligned}
$$

using here that $\|Q_E(\cdot)\|_1 \le C(E)\|Q(\cdot)\|_1$ by definition of Q_E in (64). Finally note $\|P|\psi|\|_\infty \le C\|Q\psi\|_1$ by Lemma 18.

Thus we have proved estimate (11) in case (ii).

Unconditional convergence. The series defining $I_j[\psi,\phi]h$ converges unconditionally in L^p, because

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

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

The next lemma proves convergence properties of $I_j[\psi,\phi]$ as $j\to\infty$.

Lemma 12. *Assume $\psi\in L^p$ for some $1\le p<\infty$, and that one of the following conditions holds:*

(i) *$P|\psi|\in L^p_{loc}, \phi\in L^1$, 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) *$Q\psi\in L^1, \phi\in L^q$ with ϕ having compact support, and $h\in L^{(p,\infty)}$ with*

$$
\lim_{y\to 0} h(\cdot,y)=h(\cdot,0) \qquad \text{in } L^p. \tag{19}
$$

Then (a)–(c) hold:

(a) [Upper bound]

$$
\limsup_{j\to\infty} \|I_j[\psi,\phi]h\|_p \le \|h(\cdot,0)\|_p \cdot \begin{cases} C(p)\|P|\psi|\|_{L^p(b\mathcal{C})}\|\phi\|_1 & \text{in case (i)}, \\ C(p,\operatorname{spt}\phi)\|Q\psi\|_1\|\phi\|_q & \text{in case (ii)}. \end{cases} \tag{20}
$$

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

$$
\lim_{j\to\infty} (I_j[\psi,\phi]h)(x) = h(x,0)\int_{\mathbb{R}^d}\psi(y)\,dy \int_{\mathbb{R}^d}\phi(z)\,dz \qquad \text{in } L^p. \tag{21}
$$

(c) [Scale averaging] If the dilations a_j grow exponentially, then

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

Hypothesis (19) says that $y \mapsto h(\cdot\,, y)$ is continuous as a map $\mathbb{R}^d \to L^p$, at $y = 0$.

Proof of Lemma 12. Observe $P|\psi| \in L^p_{loc}$, by hypothesis in case (i) and from Lemma 18 in case (ii). Integrating $P|\psi|$ over the period cell $b\mathcal{C}$ then shows $\psi \in L^1$. And the mean value of $P\psi$ equals

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

Thus 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}. If the dilations a_j grow exponentially then [5, Lemma 3] tells us that $\lim_{J\to\infty}\frac{1}{J}\sum_{j=1}^J g(a_jx) = 0$ in L^p_{loc}, or

$$\lim_{J\to\infty}\frac{1}{J}\sum_{j=1}^J P\psi(a_jx) = \int_{\mathbb{R}^d}\psi(y)\,dy \qquad \text{in } L^p_{loc}. \tag{22}$$

(Formula (22) is the source of all scale averaging in this paper. It is a concrete version of Mazur's theorem, which says that the weak convergence $g(a_jx) \rightharpoonup 0$ implies norm convergence of suitable convex combinations of the $g(a_jx)$.)

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

Part (a). Case (i). The estimate (16) implies that R_j is bounded by a constant independent of x, y, t and j, for all large j (using that $a_j^{-1} \to 0$). Note also $R_j(x,y,t) \to \int_{\mathbb{R}^d}|f(z)|^p\,dz$ as $j \to \infty$, for each x, y, t (by interpreting the definition of R_j in (15) as a Riemann sum and using that $f \in C_c$). Thus we may apply dominated convergence to formula (14) to obtain that

$$\limsup_{j\to\infty}\|I_j[\psi,\phi]h\|_p^p \le \int_{[0,1]}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|f(z)|^p\,dz\,|\psi(x)|(P|\psi|(x))^{p-1}|\phi(y)|\,dxdyd\omega(t)\|\phi\|_1^{p-1},$$

which implies estimate (20).

Case (ii). By dominated convergence, as $j \to \infty$ the righthand side of (18) approaches the limiting value $\int_{\mathbb{R}^d}\|h(\cdot\,,0)\|_p^p Q_E\psi(y)\,dy \cdot \|P\mathbb{1}_E\|_\infty\|P|\psi|\|_\infty^{p-1}\|\phi\|_q^p$, because $Q_E\psi \in L^1$ and $\|h(\cdot,y)\|_p \in L^\infty$ while $h(\cdot,y) \to h(\cdot,0)$ in L^p as $y \to 0$ by assumption (19). This proves (20) in case (ii), since we can now replace Q_E with Q like we did after (18).

Before considering parts (b) and (c) of the lemma, we prove (20) for a useful variant of h from case (i).

Lemma 13. *Assume $\psi \in L^p$ for some $1 \le p < \infty$, and that $P|\psi| \in L^p_{loc}, \phi \in L^1$, 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 $h^*(x,0) \equiv 0$, and $\lim_{j\to\infty}\|I_j[\psi,\phi]h^*\|_p = 0$.

Proof of Lemma 13. We have

$$\|I_j[\psi,\phi]h^*\|_p^p \le \int_{[0,1]}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}R_j^*(x,y,t)\,|\psi(x)|(P|\psi|(x))^{p-1}|\phi(y)|\,dxdyd\omega(t)\,\|\phi\|_1^{p-1} \tag{23}$$

by applying (14) 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 type sum, converging pointwise to $\int_{\mathbb{R}^d}|f(z) - f(z)|^p\,dz = 0$ as $j \to \infty$, since f is continuous with compact support. And like in the proof of Lemma 12(a) in case (i), one finds $R_j^*(x,y,t)$ is bounded by a constant independent of x, y, t and j, for all large j. Thus dominated convergence applied to (23) gives $\|I_j[\psi,\phi]h^*\|_p \to 0$ as $j \to \infty$. □

Now we return to proving Lemma 12.

Parts (b) and (c). Define

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

Then the definition of I_j in (10) implies

$$(I_j[\psi,\phi]h)(x) = (I_j[\psi,\phi]H)(x) + h(x,0)P\psi(a_j x)\int_{\mathbb{R}^d}\phi(z)\,dz. \tag{24}$$

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

$$\lim_{j\to\infty} I_j[\psi,\phi]H = 0 \qquad \text{in } L^p \tag{25}$$

by Lemma 13, because $|I_j[\psi,\phi]H| \le I_j[|\psi|,|\phi|]h^*$ pointwise.

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 (21) follows immediately from (25) and decomposition (24).

For part (c) we just use (25) and (24) and observe that

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

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 (22).

Case (ii). In this case $\lim_{j\to\infty} I_j[\psi,\phi]H = 0$ 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 (19).

Hence part (b) of the lemma again follows from the decomposition (24).

Part (c) follows like in the proof of part (i) above when $h(x,0)$ is bounded with compact support. But we can reduce part (c) to this situation by the stability estimate $\|I_j[\psi,\phi]h\|_p \le C\|h\|_{(p,\infty)}$ (proved in Lemma 11, formula (11)) in conjunction with the following density argument. Given $\epsilon > 0$, choose $\tilde{h} \in C_c(\mathbb{R}^d)$ with $\|h(\cdot\,,0) - \tilde{h}\|_p < \epsilon$, and then define

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

Then trivially $\lim_{y\to 0} h_\epsilon(\cdot\,,y) = h_\epsilon(\cdot,0)$ in L^p, while

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

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 when $y = 0$. □

7. A preliminary result: Strang–Fix implies constant periodization

Here we establish a lemma explaining Theorem 1's hypotheses on the zeros of $\widehat{\psi}$. Roughly, if the Fourier transform of ψ vanishes at every nonzero lattice point, and so do its derivatives up to order n^*, then the moments of ψ up to order n^* must all have constant periodization.

Recall $X(x) = x$ is the identity function, and $\chi(x) = 1 + |x|$.

Lemma 14. *Take integers $0 \le m^* \le m, 0 \le n^* \le n$, and suppose $\psi \in W^{m,1}$ with $\chi^n\psi^{(\rho)} \in L^1$ for all multiindices of order $|\rho| = m^*$.*

(a) Then $\widehat{\psi} \in C^n(\mathbb{R}^d \setminus \{0\})$, and $\widehat{\psi} \in C^n(\mathbb{R}^d)$ if $\chi^n\psi \in L^1$.

(b) If $D^\sigma \widehat{\psi}(\ell b^{-1}) = 0$ for all $|\sigma| \leq n^*$ and all row vectors $\ell \in \mathbb{Z}^d \setminus \{0\}$, then the periodization of $(-X)^\sigma \psi^{(\rho)}$ is constant for each $|\sigma| \leq n^*, |\rho| = m^*$, with

$$P((-X)^\sigma \psi^{(\rho)})(x) = \begin{cases} \frac{\sigma!}{(\sigma-\rho)!} \int_{\mathbb{R}^d} (-y)^{\sigma-\rho} \psi(y)\, dy & \text{if } \sigma \geq \rho, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for almost every } x.$$

Proof of Lemma 14. Part (a). Suppose $|\sigma| \leq n$ and $|\rho| = m^*$. Then $(-2\pi i X)^\sigma \psi^{(\rho)}$ 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 σ times, obtaining that $\widehat{\psi^{(\rho)}} \in C^n(\mathbb{R}^d)$ since σ was arbitrary. But $\widehat{\psi^{(\rho)}}(\xi) = (2\pi i \xi)^\rho \widehat{\psi}(\xi)$, and so $\widehat{\psi}$ has n continuous derivatives away from the set $\{\xi : \xi^\rho = 0\}$. By considering all pure multiindices (meaning $\rho = (m^*, 0, \dots, 0)$ and so on) we deduce that $\widehat{\psi}$ has n continuous derivatives away from the origin.

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

$$\begin{aligned} &\int_C P((2\pi i(-X))^\sigma \psi^{(\rho)})(bx) e^{-2\pi i \ell x}\, dx \\ &= \int_{bC} \sum_{k \in \mathbb{Z}^d} (2\pi i(-x+bk))^\sigma \psi^{(\rho)}(x-bk) e^{-2\pi i \ell b^{-1} x}\, dx && \text{by } x \mapsto b^{-1}x \text{ and definition of } P \\ &= \int_{\mathbb{R}^d} (-2\pi i x)^\sigma \psi^{(\rho)}(x) e^{-2\pi i \ell b^{-1} x}\, dx && \text{by } x \mapsto x + bk \qquad (26) \\ &= D_\xi^\sigma \int_{\mathbb{R}^d} \psi^{(\rho)}(x) e^{-2\pi i \xi x}\, dx \Big|_{\xi = \ell b^{-1}} \\ &= D_\xi^\sigma (2\pi i \xi)^\rho \widehat{\psi}(\xi) \Big|_{\xi = \ell b^{-1}} \end{aligned}$$

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

This constant value is given by the $\ell = 0$ Fourier coefficient, which by (26) equals

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

after integrating by parts ρ times. □

8. Proof of Theorem 1

First we show the hypotheses of the theorem make sense.

To start with we show $\widehat{\psi} \in C^m(\mathbb{R}^d \setminus \{0\})$, so that $D^\mu \widehat{\psi}(\ell b^{-1})$ makes sense whenever $|\mu| \leq m$ and $\ell \neq 0$. So let μ be a multiindex of order $|\mu| \leq m$. We have $P|\chi^{|\mu|} \psi^{(\mu)}| \in L^p_{loc} \subset L^1_{loc}$ (from hypothesis in case (i), and in case (ii) by using also Lemma 18). Hence $\chi^{|\mu|} \psi^{(\mu)} \in L^1$, so that $\psi \in W^{m,1}$. Lemma 14(a) with $n^* = n = m^* = m$ now tells us that $\widehat{\psi} \in C^m(\mathbb{R}^d \setminus \{0\})$.

Note also that $\chi^{|\mu|} \psi^{(\mu)} \in L^p$ by Lemma 16.

Next, $\psi \in L^1$ by above, while $\phi \in L^1$ from the hypotheses in cases (i) and (ii). Thus the normalizations on the integrals of ψ and ϕ (in the statement of Theorem 1) do make sense.

Now we commence the proof, by showing $f_j \in W^{m,p}$. Fix a multiindex ρ of order

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

If we formally take the derivative through the sum over k in the definition of f_j, in formula (1), we find that

$$D^\rho f_j(x) = |\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} f(a_j^{-1}y)\phi(y-bk)\,dy \right) a_j^\tau \psi^{(\rho)}(a_j x - bk). \tag{27}$$

To make this rigorous, let $h(x,y) = f(x+y)$ and notice the righthand side of equation (27) equals $a_j^\tau I_j[\psi^{(\rho)}, \phi]h$, which belongs to L^p by Lemma 11. That lemma proves the sum over k in (27) converges pointwise absolutely a.e. to an L^p function. Then it is straightforward to show $D^\rho f_j$ exists weakly and is given by (27). Hence $f_j \in W^{m,p}$.

Part (d). In fact $f_j \in W^{m,p}\text{-span}\{\psi_{j,k} : k \in \mathbb{Z}^d\}$, because the sum over k in (27) converges unconditionally in L^p by Lemma 11.

Parts (a)(b)(c). Our first step is to add and subtract an appropriate Taylor polynomial inside the formula (27) for $D^\rho f_j$. Specifically, we will show

$$D^\rho f_j = \mathrm{Main}_j + \mathrm{Rem}_j$$

where

$$\mathrm{Main}_j(x) = \sum_{|\sigma|\le r} \frac{f^{(\sigma)}(x)}{\sigma!} a_j^{r-s} \sum_{\tau\le\sigma} \binom{\sigma}{\tau} \int_{\mathbb{R}^d} y^{\sigma-\tau}\phi(y)\,dy \cdot P((-X)^\tau \psi^{(\rho)})(a_j x), \tag{28}$$

$$\mathrm{Rem}_j(x) = |\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} \left[f(a_j^{-1}y) - \sum_{|\sigma|\le r} \frac{f^{(\sigma)}(x)}{\sigma!}(a_j^{-1}y - x)^\sigma \right] \phi(y-bk)\,dy \right) a_j^\tau \psi^{(\rho)}(a_j x - bk), \tag{29}$$

with $s = |\sigma|$ and with $\binom{\sigma}{\tau} = \binom{\sigma_1}{\tau_1}\cdots\binom{\sigma_d}{\tau_d}$ being a product of binomial coefficients. To see this, substitute the binomial identity

$$(a_j^{-1}y - x)^\sigma = a_j^{-s} \sum_{\tau\le\sigma} \binom{\sigma}{\tau} (y-bk)^{\sigma-\tau}(bk - a_j x)^\tau$$

into $\mathrm{Rem}_j(x)$, which leads to cancellation with all the terms in $\mathrm{Main}_j(x)$ and thereby reduces us back to the known formula (27) for $D^\rho f_j$.

Remainder term. We will show $\mathrm{Rem}_j \to 0$ in L^p as $j \to \infty$. In fact we take absolute values in (29) and aim to show

$$\lim_{j\to\infty} |\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} h_r(x, a_j^{-1}y - x)|y - a_j x|^r |\phi(y-bk)|\,dy \right) |\psi^{(\rho)}(a_j x - bk)| = 0 \tag{30}$$

in L^p, where

$$h_r(x,y) = \begin{cases} \left| f(x+y) - \sum_{|\sigma|\le r} \frac{f^{(\sigma)}(x)}{\sigma!} y^\sigma \right| \Big/ |y|^r & \text{when } y \ne 0, \\ 0 & \text{when } y = 0. \end{cases} \tag{31}$$

Taylor's formula with integral remainder enables us to rewrite

$$h_r(x,y) = \left| \int_{[0,1]} \sum_{|\sigma|=r} \frac{1}{\sigma!} [f^{(\sigma)}(x+ty) - f^{(\sigma)}(x)] y^\sigma \, d\omega_r(t) \right| \Big/ |y|^r$$

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

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

Hence

$$h_r(x,y) \le H_r(x,y) := \int_{[0,1]} \sum_{|\sigma|=r} |f^{(\sigma)}(x+ty) - f^{(\sigma)}(x)|\, d\omega_r(t). \tag{32}$$

After putting $h_r \le H_r$ and the estimate

$$|y - a_j x| \le |y - bk| + |bk - a_j x| \le (1 + |y - bk|)(1 + |bk - a_j x|) \tag{33}$$

into (30), we see it's enough to prove

$$\lim_{j\to\infty} I_j[|\chi^r \psi^{(\rho)}|, \phi_r] H_r = 0 \qquad \text{in } L^p \tag{34}$$

where $\phi_r = |\chi^r \phi|$.

Our hypotheses on ϕ guarantee in case (i) that $\phi_r \in L^1$, and in case (ii) that $\phi_r \in L^q$ with compact support. Hence in case (i), the desired limit (34) follows from Lemma 13, because H_r has the form required of h^* in that lemma and $f^{(\sigma)} \in C_c$.

In case (ii), we see that (34) follows from Lemma 12(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 (32) implies

$$\begin{aligned} \|H_r(\cdot,y)\|_p &\le \sum_{|\sigma|=r} \int_{[0,1]} \|f^{(\sigma)}(\cdot + ty) - f^{(\sigma)}\|_p \, d\omega_r(t) \qquad \left(\text{which is } \le 2\sum_{|\sigma|=r} \|f^{(\sigma)}\|_p\right) \\ &\to 0 \qquad \text{as } y \to 0. \end{aligned} \tag{35}$$

This completes our proof that the remainder term Rem_j vanishes in L^p in the limit as $j \to \infty$.

Main term. Next we examine $\mathrm{Main}_j(x)$. Since $|\tau| \le |\sigma| \le r = |\rho|$, if either $\tau < \sigma$ or $|\sigma| < r$ then $0 \le |\tau| < |\rho| = r$, and so

$$P((-X)^\tau \psi^{(\rho)}) = 0 \quad \text{a.e.} \tag{36}$$

by Lemma 14(b) (with $m^* = n = r$ and $n^* = r-1$). It is here in Lemma 14 that we employ the Strang–Fix hypothesis (2) on the zeros of $\widehat{\psi}$.

Most terms in $\mathrm{Main}_j(x)$ vanish by (36). The ones that are left have $|\sigma| = r$ and $\tau = \sigma$, so that $s = r$ and $\binom{\sigma}{\tau} = 1$ and $\int_{\mathbb{R}^d} y^{\sigma-\tau}\phi(y)\,dy = \int_{\mathbb{R}^d} \phi\, dy = 1$. Thus

$$\mathrm{Main}_j(x) = \sum_{|\sigma|=r} \frac{f^{(\sigma)}(x)}{\sigma!} P((-X)^\sigma \psi^{(\rho)})(a_j x). \tag{37}$$

Proof of Part (c). Assume (ii) holds, in this part of the proof. Let $0 \le |\rho| = r \le m$. Then

$$\begin{aligned} \|\mathrm{Rem}_j\|_p &\le \|I_j[|\chi^r \psi^{(\rho)}|, \phi_r] H_r\|_p && \text{by the estimates leading up to (34)} \\ &\le C(\psi, \phi, p) \sum_{|\sigma|=r} \|f^{(\sigma)}\|_p && \text{by Lemma 11 and (35).} \end{aligned}$$

And

$$\begin{aligned} \|\mathrm{Main}_j\|_p &\le \sum_{|\sigma|=r} \|f^{(\sigma)}\|_p \|P((-X)^\sigma \psi^{(\rho)})\|_\infty && \text{by (37)} \\ &\le C \sum_{|\sigma|=r} \|f^{(\sigma)}\|_p \|Q((-X)^\sigma \psi^{(\rho)})\|_1 && \text{by Lemma 18.} \end{aligned}$$

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

Proof of Parts (a) and (b). We need only consider $f \in C_c^m$ when proving part (b): in case (i) we already assume $f \in C_c^m$, and in case (ii) we can reduce to $f \in C_c^m$ by the density of such functions in $W^{m,p}$ and the stability bound $\|f_j\|_{W^{m,p}} \leq C\|f\|_{W^{m,p}}$ proved in part (c).

To prove parts (a) and (b), we will first show

$$D^\rho f = \lim_{j\to\infty} D^\rho f_j \qquad \text{in } L^p \quad \text{if } |\rho| < m \text{ and } f \in W^{m,p}. \tag{38}$$

Then to complete the approximation formula in (a) we will show that if the Strang–Fix hypothesis (2) holds for all multiindices of order $\leq m$ (not just $< m$), then

$$D^\rho f = \lim_{j\to\infty} D^\rho f_j \qquad \text{in } L^p \quad \text{if } |\rho| = m \text{ and } f \in W^{m,p}. \tag{39}$$

To complete the approximation formula in (b) we will show (if the dilations a_j grow exponentially) that

$$D^\rho f = \lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} D^\rho f_j \qquad \text{in } L^p \quad \text{if } |\rho| = m \text{ and } f \in C_c^m. \tag{40}$$

Proof of limits (38) *and* (39). For proving the first limit (38) we suppose $|\sigma| = |\rho| = r < m$. Then

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

for almost every x, by Lemma 14(b) with $m^* = n = n^* = r$ (and recalling $\int_{\mathbb{R}^d} \psi\, dy = 1$). Hence (37) simplifies to $\mathrm{Main}_j = D^\rho f$, meaning (38) follows immediately from our remainder estimate $\mathrm{Rem}_j \to 0$.

To prove the next limit (39), just apply the same reasoning with $r = m$.

Proof of limit (40). To prove the third limit (40), suppose $|\sigma| = |\rho| = r = m$. Define the function

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

so that $g_{\sigma;\rho} \in L^p_{loc}$ is $b\mathbb{Z}^d$-periodic. Then

$$\mathrm{Main}_j(x) = D^\rho f(x) + \sum_{|\sigma|=m} f^{(\sigma)}(x) g_{\sigma;\rho}(a_j x), \tag{41}$$

by comparing with the expression (37) for $\mathrm{Main}_j(x)$. Each function $g_{\sigma;\rho}$ has mean value zero, because

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

by parts, recalling $|\sigma| = |\rho|$.

If the dilations a_j grow exponentially, as assumed for (40), then [5, Lemma 3] applies to each $g_{\sigma;\rho}$ and says that

$$\lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} g_{\sigma;\rho}(a_j x) = 0 \qquad \text{in } L^p_{loc}.$$

Then

$$\lim_{J\to\infty} f^{(\sigma)}(x) \frac{1}{J}\sum_{j=1}^{J} g_{\sigma;\rho}(a_j x) = 0 \qquad \text{in } L^p,$$

because each $f^{(\sigma)}$ is bounded and has compact support (recalling $f \in C_c^m$ for (40)). Hence

$$\lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} \mathrm{Main}_j = D^\rho f \qquad \text{in } L^p \tag{42}$$

by (41). Combining (42) with $\mathrm{Rem}_j \to 0$, we deduce the limit (40).

9. Proof of Theorem 4

Our initial task is to show $f_j^\bullet \in W^{m,p}$. Fix a multiindex ρ with $r := |\rho| \le m$. Like in Theorem 1, formally differentiating the definition (7) of $f_j^\bullet$ yields that

$$D^\rho f_j^\bullet(x) = |\det b| \sum_{k\in\mathbb{Z}^d} f(a_j^{-1}bk)a_j^r \psi^{(\rho)}(a_j x - bk). \tag{43}$$

The righthand side of this equation is exactly $a_j^r f_j^\bullet[\psi^{(\rho)}]$, where the temporary notation $f_j^\bullet[\psi^{(\rho)}]$ denotes the function obtained by replacing ψ with $\psi^{(\rho)}$ in the definition of $f_j^\bullet$. Now to show rigorously that $f_j^\bullet$ is weakly differentiable with derivative given by (43), it is enough (like in the proof of Theorem 1) to observe that the series defining $f_j^\bullet[\psi^{(\rho)}]$ converges absolutely a.e. to an L^p function, which it does by [5, Theorem 2(e) and its Remark 3]. Note $\psi^{(\rho)}$ does satisfy the hypotheses of [5, Theorem 2(e)], by using Lemma 18.

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

Part (d). In fact $f_j^\bullet \in W^{m,p}\text{-span}\{\psi_{j,k} : k \in \mathbb{Z}^d\}$, because the sum over k in (43) converges unconditionally in L^p, by [5, Theorem 2(e)] applied to $\psi^{(\rho)}$.

Parts (a)(b). We will first show

$$D^\rho f = \lim_{j\to\infty} D^\rho f_j^\bullet \qquad \text{in } L^p \quad \text{if } |\rho| < m. \tag{44}$$

Then to complete the approximation formula in (a) we will show that if hypothesis (8) holds for all multiindices of order $\le m$ (not just $< m$), then

$$D^\rho f = \lim_{j\to\infty} D^\rho f_j^\bullet \qquad \text{in } L^p \quad \text{if } |\rho| = m. \tag{45}$$

And to complete the proof of part (b) we will show (if the dilations a_j grow exponentially) that

$$D^\rho f = \lim_{J\to\infty} \frac{1}{J}\sum_{j=1}^{J} D^\rho f_j^\bullet \qquad \text{in } L^p \quad \text{if } |\rho| = m. \tag{46}$$

To begin with, we calculate from (43) that

$$D^\rho f_j^\bullet = \mathrm{Main}_j + \mathrm{Rem}_j^\bullet$$

where

$$\mathrm{Main}_j(x) = \sum_{|\sigma|=r} \frac{f^{(\sigma)}(x)}{\sigma!} P((-X)^\sigma \psi^{(\rho)})(a_j x),$$

$$\mathrm{Rem}_j^\bullet(x) = |\det b| \sum_{k\in\mathbb{Z}^d} \left[f(a_j^{-1}bk) - \sum_{|\sigma|\le r} \frac{f^{(\sigma)}(x)}{\sigma!}(a_j^{-1}bk - x)^\sigma \right] a_j^r \psi^{(\rho)}(a_j x - bk),$$

noting in this calculation that if $|\sigma| < r$ then $P((-X)^\sigma\psi^{(\rho)}) \equiv 0$ by Lemma 14(b) with $m^* = n = r$ and $n^* = r - 1$.

In formulas (44) and (45) we have $\mathrm{Main}_j = D^\rho f$, as shown in the proof of (38) and (39) in the previous section. Thus for proving (44) and (45), we have only to show $\mathrm{Rem}_j^\bullet \to 0$ in L^p.

In formula (46) we can express Main_j as in (41), and

$$f^{(\sigma)}(x)\frac{1}{J}\sum_{j=1}^{J} g_{\sigma;\rho}(a_j x) \to 0 \qquad \text{in } L^p \text{ as } J \to \infty$$

by Lemma 19, since $Q(f^{(\sigma)}) \in L^p$. Therefore (42) holds, and so to prove (46) we again have only to show $\mathrm{Rem}_j^\bullet \to 0$ in L^p.

Remainder term. We will show $\mathrm{Rem}_j^\bullet \to 0$ in L^p as $j \to \infty$. After taking absolute values inside $\mathrm{Rem}_j^\bullet(x)$, we would like to show

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

in L^p, where the function h_r was defined in (31). Since $h_r \le H_r$ by the Taylor remainder estimate (32), it suffices to prove

$$\lim_{j\to\infty} |\det b| \sum_{k\in\mathbb{Z}^d} H_r(x, a_j^{-1}bk - x)\, |(\chi^r \psi^{(\rho)})(a_j x - bk)| = 0 \qquad \text{in } L^p. \tag{47}$$

We will do this by comparing with the analogous limit that uses average rather than pointwise sampling.

So let our analyzer be $\phi = \mathbb{1}_{b\mathcal{C}}/|b\mathcal{C}|$ and subtract the quantity $I_j[|\chi^r\psi^{(\rho)}|, \phi]H_r$ from (47). This quantity tends to zero in L^p as $j \to \infty$ by Lemma 12(a), observing $H_r(\cdot, y) \to H_r(\cdot, 0) = 0$ in L^p as $y \to 0$.

After performing the subtraction of $I_j[|\chi^r\psi^{(\rho)}|, \phi]H_r$ from (47) and then taking absolute values, we see it would be enough to prove (whenever $|\sigma| = |\rho| = r \le m$) that

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

But $\phi(y - bk) \ne 0$ if and only if $y - bk \in b\mathcal{C}$, in which case $|a_j^{-1}bk - a_j^{-1}y| \le \|a_j^{-1}b\|\sqrt{d}$. Therefore the last limit would follow from

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

in L^p, where the modulus of continuity operator S is defined in Appendix A. Thus our goal is now to prove

$$\lim_{j\to\infty} I_j[|\chi^r\psi^{(\rho)}|, \phi]T_j = 0 \qquad \text{in } L^p \tag{48}$$

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

The stability estimate in Lemma 11 together with Minkowski's integral inequality implies that

$$\|I_j[|\chi^r\psi^{(\rho)}|, \phi]T_j\|_p \le C(\psi, p, r)\|T_j\|_{(p,\infty)} \le C(\psi, p, r)\|S_{a_j^{-1}b} f^{(\sigma)}\|_p \tag{49}$$
$$\to 0 \qquad \text{as } j \to \infty$$

by Lemma 20, which is valid since $Qf^{(\sigma)} \in L^p$ and $f^{(\sigma)} \in C$ by hypothesis. This proves (48), completing our proof that $\mathrm{Rem}_j^\bullet \to 0$ in L^p.

Part (c). The proof of Theorem 1(c) shows $\|\,\mathrm{Main}_j\,\|_p \le C(\psi, m, p)\|f\|_{W^{m,p}}$.

To get stability of the remainder term Rem_j, it suffices to show (in view of our proof above) that

$$\left\| I_j[|\chi^r \psi^{(\rho)}|, \phi] H_r \right\|_p \le C(\psi, r, p) \sum_{|\sigma|=r} \|f^{(\sigma)}\|_p, \tag{50}$$

$$\left\| I_j[|\chi^r \psi^{(\rho)}|, \phi] T_j \right\|_p \le C(\psi, r, p, a_{min}) \sum_{|\sigma|=r} \|Q(f^{(\sigma)})\|_p. \tag{51}$$

The first inequality follows from Lemma 11 together with the estimate $\|H_r\|_{(p,\infty)} \le \sum_{|\sigma|=r} 2\|f^{(\sigma)}\|_p$ in (35), and the second inequality follows from (49) and the fact that $\|S_{a_j^{-1}b} f^{(\sigma)}\|_p \le C(a_{min})\|Qf^{(\sigma)}\|_p$ for all $j > 0$ by Lemma 20.

10. Proof of Lemma 5. Examples.

Notice $\widehat{\phi} \in C^m$ and $\widehat{\psi} \in C^{m-1}$, since $\chi^m \phi \in L^1$ and $\chi^{m-1}\psi \in L^1$.

We adapt the reasoning in [30, p. 833] as follows. Let $K = \{k \in \mathbb{Z}^d : |k_1| + \cdots + |k_d| \le m\}$ and write $B(\xi) = \sum_{k\in K} \beta_k e^{2\pi i \xi k}$ for a trigonometric polynomial with coefficients β_k to be determined later. After checking that

$$\int_{\mathbb{R}^d} (-x)^\mu \Psi(x)\, dx = (2\pi i)^{-|\mu|} D^\mu \Big(B(\xi b)\widehat{\psi}(\xi) \Big)\Big|_{\xi=0}, \qquad |\mu| \le m-1,$$

we see the task for Ψ in (9) is to choose B 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

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

where $\theta \in \mathbb{R}^d$ is regarded as a row vector and $p_0(\xi) \equiv 1$ and where for $0 < |\mu| \le m-1$ 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

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

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

Argue similarly to construct Φ.

Examples for Lemma 5. In special cases we can argue directly to construct Φ and Ψ, rather than following the method of the proof above. Take $b = I$ for simplicity, and $\int_{\mathbb{R}^d} \phi\, dx = \int_{\mathbb{R}^d} \psi\, dx = 1$.

1. Let $m = 1$. If $\phi(x)$ is even with respect to each component x_i, then we can take $\Phi = \phi$ and $\Psi = \psi$, in Lemma 5.

2. Let $m = 2$. If $\phi(x)$ and $\psi(x)$ are even with respect to each component x_i, and $\phi(x)$ is symmetric in $x_1, \ldots, x_d$, then in Lemma 5 we can take $\Psi = \psi$ and

$$\Phi(x) = \alpha_0 \phi(x) - \alpha_1 \sum_{k:|k|_\infty = 1} \phi(x-k),$$

where $|k|_\infty := \max_{1\le i\le d} |x_i|$ and

$$\alpha_0 = 1 + (3^d - 1)\alpha_1, \qquad \alpha_1 = \frac{\int_{\mathbb{R}^d} x_1^2 \phi(x)\, dx}{2 \cdot 3^{d-1}}.$$

We leave the reader to verify that (9) holds with $m = 2$. Notice $3^d - 1 = \#\{k : |k|_\infty = 1\}$.

In dimension $d = 1$ this construction reduces to $\Phi(x) = (1 + 2\alpha_1)\phi(x) - \alpha_1\phi(x-1) - \alpha_1\phi(x+1)$ with $\alpha_1 = \frac{1}{2}\int_{\mathbb{R}} x^2 \phi(x)\, dx$, provided ϕ is an even function of one variable.

11. Proof of Theorem 6

Fix a multiindex ρ with $r := |\rho| \le m-1$.

Part (a). We decompose

$$D^\rho F_j = \operatorname{Main}_j + \operatorname{Rem}_j$$

where

$$\operatorname{Main}_j(x) = \sum_{|\sigma|\le m} \frac{f^{(\sigma)}(x)}{\sigma!} a_j^{r-s} \cdot P((-X)^\sigma \Psi^{(\rho)})(a_j x),$$

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

and $s = |\sigma|$. These quantities are identical to Main_j and Rem_j in (28) and (29) (see the proof of Theorem 1) except that here we sum over $|\sigma| \le m$ instead of $|\sigma| \le r$ and we use the moment conditions (9) on Φ to evaluate the moments $\int_{\mathbb{R}^d} y^{\sigma-\tau}\Phi(y)\,dy$. Note that the periodization $P((-X)^\sigma\Psi^{(\rho)})$ occurring in Main_j is bounded, by the hypothesis that $Q(\chi^m\Psi^{(\rho)}) \in L^1$ and Lemma 18.

Remainder term. We first show

$$\|\operatorname{Rem}_j\|_p \le C(\psi,\phi,m,p)|f|_{W^{m,p}}|a_j|^{r-m} \qquad \text{for all } j>0. \tag{52}$$

Now, Rem_j is bounded pointwise by

$$|\det b| \sum_{k\in\mathbb{Z}^d} \left(\int_{\mathbb{R}^d} h_m(x, a_j^{-1}y - x)|y - a_j x|^m |\Phi(y-bk)|\, dy \right) |\Psi^{(\rho)}(a_j x - bk)| \cdot |a_j|^{r-m} \tag{53}$$

where h_m is defined by taking "$r = m$" in (31). And after using (33) to estimate $|y - a_j x|$, we see that (53) is bounded by $|a_j|^{r-m}$ times $I_j[|\chi^m\Psi^{(\rho)}|, \Phi_m]h_m$ where $\Phi_m = |\chi^m\Phi|$.

Hence (52) follows from

$$\|I_j[|\chi^m\Psi^{(\rho)}|, \Phi_m]h_m\|_p \le C(\psi,\phi,m,p)|f|_{W^{m,p}} \qquad \text{for all } j>0,$$

which holds by the stability estimate in Lemma 11 in view of the following observations. First, $Q(\chi^m\psi^{(\rho)}) \in L^1$ by hypothesis, which implies $\chi^m\Psi^{(\rho)} \in L^1 \cap L^\infty \subset L^p$ by Lemma 18 with "$r = 1$". Second, $\phi \in L^q$ has compact support and so Φ does too, so that $\Phi_m \in L^q$ with compact support. Third, $h_m \in L^{(p,\infty)}$ with $\|h_m\|_{(p,\infty)} \le C(m,p)|f|_{W^{m,p}}$ by (32) and (35) (with r changed to m).

Main term. Next we simplify Main_j. Notice that $D^\rho\widehat{\Psi}(\ell b^{-1}) = 0$ for all row vectors $\ell \in \mathbb{Z}^d \setminus \{0\}$, because the same is assumed for ψ, in this theorem (recalling $|\rho| = r \le m-1$). Hence

$$P((-X)^\sigma\Psi^{(\rho)}) = \begin{cases} \rho! & \text{if } \sigma = \rho \\ 0 & \text{otherwise} \end{cases} \qquad \text{whenever } |\sigma| < m,$$

by Lemma 14(b) applied to Ψ (with $n^* = n = m-1$ and $m^* = r$ and with "m" in the lemma replaced by $m-1$), using also here the moment condition (9) on Ψ.

Thus the only terms in Main_j that can make a nonzero contribution are those either with $|\sigma| = m$ or else with $|\sigma| < m$ and $\sigma = \rho$. Hence

$$\operatorname{Main}_j(x) = f^{(\rho)}(x) + \sum_{|\sigma|=m} \frac{f^{(\sigma)}(x)}{\sigma!} a_j^{r-m} \cdot P((-X)^\sigma\Psi^{(\rho)})(a_j x),$$

so that

$$\begin{aligned}\| \operatorname{Main}_j - D^\rho f\|_p &\le \sum_{|\sigma|=m} \|f^{(\sigma)}\|_p \|P((-X)^\sigma \Psi^{(\rho)})\|_\infty |a_j|^{r-m} \\ &\le C(\psi, m, p)|f|_{W^{m,p}}|a_j|^{r-m} \end{aligned} \tag{54}$$

By putting together (52) and (54) we get

$$\|D^\rho F_j - D^\rho f\|_p \le C(\psi, \phi, m, p)|f|_{W^{m,p}}|a_j|^{r-m},$$

which proves part (a) of the theorem.

Part (b). Similar to part (a) we decompose

$$D^\rho F_j^\bullet = \operatorname{Main}_j + \operatorname{Rem}_j^\bullet$$

where

$$\operatorname{Rem}_j^\bullet(x) = |\det b| \sum_{k\in\mathbb{Z}^d} \left[f(a_j^{-1}bk) - \sum_{|\sigma|\le m} \frac{f^{(\sigma)}(x)}{\sigma!}(a_j^{-1}bk - x)^\sigma \right] a_j^r \Psi^{(\rho)}(a_j x - bk).$$

The term Main_j was estimated already in part (a), leading to (54). Hence to prove part (b) it suffices to show the remainder estimate

$$\| \operatorname{Rem}_j^\bullet \|_p \le C(\psi, m, p, a_{min}) \sum_{|\sigma|=m} \|Qf^{(\sigma)}\|_p |a_j|^{r-m} \qquad \text{for all } j > 0.$$

Notice $\operatorname{Rem}_j^\bullet$ is bounded pointwise by

$$|\det b| \sum_{k\in\mathbb{Z}^d} h_m(x, a_j^{-1}bk - x)|bk - a_j x|^m |\Psi^{(\rho)}(a_j x - bk)| \cdot |a_j|^{r-m}.$$

After using $h_m \le H_m$ like in (32), where $H_m(x,y) = \sum_{|\sigma|=m} \int_{[0,1]} |f^{(\sigma)}(x+ty) - f^{(\sigma)}(x)|\, d\omega_m(t)$, we reduce the remainder estimate to showing

$$\left\| |\det b| \sum_{k\in\mathbb{Z}^d} H_m(x, a_j^{-1}bk - x)\, |(\chi^m \Psi^{(\rho)})(a_j x - bk)| \right\|_p \le C(\Psi, m, p, a_{min}) \sum_{|\sigma|=m} \|Qf^{(\sigma)}\|_p \tag{55}$$

for all $j > 0$. Next we let $\phi = \mathbb{1}_{bC}/|bC|$, and subtract and add the quantity $I_j[|\chi^m\Psi^{(\rho)}|, \phi]H_m$ inside the L^p norm on the left of (55). By reasoning like we did leading up to (48), we deduce (55) will follow once we verify

$$\left\| I_j[|\chi^m \Psi^{(\rho)}|, \phi] H_m \right\|_p \le C(\Psi, m, p) \sum_{|\sigma|=m} \|f^{(\sigma)}\|_p, \tag{56}$$

$$\left\| I_j[|\chi^m \Psi^{(\rho)}|, \phi] T_j \right\|_p \le C(\Psi, m, p, a_{min}) \sum_{|\sigma|=m} \|Qf^{(\sigma)}\|_p, \tag{57}$$

where $T_j(x,y) = \int_{[0,1]} (S_{a_j^{-1}b} f^{(\sigma)})(x+ty)\, d\omega_m(t)$, $|\sigma| = m$. But inequalities (56)–(57) are essentially the same as (50)–(51) except with $r = m$, and so they are proved already by the paragraph after (50)–(51).

12. Proof of Proposition 9

If $f \in W^{m,p}$ then f can be approximated arbitrarily well in the $W^{m,p}$-norm by linear combinations of functions in $\mathcal{H}$. In the process, $D^\nu f$ gets approximated arbitrarily well in the $W^{m-|\nu|,p}$-norm by linear combinations of functions in $D^\nu \mathcal{H}$.

Thus we need only prove that the collection $\{D^\nu f : f \in W^{m,p}\}$ is dense in $W^{m-|\nu|,p}$. The next lemma does this. Write $\mathcal{S}$ for the Schwartz class.

Lemma 15. *$\{D^\sigma f : f \in \mathcal{S}\}$ is dense in $W^{n,p}$ for all $1 < p < \infty, n \in \mathbb{N} \cup \{0\}$ and multiindices σ.*

Proof of Lemma 15. For $\sigma = 0$, the claim is simply that the Schwartz class is dense in $W^{n,p}$, which is well known.

Now we use induction on σ. The task is to show that if $\{D^\sigma f : f \in \mathcal{S}\}$ is dense in $W^{n,p}$ for all $1 < p < \infty, n \in \mathbb{N} \cup \{0\}$, then the same is true for the multiindex $\sigma + e_t$ for each $t = 1, \ldots, d$. Without loss of generality we can suppose $t = 1$, so that $e_1 = (1, 0, \ldots, 0)$.

Let $1 < p < \infty, n \in \mathbb{N} \cup \{0\}$, and take $u \in \mathcal{S}$ and $\varepsilon > 0$. The induction hypothesis implies that $\|u - D^\sigma f\|_{W^{n+1,p}} < \varepsilon$ for some $f \in \mathcal{S}$. In particular, $\|D_1 u - D^{\sigma+e_1} f\|_{W^{n,p}} < \varepsilon$.

Thus we have only to show that $\{D_1 u : u \in \mathcal{S}\}$ is dense in $W^{n,p}$. Suppose to the contrary that it is not dense. Then by the Hahn–Banach theorem there exists a functional $g \in (W^{n,p})^* \setminus \{0\}$ such that $g[D_1 u] = 0$ for all $u \in \mathcal{S}$.

The functional g can be written as a sum of distributional derivatives, with $g = \sum_{|\tau| \le n} c_\tau D^\tau g_\tau$ for some functions $g_\tau \in L^q$, by the standard representation of the dual space $(W^{n,p})^*$ (see [1, Theorem 3.8]). Hence if η is a mollifier then the mollified distribution

$$g^{(\varepsilon)} = \eta_\varepsilon * g = \sum_{|\tau| \le n} c_\tau \varepsilon^{-|\tau|} (D^\tau \eta)_\varepsilon * g_\tau$$

is a smooth function belonging to L^q, for each $\varepsilon > 0$.

We know $D_1 g^{(\varepsilon)} = \eta_\varepsilon * D_1 g = 0$, because $D_1 g = 0$ as a distribution by construction above. Thus the function $g^{(\varepsilon)}$ is constant in the x_1-direction. Since $g^{(\varepsilon)}$ is also L^q-integrable, it must be identically zero. Letting $\varepsilon \to 0$ gives $g = 0$ as a distribution, and hence by density of $\mathcal{S}$ we see $g = 0$ as a functional on $W^{n,p}$. This contradicts the construction of g, completing the proof. □

13. Proof of Theorem 10

Write $z = b\kappa$. Clearly

$$W^{m,p}\text{-span}\{\psi_{j,k} : k \in \mathbb{Z}^d\} \supset W^{m,p}\text{-span}\{(\Delta_{c,z}\psi)_{j,k} : k \in \mathbb{Z}^d\} \tag{58}$$

because

$$\Delta_{c,z}\psi(a_j x - bk) = \psi(a_j x - bk) - c\psi(a_j x - b(k+\kappa)) \tag{59}$$

and $k + \kappa \in \mathbb{Z}^d$.

To prove the reverse inclusion in (58), first consider the case $|c| < 1$. Temporarily fix $k \in \mathbb{Z}^d$. For $n \ge 1$, examine the linear combination

$$\begin{aligned} \sum_{\ell=0}^{n-1} c^\ell (\Delta_{c,z}\psi)_{j,k+\kappa\ell} &= \sum_{\ell=0}^{n-1} c^\ell [\psi_{j,k+\kappa\ell} - c\psi_{j,k+\kappa(\ell+1)}] && \text{using (59)} \\ &= \psi_{j,k} - c^n \psi_{j,k+\kappa n} && \text{by telescoping} \\ &\to \psi_{j,k} && \text{in } W^{m,p} \text{ as } n \to \infty, \end{aligned}$$

because $|c| < 1$. Thus $\psi_{j,k}$ belongs to the righthand side of (58), so that equality holds in (58).

Next consider $|c| > 1$. (We will reduce to the case "$|c| < 1$".) Notice

$$\begin{aligned}\Delta_{c,z}\psi(x) &= -c[\psi(x-b\kappa) - c^{-1}\psi(x)]\\ &= -c\Delta_{c^{-1},-z}\psi(x-b\kappa)\end{aligned}$$

and hence $(\Delta_{c,z}\psi)_{j,k} = -c(\Delta_{c^{-1},-z}\psi)_{j,k+\kappa}$. Thus

$$W^{m,p}\text{-span}\{(\Delta_{c,z}\psi)_{j,k} : k \in \mathbb{Z}^d\} = W^{m,p}\text{-span}\{(\Delta_{c^{-1},-z}\psi)_{j,k} : k \in \mathbb{Z}^d\}.$$

Since $|c^{-1}| < 1$, the previous case now implies equality in (58).

We have handled all cases $|c| \neq 1$. To complete the proof of the theorem, we now suppose $1 < p < \infty$ and $|c| \leq 1$. To show equality holds in (58), we take $n \geq 1$ and examine the linear combination

$$\begin{aligned}\sum_{\ell=0}^{n-1}\frac{n-\ell}{n}c^\ell(\Delta_{c,z}\psi)_{j,k+\kappa\ell} &= \sum_{\ell=0}^{n-1}\frac{n-\ell}{n}c^\ell[\psi_{j,k+\kappa\ell} - c\psi_{j,k+\kappa(\ell+1)}]\\ &= \sum_{\ell=0}^{n-1}\frac{n-\ell}{n}c^\ell\psi_{j,k+\kappa\ell} - \sum_{\ell=1}^{n}\frac{n-\ell+1}{n}c^\ell\psi_{j,k+\kappa\ell}\\ &= \psi_{j,k} - \frac{1}{n}\sum_{\ell=1}^{n}c^\ell\psi_{j,k+\kappa\ell}.\end{aligned}$$

Thus to show that $\psi_{j,k}$ lies in the closed $W^{m,p}$-span of $\{(\Delta_{c,z}\psi)_{j,k'} : k' \in \mathbb{Z}^d\}$, we need only show

$$\frac{1}{n}\sum_{\ell=1}^{n}c^\ell\psi_{j,k+\kappa\ell} \to 0 \qquad \text{in } W^{m,p} \text{ as } n \to \infty. \tag{60}$$

Take $\varepsilon > 0$ and choose $u \in W^{m,p}$ with compact support and satisfying $\|\psi(a_jx) - u(a_jx)\|_{W^{m,p}} < \varepsilon|a_j|^{-d/p}$ (here we use that $p < \infty$ and j is fixed). Then

$$\left\|\frac{1}{n}\sum_{\ell=1}^{n}c^\ell\psi_{j,k+\kappa\ell} - \frac{1}{n}\sum_{\ell=1}^{n}c^\ell u_{j,k+\kappa\ell}\right\|_{W^{m,p}} \leq \frac{1}{n}\sum_{\ell=1}^{n}\|\psi_{j,k+\kappa\ell} - u_{j,k+\kappa\ell}\|_{W^{m,p}} < \varepsilon$$

for all n (using $|c| \leq 1$). Thus for (60) it remains only to show $\lim_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n}c^\ell u_{j,k+\kappa\ell} = 0$ in $W^{m,p}$ whenever $u \in W^{m,p}$ has compact support.

We may further choose u to be supported in a set of the form $y + bC$ for some $y \in \mathbb{R}^d$ (just by decomposing the original u into a finite sum of functions with such supports, by a partition of unity). Then the functions $u_{j,k+\kappa\ell}$ for $\ell = 1, \ldots, n$ have disjoint supports, so that

$$\begin{aligned}\left\|\frac{1}{n}\sum_{\ell=1}^{n}c^\ell u_{j,k+\kappa\ell}\right\|_{W^{m,p}} &\leq \frac{1}{n}n^{1/p}\|u(a_jx)\|_{W^{m,p}}\\ &\to 0 \qquad \text{as } n \to \infty, \text{ because } p > 1.\end{aligned}$$

Acknowledgments

Part of this paper was researched at the Institute for Mathematical Sciences, National University of Singapore, during the program on "Mathematics and Computation in Imaging Science and Information Processing" in 2004. We thank the IMS for its support. Also we thank Ilya Krishtal for stimulating discussions on frames and the Mexican hat function.

Appendix A. The operators P, Q and S

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

Lemma 16. *If $P|f| \in L^p_{loc}$ for some $1 \le p \le \infty$, then $f \in L^p$.*

Proof of Lemma 16. The result is clear when $p = \infty$, because $|f(x)| \le \sum_{k\in\mathbb{Z}^d} |f(x-bk)|$.
Suppose $1 \le p < \infty$, so that

$$\sum_{k\in\mathbb{Z}^d} |f(x-bk)|^p \le \Big(\sum_{k\in\mathbb{Z}^d} |f(x-bk)|\Big)^p.$$

Hence if $P|f| \in L^p_{loc}$ then $P(|f|^p) \in L^1_{loc}$, which implies $|f|^p \in L^1$ or $f \in L^p$. □

Recall the local supremum operator

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

where the choice of radius $\sqrt{d}$ is convenient but not essential. Obviously $0 \le Qf(x) \le \infty$, and the function Qf is measurable because it is lower semicontinuous.

Incidentally, the norm equivalence

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

is not difficult to show, where $W(L^\infty, \ell^p)$ is the Wiener amalgam space considered by Feichtinger and others (see e.g. [2], [3]).

We prove Lemmas 2 and 3.

Proof of Lemma 2. First we extend the decay condition (4) to all lower derivatives of ψ:

$$|\psi^{(\mu)}(x)| \le C'|x|^{-d-|\mu|-\epsilon} \qquad \text{for each } |\mu| \le m \text{ and almost every } x \text{ with } |x| > R, \tag{61}$$

for some constant $C' > 0$.

For this, it suffices by induction on (4) to show that if $f \in W^{1,1}_{loc}$ and $|Df(x)| \le C|x|^{-n-1}$ for almost every x with $|x| > R$, for some constants $C, n > 0$, then $|f(x)| \le (C/n)|x|^{-n}$ for almost every x with $|x| > R$. The proof goes by radial integration: for almost every direction $v \in S^{d-1}$, the function $F(r) = f(rv)$ belongs to $W^{1,1}_{loc}(0,\infty)$, with $|F(r)| \le \int_r^\infty |Df(sv)|\,ds \le (C/n)r^{-n}$ for almost every $r > R$.

The lemma will now follow once we prove

$$f \in L^p_{loc} \quad \text{and} \quad |f(x)| \le C|x|^{-d-\epsilon} \text{ for almost all } |x| > R \qquad \Longrightarrow \qquad P|f| \in L^p_{loc}, \tag{62}$$

because when $f = \chi^{|\mu|}\psi^{(\mu)}$ we know $f \in L^p_{loc}$ (since $\psi \in W^{m,p}$) while $|f(x)| \le C|x|^{-d-\epsilon}$ for almost all $|x| > R$ by (61).

To prove (62), write $g = \mathbb{1}_{|x|\le R} f$ and $h = \mathbb{1}_{|x|>R} f$, so that $g + h = f$. It is easy to show $P|g| \in L^p_{loc}$, because $g \in L^p$ has compact support. And h has a bounded radially decreasing L^1 majorant of the form $C(1+|x|)^{-d-\epsilon}$, by construction, so that $P|h| \in L^\infty$ by a Riemann sum argument (see [5, Lemma A.2]). Therefore $P|f| \le P|g| + P|h| \in L^p_{loc}$, which proves (62). □

Proof of Lemma 3. We need only show that

$$f \in L^\infty_{loc} \quad \text{and} \quad |f(x)| \le C|x|^{-d-\epsilon} \text{ for almost all } |x| > R \qquad \Longrightarrow \qquad Qf \in L^1, \tag{63}$$

because when $f = \chi^{|\mu|}\psi^{(\mu)}$ we know $f \in L^\infty_{loc}$ (since $\psi \in W^{m,\infty}$) while hypothesis (4) ensures (61) and so $|f(x)| \le C|x|^{-d-\epsilon}$ for almost all $|x| > R$.

To prove (63), write $f = g + h$ like in the previous proof. Then $Qg \in L^1$ because g is bounded with compact support, and $Qh \in L^1$ because h has a bounded radially decreasing L^1 majorant (cf. [6, Lemma 21]). Therefore $Qf \le Qg + Qh \in L^1$. □

Next we derive pointwise relations for f and Qf.

Lemma 17. *We have $|f| \le Qf$ a.e. And if E is a bounded set in $\mathbb{R}^d$ then*

$$|f(x)| \le Q_E f(y) := \sum_{k:|k|<\operatorname{diam}(E)+\sqrt{d}} Qf(y+k) \tag{64}$$

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

Proof of Lemma 17. 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 essentially locally bounded and 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 $Qf(x) = \infty \ge |f(x)|$.

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 $Qf(y+k) = \infty$ (since otherwise $y + k \in F$, which implies $x \in G$). Either way, we have proved (64). $\square$

Now we prove norm relations between P and Q.

Lemma 18.

$$\|f\|_q \le \|Qf\|_r \qquad \textit{for all } 1 \le r \le q \le \infty, \textit{ and}$$
$$\|Pf\|_\infty \le \|P|f|\|_\infty \le C\|Qf\|_1.$$

Proof of Lemma 18. We will show $\|f\|_\infty \le \|Qf\|_r$ for all $1 \le r \le \infty$, which proves the first inequality in the lemma for $q = \infty$. Then for $q < \infty$,

$$|f|^q = |f|^{q-r}|f|^r \le \|f\|_\infty^{q-r}|Qf|^r \quad \text{a.e.}$$

by Lemma 17, and so $\|f\|_q \le \|Qf\|_r$ as desired.

To show $\|f\|_\infty \le \|Qf\|_r$, suppose $1 \le r < \infty$ (noting the case $r = \infty$ follows from $|f| \le Qf$ in Lemma 17). For each $y \in \mathbb{R}^d$,

$$\begin{aligned}
\|f\|_\infty^r &\le \sup_{k\in\mathbb{Z}^d} \|f\|^r_{L^\infty(B(y+k,\sqrt{d}))} \\
&\le \sum_{k\in\mathbb{Z}^d} \|f\|^r_{L^\infty(B(y+k,\sqrt{d}))} \\
&= \sum_{k\in\mathbb{Z}^d} Qf(y+k)^r.
\end{aligned}$$

Integrating over $y \in \mathcal{C}$ gives $\|f\|_\infty^r \le \|Qf\|_r^r$, as we wanted.

For the second inequality in the lemma, we have $|f(x - bk)| \le Q_E f(x - bk - y)$ for all $k \in \mathbb{Z}^d$ and almost every $(x, y) \in \mathbb{R}^d \times E$ by applying Lemma 17 with $E = b\mathcal{C}$. For such x and y values,

$$|\det b|^{-1} P|f|(x) \le \sum_{k\in\mathbb{Z}^d} Q_E f(x - bk - y).$$

Integrating over $y \in b\mathcal{C}$ yields that for almost every x,

$$P|f|(x) \le \int_{b\mathcal{C}} \sum_{k \in \mathbb{Z}^d} Q_E f(x - bk - y)\, dy$$
$$= \int_{\mathbb{R}^d} Q_E f(x-y)\, dy = \|Q_E f\|_1 \le C(b\mathcal{C}) \|Qf\|_1$$

by definition of Q_E in (64). $\square$

Next we investigate norm convergence of scale averages of rapidly oscillating functions, as used in the proof of Theorem 4.

Lemma 19. *Suppose that $g \in L^p_{loc}$ (for some $1 \le p < \infty$) is $b\mathbb{Z}^d$-periodic with mean value zero. If the dilations a_j grow exponentially and $f \in L^p$ with $Qf \in L^p$, then*

$$\lim_{J\to\infty} f(x) \frac{1}{J} \sum_{j=1}^{J} g(a_j x) = 0 \qquad \text{in } L^p.$$

Proof. We have

$$\lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} g(a_j x) = 0 \qquad \text{in } L^p_{loc} \tag{65}$$

by [5, Lemma 3]. And $f \in L^\infty$ by Lemma 18, since $Qf \in L^p$.

Let $R > 0$ be arbitrary. Then

$$\int_{B(0,R)} \left| f(x) \frac{1}{J} \sum_{j=1}^{J} g(a_j x) \right|^p dx \le \|f\|_\infty^p \int_{B(0,R)} \left| \frac{1}{J} \sum_{j=1}^{J} g(a_j x) \right|^p dx$$
$$\to 0 \qquad \text{as } J \to \infty, \text{ by (65)}.$$

Furthermore, $|f(x)| \le Qf(k)$ for almost every $x \in k + \mathcal{C} \subset B(k, \sqrt{d})$, by definition of Q. Thus for each J,

$$\begin{aligned}
\int_{\mathbb{R}^d \setminus B(0,R)} \left| f(x) \frac{1}{J} \sum_{j=1}^{J} g(a_j x) \right|^p dx &\le \sum_{|k| > R - \sqrt{d}} Qf(k)^p \int_{k+\mathcal{C}} \left| \frac{1}{J} \sum_{j=1}^{J} g(a_j x) \right|^p dx \\
&\le \sum_{|k| > R - \sqrt{d}} Qf(k)^p \left(\frac{1}{J} \sum_{j=1}^{J} \left(\int_{k+\mathcal{C}} |g(a_j x)|^p\, dx \right)^{1/p} \right)^p \\
&\le \sum_{|k| > R - \sqrt{d}} Qf(k)^p \left(\frac{1}{J} \sum_{j=1}^{J} \left(|a_j|^{-d} \int_{a_j(k+\mathcal{C})} |g(x)|^p\, dx \right)^{1/p} \right)^p \\
&\le \sum_{|k| > R - \sqrt{d}} Qf(k)^p \cdot C \|g\|^p_{L^p(b\mathcal{C})} \qquad (66)
\end{aligned}$$

since the mean value of the $b\mathbb{Z}^d$-periodic function $|g|^p$ over the set $a_j(k+\mathcal{C})$ is bounded by a constant times its mean value over the period cell $b\mathcal{C}$ (see for example [6, Lemma 25]; the constant C depends on $\min_{j>0} |a_j|$). The expression (66) can be made as small as we like by choosing R sufficiently large, because $\sum_{k \in \mathbb{Z}^d} Qf(k)^p < \infty$ as explained below. Letting $R \to \infty$ then proves the lemma.

We have

$$B(x, \sqrt{d}) \subset B(0, 2\sqrt{d}) \subset \cup_{|\ell| < 3\sqrt{d}} B(\ell, \sqrt{d}) \qquad \text{whenever } |x| < \sqrt{d},$$

and translating these balls by $k-x$ shows $Qf(k)^p \leq \sum_{|\ell|<3\sqrt{d}} Qf(\ell+k-x)^p$. Integrating this over $x \in C$ and then summing over k gives that

$$\begin{aligned}\sum_{k\in\mathbb{Z}^d} Qf(k)^p &\leq \sum_{k\in\mathbb{Z}^d} \int_C \sum_{|\ell|<3\sqrt{d}} Qf(\ell+k-x)^p\,dx \\ &= \sum_{|\ell|<3\sqrt{d}} \|Qf\|_p^p < \infty.\end{aligned}$$

Our final lemma concerns the *modulus of continuity* function

$$S_a f(x) = \|f(x) - f(\cdot)\|_{L^\infty(B(x,\|a\|\sqrt{d}))},$$

where a is a $d \times d$ matrix.

Lemma 20. *Let a be a $d\times d$ matrix, and $1 \leq p \leq \infty$. Then*

$$\|S_a f\|_p \leq C(\|a\|)\|Qf\|_p.$$

If $Qf \in L^p$ and f is continuous, then $S_a f \to 0$ in L^p as $\|a\| \to 0$.

Proof of Lemma 20. Let $\mathcal{K}(a)$ be a finite collection of lattice points in $\mathbb{Z}^d$ such that $B(0,\|a\|\sqrt{d}) \subset \cup_{k\in\mathcal{K}(a)} B(k,\sqrt{d})$. Then

$$S_a f(x) \leq |f(x)| + \|f\|_{L^\infty(B(x,\|a\|\sqrt{d}))} \leq Qf(x) + \sum_{k\in\mathcal{K}(a)} Qf(x+k) \quad \text{a.e.}$$

because $|f| \leq Qf$ a.e. by Lemma 17. Hence $\|S_a f\|_p \leq (1+|\mathcal{K}(a)|)\|Qf\|_p$.

Now suppose $Qf \in L^p$ and f is continuous. Then $S_a f(x) \to 0$ for each x as $\|a\| \to 0$, by definition of $S_a f$, so that $S_a f \to 0$ in L^p by dominated convergence (with dominating function $|f| + Qf$ for all $\|a\| \leq 1$, since $\mathcal{K}(a) = \{0\}$ when $\|a\| \leq 1$). □

References

[1] R. A. Adams. Sobolev spaces. Academic Press, New York, NY, 1975.

[2] A. Aldroubi and H.G. Feichtinger. *Exact iterative reconstruction algorithm for multivariate irregularly sampled functions in spline-like spaces: The L^p-theory.* Proc Amer Math Soc 126:2677–2686, 1998.

[3] A. Aldroubi and K. Gröchenig. *Nonuniform sampling and reconstruction in shift-invariant spaces.* SIAM Review 43:585–620, 2001.

[4] I. Babuška. *Approximation by hill functions.* Comment Math Univ Carolinae 11:787–811, 1970.

[5] H.-Q. Bui and R. S. Laugesen. *Affine systems that span Lebesgue spaces.* J. Fourier Anal. Appl., vol. 11, appeared online `http://dx.doi.org/10.1007/s00041-005-4049-2`

[6] H.-Q. Bui and R. S. Laugesen. *Spanning and sampling in Lebesgue and Sobolev spaces.* University of Canterbury Research Report UCDMS2004/8, 2004. `www.math.uiuc.edu/~laugesen/publications.html`

[7] H.-Q. Bui and R. S. Laugesen. *Approximation and spanning in the Hardy space, by affine systems.* Preprint. `www.math.uiuc.edu/~laugesen/publications.html`

[8] H.-Q. Bui and M. Paluszyński. *On the phi and psi transforms of Frazier and Jawerth.* University of Canterbury Research Report UCDMS2004/11, 18 pages, 2004.

[9] P. G. Casazza and O. Christensen. *Weyl–Heisenberg frames for subspaces of $L^2(\mathbb{R})$.* Proc Amer Math Soc 129:145–154, 2001.

[10] O. Christensen. An introduction to frames and Riesz bases. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Boston, MA, 2003.

[11] I. Daubechies. *The wavelet transform, time-frequency localization and signal analysis.* IEEE Trans Inform Theory 36:961–1005, 1990.

[12] 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.

[13] G. Fix and G. Strang. *Fourier analysis of the finite element method in Ritz–Galerkin theory.* Stud Appl Math 48:265–273, 1969.

[14] M. Frazier, B. Jawerth and G. Weiss. Littlewood-Paley theory and the study of function spaces. CBMS Reg Conf Ser in Math, No. 79, Amer. Math. Soc., Providence, RI, 1991.

[15] 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.* Memoirs Amer Math Soc 156, no. 742, 74 pp., 2002.

[16] F. di Guglielmo. *Construction d'approximations des espaces de Sobolev sur des réseaux en simplexes.* Calcolo 6:279–331, 1969.

[17] E. Hernández and G. Weiss. A first course on wavelets. CRC Press, Boca Raton, Florida, 1996.

[18] O. Holtz and A. Ron. *Approximation orders of shift-invariant subspaces of* $W_2^s(\mathbb{R}^d)$. J Approx Theory 132:97–148, 2005.

[19] K. Jetter and D.-X. Zhou. *Seminorm and full norm order of linear approximation from shift-invariant spaces.* Rend Sem Mat Fis Milano 65 (1995): 277–302, 1997.

[20] R.-Q. Jia. *Approximation with scaled shift-invariant spaces by means of quasi-projection operators.* J Approx Theory 131:30–46, 2004.

[21] R.-Q. Jia, J. Wang and D.-X. Zhou. *Compactly supported wavelet bases for Sobolev spaces.* Appl Comput Harmon Anal 15:224–241, 2003.

[22] M. J. Johnson. *On the approximation order of principal shift-invariant subspaces of* $L^p(\mathbf{R}^d)$. J Approx Theory 91:279–319, 1997.

[23] V. Maz'ya and G. Schmidt. *On approximate approximations using Gaussian kernels.* IMA J Numer Anal 16:13–29, 1996.

[24] V. Maz'ya and G. Schmidt. *On quasi-interpolation with non-uniformly distributed centers on domains and manifolds.* J Approx Theory 110:125–145, 2001.

[25] Y. Meyer. Wavelets and operators. Cambridge University Press, Cambridge, 1992.

[26] S. G. Mikhlin. Approximation on a rectangular grid. Translated by R. S. Anderssen and T. O. Shaposhnikova. Sijthoff & Noordhoff, The Netherlands, 1979.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

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`