Non-uniform weighted average sampling and reconstruction in shift-invariant and wavelet spaces

Akram Aldroubi

Department of Mathematics, Vanderbilt University, Nashville, TN 37240, USA

Received 1 May 2001; revised 7 June 2002

Communicated by Hans Feichtinger

Abstract

The problem of reconstructing a function \( f \) from a set of non-uniformly distributed, weighted-average sampled values \( \{ \int_{\mathbb{R}^d} f(x) \psi_{x_j}(x) \, dx : j \in J \} \) is studied in the context of shift-invariant subspaces of \( L^2(\mathbb{R}^d) \). Necessary density conditions on the sampling set \( X = \{ x_j : j \in J \} \) for stable reconstruction are obtained, and fast iterative algorithms are described. The performance of the algorithms are analyzed when the data are corrupted by noise. Estimates are derived for the convergence rates of the algorithms in terms of the sampling density and the diameters of the sampling functionals \( \{ \psi_{x_j} : x_j \in X \} \). The results provide a mathematical framework for situations arising frequently in applications, e.g., when the sample values are not precise because they are gathered by real-world acquisition devices.

Keywords: Irregular sampling; Frame; Reproducing kernel Hilbert space; Amalgam spaces; Wavelet

1. Introduction

In the classical sampling problem, the objective is to recover a function \( f \) on \( \mathbb{R}^d \) from its samples \( \{ f(x_j) : j \in J \} \), where \( J \) is a countable indexing set. For this problem to be well-posed, the function \( f \) is assumed to be bandlimited, or to belong to a shift invariant space of the form

\[
V^2(\phi) = \left\{ \sum_{k \in \mathbb{Z}^d} c_k \phi(\cdot - k) : c \in \ell^2 \right\},
\]

(1.1)

see, for example, [2–6,9,10,17,18,21–25] and the references therein. Obviously, the space \( V^2(\phi) \) is not a space of bandlimited functions unless the generator \( \phi \) is bandlimited.

Well posedness of the sampling problem implies that the following inequalities must hold:

\[
c \|f\|_{L^2} \leq \left( \sum_{x_j \in X} |f(x_j)|^2 \right)^{1/2} \leq C \|f\|_{L^2},
\]

(1.2)

where \( c \) and \( C \) are positive constants independent of \( f \in V^2(\phi) \).
In practice, the assumption that the samples \( \{ f(x_j): j \in J \} \) can be measured exactly is not realistic. A better assumption is that a weighted-average value in the neighborhood of \( x_j \) is obtained. This means that the sampled data are of the form

\[
g_{x_j} = \langle f, \psi_{x_j} \rangle = \int_{\mathbb{R}^d} f(x) \psi_{x_j}(x) \, dx,
\]

where \( \{ \psi_{x_j}: x_j \in X \} \) is a set of functionals that act on the function \( f \) to produce the data \( \{ g_{x_j}: x_j \in X \} \). The functionals \( \{ \psi_{x_j}: x_j \in X \} \) may reflect the characteristics of the sampling devices.

For this case, the well posedness condition (1.2) must be changed to

\[
c \| f \|_{L^2} \leq \left( \sum_{x_j \in X} |g_{x_j}(f)|^2 \right)^{1/2} \leq C \| f \|_{L^2}, \quad \text{for all } f \in V^2(\phi),
\]

where \( g_{x_j} \) is defined by (1.3) and where \( c \) and \( C \) are positive constants independent of \( f \).

A particular case is when all the measurements are obtained from a single device with impulse response \( \psi \). For this case, the functionals are of the form \( \psi_{x_j}(\cdot) = \psi(\cdot - x_j) \).

In one dimension and for the special case of bandlimited functions, Gröchenig [15] proved that if \( |x_{j+1} - x_j| \leq \delta < \sqrt{2}/2 \), then any band-limited function \( f \) with \( \text{supp}(\hat{f}) \subset [-1/2, 1/2] \) is uniquely determined from its averages \( \langle f, \psi_{x_j} \rangle \), provided that

\[
\text{supp} \psi_{x_j} \subset \left[ x_j - \frac{\delta}{2}, x_j + \frac{\delta}{2} \right], \quad \psi_{x_j} \geq 0.
\]

Under the same assumptions, Sun and Zhou [19] extended these results, and derived frame algorithms for the reconstruction. They also gave bounds on the error of reconstruction when a non-band-limited function is reconstructed by the frame algorithms. For dimension one, Sun and Zhou also showed that if the maximal gap between consecutive sampling points is smaller than a characteristic length, then a function in a spline subspace is uniquely determined from local averages obtained from averaging functions satisfying (1.5) [20].

In this article, we will investigate the problem of reconstructing a function \( f \) from a set of non-uniformly distributed weighted-average samples. The problem is studied in \( \mathbb{R}^d \) and in the context of shift-invariant spaces. Reconstruction from averages and its connection with frames is considered in Section 2. Amalgam spaces and shift-invariant spaces are reviewed in Section 3.1. Iterative algorithms for the reconstruction are described in Section 4.1. Performance of the algorithms in the presence of noise and the density conditions for reconstruction are discussed in Sections 4.2 and 4.3. Proofs of some of the results are postponed to Section 5.

2. Weighted-average sampling, reconstruction, and frames

For a shift-invariant space

\[
V^2(\phi) = \left\{ \sum_{k \in \mathbb{Z}^d} c_k \phi(\cdot - k): c \in \ell^2 \right\};
\]

the standard assumption on the generator \( \phi \) is that there exists two constants \( m > 0 \) and \( M > 0 \) such that

\[
0 < m \leq \hat{\phi}(\xi) = \sum_{j \in \mathbb{Z}^d} |\hat{\phi}(\xi + j)|^2 \leq M < \infty \quad \text{a.e. } \xi.
\]

Under the condition above, the space \( V^2(\phi) \) has a Hilbert space structure, it is a subspace of \( L^2 \), and the set \( \{ \phi(\cdot - k): k \in \mathbb{Z}^d \} \) forms a Riesz basis for \( V^2(\phi) \).
The well-posedness condition (1.4) is similar to a frame condition. However, the set \{\psi_{x_j} : x_j \in X\} does not necessarily form a frame for \(V^2(\phi)\) since, in general, the functions \(\psi_{x_j}\) do not belong to \(V^2(\phi)\).

Consider the orthogonal projection \(\theta_{x_j} = \mathbf{P} \psi_{x_j}\) of \(\psi_{x_j}\) on \(V^2(\phi)\). Then for any \(f \in V^2(\phi)\), we have
\[
(f, \psi_{x_j}) = (\mathbf{P} f, \psi_{x_j}) = (f, \mathbf{P} \psi_{x_j}) = (f, \theta_{x_j}).
\]

It follows that (1.4) holds if and only if
\[
c \|f\|_{L^2} \leq \left( \sum_{x_j \in X} \| (f, \theta_{x_j}) \|^2 \right)^{1/2} \leq C \|f\|_{L^2} \quad \text{for all } f \in V^2(\phi).
\]
Thus (1.4) holds if and only if \(\{\theta_{x_j} : x_j \in X\} \subset V^2(\phi)\) constitutes a frame for \(V^2(\phi)\).

**Proposition 2.1.** Let \(\mathbf{P}\) be the orthogonal projector from \(L^2\) onto \(V^2(\phi)\). Then the set \(\{\theta_{x_j} = \mathbf{P} \psi_{x_j} : x_j \in X\}\) is a frame for \(V^2(\phi)\) if and only if Condition (1.4) holds. Consequently, if (1.4) holds, then there exists a dual frame \(\{\tilde{\theta}_{x_j}, x_j \in X\}\) for \(V^2(\phi)\), that allows us to write the reconstruction formula
\[
f = \sum_{j \in J} (f, \theta_{x_j}) \tilde{\theta}_{x_j} = \sum_{j \in J} (f, \mathbf{P} \psi_{x_j}) \tilde{\theta}_{x_j} = \sum_{j \in J} (f, \psi_{x_j}) \tilde{\theta}_{x_j}.
\]

If the sampling functions \(\psi_{x_j}\) are generated by shifts of a single function, e.g., \(\psi_{x_j} = \psi(\cdot - x_j)\), then the orthogonal projection \(\theta_{x_j} = \mathbf{P} \psi_{x_j}\) is given by
\[
\theta_{x_j}(x) = \sum_{k \in \mathbb{Z}^d} \hat{\phi}(\cdot - k) \hat{\phi}(x - k),
\]
where \(\hat{\phi} \in V^2(\phi)\) is the dual of the generator \(\phi\), i.e., \(\hat{\phi} \in V^2(\phi)\) is the unique function that satisfies \(\langle \hat{\phi}(\cdot), \phi(\cdot - k) \rangle = \delta_{k,0}, k \in \mathbb{Z}^d\). The dual generator \(\hat{\phi} \in V^2(\phi)\) can be written as
\[
\hat{\phi}(\cdot) = \sum_{k \in \mathbb{Z}^d} \alpha_k \hat{\phi}(\cdot - k),
\]
where \(\alpha_k\) are the Fourier coefficients of \(\hat{\phi} = (1/\hat{\alpha}_\phi) \in L^2(0, 1)^d\) (see Definition 2.2 for \(\hat{\alpha}_\phi\)). By writing the inner product in expression (2.3) in terms of convolution, we obtain
\[
\theta_{x_j}(x) = \sum_{k \in \mathbb{Z}^d} (\psi * \hat{\phi}^*)(k - x_j) \phi(x - k),
\]
where \(\hat{\phi}^*(x) = \overline{\phi(-x)}\). Therefore, the frame \(\{\theta_{x_j} = \mathbf{P} \psi_{x_j} : x_j \in X\}\) can be expressed explicitly in terms of the kernel
\[
K_x(y) = \sum_{k \in \mathbb{Z}^d} (\psi * \hat{\phi}^*)(k - x) \phi(y - k),
\]
where \(\hat{\phi}^*(x) = \overline{\phi(-x)}\), and we have the following theorem.

**Theorem 2.2.** If \(\phi\) satisfies condition (2.2), and \(\psi_{x_j}(\cdot) = \psi(\cdot - x_j)\) for some \(\psi \in L^2\), then the set \(\{\theta_{x_j} = \mathbf{P} \psi_{x_j} : x_j \in X\}\) can be obtained in terms of the kernel \(K_x\) as \(\theta_{x_j} = K_{x_j}\) for all \(x_j \in X\).

Except in special cases (e.g., uniform sampling), Proposition 2.1 or Theorem 2.2 are not useful in practice because the construction of the frame \(\{\theta_{x_j} : x_j \in X\}\) from \(\{\psi_{x_j} : x_j \in X\}\) is a lengthy task, and it must be repeated for every new sampling set \(X\). The construction of the dual frame \(\{\tilde{\theta}_{x_j} : x_j \in X\}\) is even more challenging. However, Proposition 2.1 and Theorem 2.2 are useful for theoretical purposes. For example, we will use them in Section 4.3 to find conditions on the sampling density of \(X\) for exact and stable reconstruction.
3. Notation and preliminaries

For the sampling problem we need to impose regularity requirements on the space \( V^2(\phi) \) (e.g., for ideal sampling, \( V^2(\phi) \) must be a space of continuous functions). Wiener amalgam spaces \([11,14,16]\) are useful in this context and they are defined as follows: a measurable function \( f \) belongs to \( W^p \) if it satisfies
\[
\| f \|_{W^p} = \sum_{k \in \mathbb{Z}^d} \text{ess sup} \{ |f(x + k)|^p : x \in [0, 1]^d \} < \infty.
\] (3.1)

If \( p = \infty \), a measurable function \( f \) belongs to \( W^\infty \) if it satisfies
\[
\| f \|_{W^\infty} = \sup_{k \in \mathbb{Z}^d} \{ \text{ess sup} \{ |f(x + k)| : x \in [0, 1]^d \} \} < \infty.
\] (3.2)

In this case \( W^\infty \) coincides with \( L^\infty \).

Endowed with this norm, \( W^p \) becomes a Banach space \([12]\). The subspace of continuous functions \( W^p_0 \subset W^p \) coincides with the closure of the test functions in \( W^p \) for \( 1 \leq p < \infty \), hence it is a closed subspace of \( W^p \) and thus also a Banach space \([12]\). We have the following inclusions between the various spaces:
\[
W^p_0 \subset W^p_0 \subset W^q \subset L^q, \quad 1 \leq p \leq q \leq \infty.
\] (3.3)

The following convolution relations are useful \([3]\):

(i) Let \( f \in L^p \) and \( g \in W^1 \), then \( f * g \in W^p \) and we have
\[
\| f * g \|_{W^p} \leq C \| f \|_{L^p} \| g \|_{W^1}, \quad \text{for all } 1 \leq p \leq \infty.
\] (3.4)

(ii) Let \( c \in \ell^p \) and \( \varphi \in W^1 \), then
\[
\left\| \sum_{k \in \mathbb{Z}^d} c_k \varphi(\cdot - k) \right\|_{W^p} \leq \| c \|_{\ell^p} \| \varphi \|_{W^1}, \quad \text{for all } 1 \leq p \leq \infty.
\] (3.5)

(iii) If \( f \in L^p \) and \( g \in W^1 \), then the sequence \( d \) defined by \( d_k = \int_{\mathbb{Z}^d} f(x)g(x - k) \, dx \), \( k \in \mathbb{Z}^d \), belongs to \( \ell^p \) and we have
\[
\| d \|_{\ell^p} \leq \| f \|_{L^p} \| g \|_{W^1}, \quad \text{for all } 1 \leq p \leq \infty.
\] (3.6)

3.1. Shift-invariant spaces

In addition to the requirement that the generator of \( V^2(\phi) \) in (2.1) satisfies (2.2), we also require the generator \( \phi \) to belong to \( W^1_0 \). With these requirements, it is well known that the space \( V^2(\phi) \) is a space of continuous functions, and we have the following properties \([3]\):

(i) The space \( V^2(\phi) \) is a closed subspace of \( L^2 \) and of \( W^2_0 \), and \( \{ \varphi(\cdot - k) : k \in \mathbb{Z}^d \} \) is a Riesz basis for \( V^2(\phi) \), i.e., there exist constants \( m > 0 \), \( M > 0 \) such that
\[
m \| c \|_{\ell^2} \leq \left\| \sum_{k \in \mathbb{Z}^d} c_k \varphi(\cdot - k) \right\|_{L^2} \leq M \| c \|_{\ell^2}, \quad \forall c = (c_k) \in \ell^2(\mathbb{Z}^d),
\] (3.7)

and we have the norm equivalences \( \| f \|_{L^2} \approx \| c \|_{\ell^2} \approx \| f \|_{W^2} \).

(ii) The biorthogonal system is given by \( \mathbb{Z}^d \) translates of the dual function \( \tilde{\varphi} \) which also belongs to \( W^1_0 \).

(iii) If \( X = \{ x_j : j \in J \} \) is such that \( \inf_{j,l} |x_j - x_l| = \delta > 0 \), then
\[
\left( \sum_{x_j \in X} |f(x_j)|^2 \right)^{1/2} \leq C_\delta \| f \|_{L^2}, \quad \text{for all } f \in W^2_0.
\] (3.8)

In particular, (3.8) holds for all \( f \in V^2(\phi) \).
4. Main results

For the remainder of this article, we will assume that the sampling functions \{\psi_{x_j}: x_j \in X\} have compact support and satisfy the following properties:

(i) \(\text{supp} \psi_{x_j} \subset x_j + [-a, a]^d\) for all \(x_j \in X\);
(ii) Each function \(|\psi_{x_j}|\) is bounded above by some constant \(U_j\);
(iii) \(\int_{\mathbb{R}^d} |\psi_{x_j}| \leq M\) for all \(x_j \in X\); and
(iv) \(\int_{\mathbb{R}^d} \psi_{x_j} = 1\).

Condition (iv) is only a convenient normalization for making the proofs more transparent, and for avoiding complicated expressions with explicit values of the integral \(\int_{\mathbb{R}^d} \psi_{x_j}\).

4.1. Fast iterative reconstruction algorithms

Fast iterative schemes for the reconstruction of functions from their samples has been introduced by Feichtinger and Gröchenig [13] for the case of bandlimited functions. These schemes have been extended by Aldroubi and Feichtinger [1] to general shift-invariant spaces. In this article, we will develop the theory of fast iterative reconstruction schemes for the case of average sampling in shift-invariant spaces. First, we need to introduce the notion of \(\gamma\) density useful in this regards.

**Definition 4.1.** A set \(X = \{x_j: j \in J\}\) is \(\gamma\) dense in \(\mathbb{R}^d\) if
\[
\mathbb{R}^d = \bigcup_j B_r(x_j), \quad \text{for every } r > \gamma,
\]
where \(B_r(x_j)\) are balls centered on \(x_j\), and with radius \(r\).

This definition implies that the distance of any sampling point to its next neighbor is at most \(2\gamma\). Thus strictly speaking, \(\gamma\) is the inverse of a density, i.e., if \(\gamma\) increases, the number of points per unit cube decreases.

The iterative algorithm that we develop uses a quasi-reconstruction operator \(A_X\) in the iteration scheme. To define this operator, we start from a partition of unity \(\{\beta_j\}_{j \in J}\) defined as follows.

**Definition 4.2.** A bounded partition of unity (BPU) adapted to \(\{B_{\gamma}(x_j)\}_{j \in J}\) is a set of functions \(\{\beta_j\}_{j \in J}\) that satisfy:

1. \(0 \leq \beta_j \leq 1, \forall j \in J\);
2. \(\text{supp} \beta_j \subset B_{\gamma}(x_j)\); and
3. \(\sum_{j \in J} \beta_j = 1\).

The constructions of such BPUs can be obtained by well-known standard techniques. The operator \(A_X\) is then defined by
\[
A_X f = \sum_{j \in J} (f, \psi_{x_j}) \beta_j,
\]
where as before \(\text{supp} \psi_{x_j} \subset x_j + [-a, a]^d\). Obviously the quasi-reconstruction \(A_X f\) does not belong to the space \(V^2(\phi)\). However, we can use this reconstruction in an iterative scheme to recover the exact function \(f \in V^2(\phi)\) as follows.

**Theorem 4.1.** Let the generator \(\phi \in W^1_0\) be given. Then there exists a density \(\gamma = \gamma(\phi) > 0\) and \(a_0 > 0\) such that any \(f \in V^2(\phi)\) can be recovered from the data \(\{(f, \psi_{x_j}): j \in J\}\) on any \(\gamma\)-dense set \(X = \{x_j: j \in J\}\) and for any support size condition (for \(\psi_{x_j}\)) \(0 < a \leq a_0\).
by the following iterative algorithm:

$$f_1 = P A_X f, \quad f_{n+1} = P A_X(f - f_n) + f_n, \quad (4.3)$$

where $P$ is the orthogonal projector from $L^2$ onto $V^2(\phi)$. In this case, the iterate $f_n$ converges to $f$ in the $W^2$ norm, hence both in the $L^2$ norm, and uniformly. The convergence is geometric, that is,

$$\|f - f_n\|_{L^2} \leq \|f - f_n\|_{W^2} \leq C_1 \alpha^n \|f - f_1\|_{W^2}$$

for some $\alpha = \alpha(\gamma, a_0, \phi) < 1$ and $C_1 < \infty$.

Obviously, the case $\psi_{x_j}(\cdot) = (1/a^d)\psi((\cdot - x_j)/a)$ where $\psi$ has compact support is just a special case and Theorem 4.1 applies.

**Remark 4.1.** (i) Theorem 4.1 does not rule out sampling point clustering. Thus in principle, algorithm (4.3) still works even in the presence of clustering. However, if the sampling set $X$ is separated, i.e., $\inf|x_j - x_i| = \delta > 0$, then (1.4) holds and we have the following norm equivalence

$$c \|f\|_{L^2} \leq \left( \sum_{x_j \in X} |\langle f, \psi_{x_j} \rangle|^2 \right)^{1/2} \leq C \|f\|_{L^2},$$

where $c$ and $C$ are positive constants that may depend on $\delta$ but that are independent of $f$.

### 4.2. Reconstruction in presence of noise

In practice, the sampled data are often corrupted by noise. Moreover, the assumption that the function $f$ belongs to some specific space $V^2(\phi)$ is often an idealization. Thus, it is important to know whether algorithm (4.3) still converges under non-ideal circumstances. To investigate these situations, we only assume that the data $f' = \{f'_j: j \in J\}$ belong to $\ell^2$, but we do not assume that $f' = \{f'_j: j \in J\}$ are samples of a function $f \in V^2(\phi)$. For this case we use the initialization

$$f_1 = P Q_X \{f'_j\} := P \left( \sum_{j \in J} f'_j \beta_j \right), \quad (4.4)$$

where $\{\beta_j: j \in J\}$ is the BPU in Definition 4.2. Algorithm (4.3) becomes

$$f_{n+1} = f_1 + (I - P A_X) f_n, \quad (4.5)$$

and we have the following theorem.

**Theorem 4.2.** Under the same assumptions as in Theorem 4.1, the algorithm (4.5) converges to a function $f_\infty \in V^2(\phi)$ which satisfies $P(A_X f_\infty - Q_X \{f'_j\}) = 0$.

### 4.3. Density for exact reconstruction

Obviously, to reconstruct a function from its weighted samples $g_{x_j} = \langle f, \psi_{x_j} \rangle$ the sampling set $X$ must be sufficiently dense. If the supports of $\phi$ and $\{\psi_{x_j}: x_j \in X\}$ are compact, then we have the following lower bound on the density of points in any cube $C = (r, s)^d \subset \mathbb{R}^d$.

**Theorem 4.3.** Assume that $\text{supp} \psi_{x_j} \subset x_j + [-a, a]^d$ for all $x_j \in X$, and that $\text{supp} \phi \subset [-b, b]^d$, where $a$ and $b$ are positive integers. If the exact stable reconstruction condition (1.4) is satisfied, then every open cube $C = (r, s)^d$ with side length $(s - r) \geq 2a + 2b$ contains at least $((s - r) - 2a - 2b)^d$ points of $X$ (here $\lfloor t \rfloor$ denotes the greatest integer less than or equal to $t$).
As a corollary, we immediately obtain a lower bound on the Beurling density defined by [7,8]

\[ D^-(X) = \lim_{r \to \infty} \inf_{y \in d(\#X \cap (y + [0, r]^d))} \min_{y \in d(r)} \]  

(4.6)

**Corollary 4.4.** Under the same assumption as Theorem 4.3, if the stable reconstruction condition (1.4) is satisfied, then \( D^-(X) \geq 1 \).

Note that if a set \( X \) is \( \gamma \)-dense, then its Beurling density defined by (4.6) satisfies \( D^-(X) \geq \gamma^{-1} \). This last relation states that \( \gamma \)-density imposes more constraints on a sampling set \( X \) than the Beurling density \( D^-(X) \).

An estimate of the convergence rate \( \alpha \) in Theorem 4.1 in terms of the \( \gamma \) density, \( a \), and \( \phi \) is given by the following theorem.

**Theorem 4.5.** Assume that \( \phi \in W^1_0 \) and \( |\nabla \phi| \in W^1_0 \). Let \( M \) be such that \( \int_{\mathbb{R}^d} |\psi_{x_j}| \leq M \).

Then the convergence rate \( \alpha \) in Theorem 4.1 satisfies

\[ \alpha \leq \frac{1}{m} \left( \gamma (1 + 2\lceil \gamma \rceil)^d + M((1 + 2\lceil \gamma \rceil)^d + 2) a (1 + 2\lceil a \rceil)^d \right) \| \nabla \phi \|_{W^1}, \]

where \( m \) is the lower bound constant in (3.7) (here \( \lceil t \rceil \) denotes the smallest integer bigger than or equal to \( t \)).

**Remark 4.2.** The relevance of Theorem 4.5 is that it allows to find the density \( \gamma \) and the value \( a_0 \) needed for the reconstruction algorithm (4.3) to converge. In addition, Theorem 4.5 gives us an estimate of the convergence rate \( \alpha \) (in terms of \( \phi \), \( a \), and \( M \)) in Theorem 4.1.

5. Proofs

5.1. Proof of Theorem 4.1

To prove Theorem 4.1, we need to introduce the quasi-interpolant \( Q_X \) of the sampled values \( f|X \) of a function \( f \in W^2_0 \). Given a partition of unity \( \{\beta_j, j \in J\} \) associated with a sampling set \( X \) as in Definition 4.2, we define a quasi-interpolant \( Q_X c \) on sequences by

\[ Q_X c = \sum_{j \in J} c_j \beta_j. \]

If \( f \in W^2_0 \), we write

\[ Q_X f = \sum_{j \in J} f(x_j) \beta_j \]

for the quasi-interpolant of the sequence \( c_j = f(x_j) \). We will need the following property of the quasi-interpolant \( Q_X \).

**Lemma 5.1.** Let \( X \) be any sampling set with \( \gamma \)-density \( \gamma(X) \) and let \( \{\beta_j: j \in J\} \) be a BPU associated with \( X \). If \( \psi \in W^1_0(\mathbb{R}^d) \) then for any \( f(\cdot) = \sum_k c_k \psi(\cdot - k) \), we have

\[ \| Q_X f \|_{L^2} \leq \| Q_X f \|_{W^2} \]

\[ \leq \left( (1 + 2\lceil \gamma \rceil)^d + 2 \right) \| c \|_{\ell^2} \| \psi \|_{W^1}, \quad \forall c = (c_k) \in \ell^2. \]

(5.1)

If in addition \( |\nabla \psi| \in W^1_0(\mathbb{R}^d) \), then we have

\[ \| Q_X f \|_{L^2} \leq \| Q_X f \|_{W^2} \leq \| c \|_{\ell^2} \left( \gamma (2\lceil \gamma \rceil + 1)^d \right) \| \nabla \psi \|_{W^1} + \| \psi \|_{W^1}. \]

(5.2)
To prove this Lemma, we need the following result on the oscillation operator (or the modulus of continuity operator) defined as

$$\text{osc}_\gamma(f)(x) = \sup_{|y| \leq \gamma} |f(x + y) - f(x)|.$$ 

Lemma 5.2. Let $\varphi \in W^1_0(\mathbb{R}^d)$, and let $f(\cdot) = \sum_k c_k \varphi(\cdot - k)$ where $c = (c_k) \in \ell^2$. Then:

(i) the oscillation $\text{osc}_\gamma(f)$ belongs to $W^2$;
(ii) the oscillation $\text{osc}_\gamma(\varphi)$ satisfies

$$\|\text{osc}_\gamma(\varphi)\|_{W^1} \leq \gamma \sup_{|z| \leq |y|} \|\nabla \varphi(x + z)\|,$$

and $\|\text{osc}_\gamma(\varphi)\|_{W^2} \to 0$ as $\gamma \to 0$. If in addition $|\nabla \varphi| \in W^1_0(\mathbb{R}^d)$, then

$$\|\text{osc}_\gamma(\varphi)\|_{W^1} \leq \gamma \left( 2 \gamma + 1 \right)^d \|\nabla \varphi\|_{W^1};$$

(iii) the oscillation $\text{osc}_\gamma(f)$ satisfies

$$\|\text{osc}_\gamma(f)\|_{W^2} \leq \|c\|_{\ell^2} \|\text{osc}_\gamma(\varphi)\|_{W^1}, \quad \text{for all } c \in \ell^2.$$ 

In particular, $\|\text{osc}_\gamma(f)\|_{W^2} \to 0$ as $\gamma \to 0$.

Remark 5.1. The bound on $\|Q_X f\|$ in (5.1) depends on the density $\gamma$ of the set $X$, but it does not depend explicitly on the sampling points $x_j \in X$, or the partition of unity in Definition 4.2.

Proof of Lemma 5.2. We first prove (5.4). Let $\varphi \in W^1_0$ and $|\nabla \varphi| \in W^1_0$. Then

$$|\varphi(x + y) - \varphi(x)| = \int_0^1 \nabla \varphi(x + sy) \cdot y \, ds \leq \int_0^1 |\nabla \varphi(x + sy) \cdot y| \, ds \leq |y| \sup_{|z| \leq |y|} |\nabla \varphi(x + z)|.$$

Hence we obtain

$$\sup_{x \in [0,1]^d} \text{osc}_\gamma(\varphi)(x + k) \leq \gamma \sup_{x \in [0,1]^d} \sup_{|z| \leq |y|} |\nabla \varphi(x + z + k)| \leq \gamma \sup_{x \in E} |\nabla \varphi(x + k)|,$$

where $E = [-\lceil \gamma \rceil, \lceil \gamma \rceil + 1]^d$. From this estimate we get that

$$\|\text{osc}_\gamma(\varphi)\|_{W^1} \leq \gamma \left( 2 \lceil \gamma \rceil + 1 \right)^d \|\nabla \varphi\|_{W^1}$$

which proves (5.4) of Lemma 5.2. The proof of (5.3) can be found in [1].

Let $f(\cdot) = \sum_k c_k \varphi(\cdot - k)$ where $c \in \ell^2$. Then

$$\text{osc}_\gamma(f) \leq \sum_k |c_k| \text{osc}_\gamma \varphi(\cdot - k).$$

This pointwise estimate together with (3.5) and (5.3) or (5.4) imply (5.5), and hence part (iii) as well part (i) of Lemma 5.2.

Proof of Lemma 5.1. Let $f(\cdot) = \sum c_k \varphi(\cdot - k)$ where $c \in \ell^2$. From (3.5) and property (i) of shift-invariant spaces in Section 3.1, $f \in W^2_0$ and we have

$$|f(x) - (Q_X f)(x)| = \left| f(x) - \sum_{j \in J} f(x_j) \beta_j(x) \right|$$

$$= \left| f(x) \sum_{j \in J} \beta_j(x) - \sum_{j \in J} f(x_j) \beta_j(x) \right|.$$
\[ \sum_{j \in J} |f(x) - f(x_j)| \beta_j(x) \leq \sum_{j \in J} \text{osc}_\gamma(f)(x) \beta_j(x) \]
\[ \leq \text{osc}_\gamma(f)(x) \sum_{j \in J} \beta_j(x) = \text{osc}_\gamma(f)(x). \]

From this pointwise estimate and Lemma 5.2 we get that
\[ \| f - Q_X f \|_{L^2} \leq \| \text{osc}_\gamma(f) \|_{W^2} \leq \| c \|_{L^2} \| \text{osc}_\gamma(\phi) \|_{W^1}. \] (5.6)

Using the estimate
\[ \| Q_X f \|_{L^2} \leq \| f - Q_X f \|_{L^2} + \| f \|_{L^2} \] (5.7)
and using (3.5), (5.3), (5.4), and (5.6) we obtain (5.1) and (5.2).

For the proof of Theorem 4.1 we will also need the following lemma.

**Lemma 5.3.** Let \( P \) be the orthogonal projection from \( L^2 \) onto \( V^2(\phi) \). Then there exist \( \gamma_0 > 0 \) and \( a_0 > 0 \) such that for any \( a, 0 < a \leq a_0 \), the operator \( I - PA_X \) is a contraction on \( V^2(\phi) \) for every \( \gamma \)-dense set \( X \) with \( \gamma \leq \gamma_0 \).

**Proof.** For \( f = \sum_k c_k \phi(-k) \in V^2(\phi) \) we have
\[ \| f - PA_X f \|_{L^2} = \| f - PQ_X f + PQ_X f - PA_X f \|_{L^2} \]
\[ \leq \| f - PQ_X f \|_{L^2} + \| PQ_X f - PA_X f \|_{L^2} \]
\[ \leq \| f - Q_X f \|_{L^2} + \| Q_X f - A_X f \|_{L^2}. \] (5.8)

Using (5.6) and the lower bound inequality of (3.7), the first term of the last inequality in (5.8) can be estimated as follows:
\[ \| f - Q_X f \|_{L^2} \leq \| f - Q_X f \|_{W^2} \leq \frac{1}{m} \| \text{osc}_\gamma(\phi) \|_{W^1} \| f \|_{L^2}. \] (5.9)

For the second term \( \| Q_X f - A_X f \|_{L^2} \) of (5.8) we have the pointwise estimate
\[ |(Q_X f - A_X f)(x)| = \left| \sum_j \left( f(x_j) - \langle f, \psi_{x_j} \rangle \right) \beta_j(x) \right| \]
\[ = \left| \sum_j \left( \int_{\mathbb{R}^d} \left( f(x_j) - f(\xi) \right) \psi_{x_j}(\xi) d\xi \right) \beta_j(x) \right| \]
\[ \leq \sum_j \int_{\mathbb{R}^d} |f(x_j) - f(\xi)||\psi_{x_j}(\xi)| d\xi \beta_j(x) \]
\[ \leq \sum_j \text{osc}_a(f)(x_j) \int_{\mathbb{R}^d} |\psi_{x_j}(\xi)| d\xi \beta_j(x) \]
\[ \leq M \sum_j \text{osc}_a(f)(x_j) \beta_j(x) \]
\[ \leq M \sum_j \left( \sum_k |c_k| \text{osc}_a(\phi)(x_j - k) \right) \beta_j(x). \] (5.10)

From this pointwise estimate and Lemma 5.1, it follows that:
\[ \| Q_X f - A_X f \|_{L^2} \leq M \left( (1 + 2[\gamma])^d + 2 \| c \|_{L^2} \| \text{osc}_\gamma(\phi) \|_{W^1} \right). \] (5.11)

By combining (5.8), (5.9), and (5.11), we get
\[ \| f - PA_X f \|_{L^2} \]
\[ \leq \left( \| \text{osc}_\gamma(\phi) \|_{W^1} + M \left( (1 + 2[\gamma])^d + 2 \| \text{osc}_\gamma(\phi) \|_{W^1} \right) \right) \frac{\| f \|_{L^2}}{m}. \] (5.12)
Let \( \epsilon > 0 \) be any positive real number. Using Lemma 5.2 (ii), we may choose \( \gamma_0 \) so small so that \( \| \text{osc}_\gamma(\phi) \|_{W^1} \leq \epsilon/2 \), for all \( \gamma \leq \gamma_0 \). Then, by Lemma 5.2 (ii), we may choose \( a_0 \) so small that \( M((1 + 2[\gamma_0])d + 2)\| \text{osc}_a(\phi) \|_{W^1} \leq \epsilon/2 \) for \( 0 < a \leq a_0 \). Therefore, we can choose \( \gamma_0 \) and \( a_0 \) so that for any \( \gamma \leq \gamma_0 \) and \( a \leq a_0 \), we have
\[
\| f - \mathcal{P} \mathcal{X} f \|_{L^2} \leq \frac{\epsilon}{m} \| f \|_{L^2} \quad \text{for all } f \in V^2(\phi). \tag{5.13}
\]
To get a contraction, we choose \( \epsilon/m < 1 \).

**Proof of Theorem 4.1.** Let \( e_n = f - f_n \) be the error after \( n \) iterations of algorithm (4.3). Then the sequence \( e_n \) satisfies the recursion
\[
e_{n+1} = f - f_{n+1} = f - f_n - \mathcal{P} \mathcal{X} (f - f_n) = (I - \mathcal{P} \mathcal{X})e_n. \tag{5.14}
\]
Using Lemma 5.3, we may choose \( \gamma_0 \) and \( a_0 \) so small that \( \|I - \mathcal{P} \mathcal{X}\|_{op} = \alpha < 1 \). Therefore by (5.14) we obtain
\[
\|e_{n+1}\|_{L^2} \leq \alpha \|e_n\|_{L^2} \tag{5.15}
\]
and
\[
\|e_n\|_{L^2} \leq \alpha^n \|e_0\|_{L^2}.
\]
Thus \( \|e_n\|_{L^2} \to 0 \) as \( n \to \infty \). Since for \( V^2(\phi) \) the \( W^2 \) norm and the \( L^2 \) norm are equivalent, the inequality above also holds in the \( W^2 \) norm and the proof is complete.

5.2. **Proof of Theorem 4.2**

**Proof.** By Lemma 5.3, the operator \( I - \mathcal{P} \mathcal{X} \) is a contraction. It follows that the sequence of functions \( f_n \) in (4.5) is convergent to a function \( f_\infty \). By taking the limits of both sides of (4.5), and using (4.4), we get \( \mathcal{P}(\mathcal{X} f_\infty - \mathcal{Q} \{f_j^\prime\}) = 0 \).

5.3. **Proof of Theorem 4.3**

**Proof.** Since \( \{g_{x_j} = \langle f, \psi_{x_j} \rangle \} \) satisfy (1.4), Proposition 2.1 implies that every function \( f \in V^2(\phi) \) has the expansion
\[
f(x) = \sum_{j \in J} \langle f, \psi_{x_j} \rangle \tilde{\theta}_{x_j}(x),
\]
where \( \{\tilde{\theta}_{x_j}(x)\} \) is a dual frame of \( \{\theta_{x_j} = \mathcal{P} \psi_{x_j}\} \). For any cube \( C = (r, s)^d \) of side length \( (s - r) \geq 2a + 2b \), we define the finite-dimensional subspaces
\[
V_{\phi}(C) = \text{span} \{\phi(\cdot - k): k \in [r + a + b, s - a - b]^d\},
\]
and
\[
\tilde{V}_{\phi}(C) = \text{span} \{\tilde{\theta}_{x_j}: x_j \in C\}.
\]
Expanding the basis functions \( \phi(\cdot - k) \in V_{\phi}(C) \) in terms of the frame \( \{\tilde{\theta}_{x_j}: x_j \in C\} \), and using the assumptions on the supports of \( \phi \) and \( \psi_{x_j} \), we get
\[
\phi(x - k) = \sum_{j \in \mathbb{Z}^d} \langle \phi(\cdot - k), \psi_{x_j} \rangle \tilde{\theta}_{x_j}(x)
= \sum_{x_j \in C} \langle \phi(\cdot - k), \psi_{x_j} \rangle \tilde{\theta}_{x_j}(x).
\]
It follows that \( V_{\phi}(C) \subseteq \tilde{V}_{\phi}(C) \). But
\[
\dim(V_{\phi}(C)) \geq (|s - r| - 2a - 2b)^d,
\]
and \( \dim \tilde{V}_{\phi}(C) \leq \#(C \cap X) \). Hence the theorem follows.
5.4. Proof of Theorem 4.5

Proof. The proof of this theorem is an immediate consequence of (5.12) and (5.4). □

Acknowledgments

I thank Professor Wai Shing Tang for many stimulating discussions, insightful remarks, and editorial help. I also thank Professor Gieri Simonett for helpful comments.

References