Bases and Frames for Signal Representations

Minh N. Do

1 Introduction

As with finite dimensional vector spaces, when working in an infinite dimensional vector space $\mathcal{X}$, it is often desirable to be able to expand each vector $x \in \mathcal{X}$ as a linear combination of elementary vectors $\{e_i\}$,

$$x = \sum_{i=1}^{\infty} \alpha_i e_i.$$  

(1)

Such expansion is called a linear expansion of $\mathcal{X}$. Because $\mathcal{X}$ is infinite dimensional, the right side of (1) is necessary an infinite summation, and thus has to be interpreted as a convergent series. We will consider several linear expansion systems for Hilbert spaces that are often used signal processing, including orthogonal bases, Riesz bases, and frames.

2 Orthogonal Bases

The most simple (and useful) linear expansions for Hilbert spaces are the ones that are based on orthonormal bases. Recall that a set of vectors $\{x_\alpha\}$ in an inner product space (IPS) $\mathcal{X}$ is said to be orthogonal if $x_\alpha \perp x_\beta$ whenever $\alpha \neq \beta$. In addition, if each vector in the set is normalized to have unit norm then the set is said to be orthonormal. Orthonormal sets have many nice properties that make them easy to deal with. We begin by pointing some properties for finite sets of orthonormal vectors.

Lemma 1. Let $\{e_i\}_{i=1}^n$ be an orthonormal set in an IPS $\mathcal{X}$. Then:

1. $\{e_i\}_{i=1}^n$ is linearly independent.

2. The orthogonal projection of $\mathcal{X}$ onto the subspace $\mathcal{S}$ spanned by $\{e_i\}_{i=1}^n$ is given by

$$P_\mathcal{S}x = \sum_{i=1}^{n} \langle x, e_i \rangle e_i.$$ 

3. $\|\sum_{i=1}^{n} \alpha_i e_i\|^2 = \sum_{i=1}^{n} |\alpha_i|^2$, where $\alpha_i$ are scalars.

Proof. 1. We need to show that the only solution of $\alpha_1 e_1 + \ldots + \alpha_n e_n = 0$ is $\alpha_1 = \ldots = \alpha_n = 0$. Indeed, taking inner product with both sides of the equation with $e_i$ and using the fact that $\langle e_i, e_j \rangle = \delta[i - j]$ (where $\delta[n]_{n \in \mathbb{Z}}$ denotes the Kronecker delta sequence), we have

$$0 = \langle 0, e_i \rangle = \langle \alpha_1 e_1 + \ldots + \alpha_n e_n, e_i \rangle = \alpha_i.$$
2. Since \( P_S x \in S \), we can write \( P_S x = \sum_{i=1}^{n} \alpha_i e_i \), where \( \alpha_i \) are scalars depending on \( x \). The orthogonal projection condition implies \( \langle x - P_S x, e_i \rangle = 0 \), for \( i = 1, 2, \ldots, n \). Thus we have
\[
\langle x, e_i \rangle = \langle P_S x, e_i \rangle = \sum_{j=1}^{n} \alpha_j e_j, e_i = \sum_{i=1}^{n} \alpha_j \langle e_j, e_i \rangle = \alpha_i.
\]

3. Again using \( \langle e_i, e_j \rangle = \delta[i - j] \), we have
\[
\left\| \sum_{i=1}^{n} \alpha_i e_i \right\|^2 = \left\langle \sum_{i=1}^{n} \alpha_i e_i, \sum_{i=1}^{n} \alpha_i e_i \right\rangle
= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \langle e_i, e_j \rangle
= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \delta[i - j]
= \sum_{i=1}^{n} |\alpha_i|^2.
\]

Proof. Denote \( \alpha_i = \langle x, e_i \rangle \). From Lemma 1, \( \sum_{i=1}^{n} \alpha_i e_i \) is the orthogonal projection of \( x \) onto the subspace spanned by an orthonormal set \( \{e_i\}_{i=1}^{n} \), so
\[
0 \leq \|x - \sum_{i=1}^{n} \alpha_i e_i\|^2 = \|x\|^2 - \|\sum_{i=1}^{n} \alpha_i e_i\|^2 = \|x\|^2 - \sum_{i=1}^{n} |\alpha_i|^2.
\]
Thus, for all \( n \)
\[
\sum_{i=1}^{n} |\alpha_i|^2 \leq \|x\|^2.
\]
Hence (2) also holds for infinite sum. \( \square \)

In the above lemma, we start seeing the role of \( \langle x, e_i \rangle \) as coefficients of the linear expansion by an orthonormal basis. In addition, if \( x \in S \) then \( \|x\|^2 = \sum_{i=1}^{n} |\langle x, e_i \rangle|^2 \). Now let’s turn to infinite sets of orthonormal vectors.

**Lemma 2 (Bessel’s Inequality).** Let \( \{e_i\} \) be an orthonormal sequence in an IPS \( X \). The for any \( x \in X \) one has
\[
\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leq \|x\|^2. \tag{2}
\]

Proof. Denote \( \alpha_i = \langle x, e_i \rangle \). From Lemma 1, \( \sum_{i=1}^{n} \alpha_i e_i \) is the orthogonal projection of \( x \) onto the subspace spanned by an orthonormal set \( \{e_i\}_{i=1}^{n} \), so
\[
0 \leq \|x - \sum_{i=1}^{n} \alpha_i e_i\|^2 = \|x\|^2 - \|\sum_{i=1}^{n} \alpha_i e_i\|^2 = \|x\|^2 - \sum_{i=1}^{n} |\alpha_i|^2.
\]
Thus, for all \( n \)
\[
\sum_{i=1}^{n} |\alpha_i|^2 \leq \|x\|^2.
\]
Hence (2) also holds for infinite sum. \( \square \)

Recall that for infinite dimensional Hilbert spaces the right side of (1) is an infinite summation, and thus has to be interpreted as a convergent series. We now establish a necessary and sufficient condition for an infinite series of orthonormal vectors to converge in Hilbert space.

**Theorem 1 (Riesz-Fischer).** Let \( \{e_i\} \) be an orthonormal set in a Hilbert space \( X \). A series of the form \( \sum_{i=1}^{\infty} \alpha_i e_i \) converges to a vector \( x \in X \) if and only if \( \sum_{i=1}^{\infty} |\alpha_i|^2 < \infty \) and, in that case, we have \( \alpha_i = \langle x, e_i \rangle \).
Proof. Suppose that $\sum_{n=1}^{\infty} |\alpha_n|^2 < \infty$ and define $s_n = \sum_{i=1}^{n} \alpha_i e_i$. Then,

$$\|s_n - s_m\|^2 = \left\| \sum_{i=m+1}^{n} \alpha_i e_i \right\|^2 = \sum_{i=m+1}^{n} |\alpha_i|^2,$$

which converges to 0 as $n, m \to \infty$. Therefore $\{s_n\}$ is a Cauchy sequence so by the completeness of $X$ there is a vector $x \in X$ such that $s_n \to x$. Because $\langle s_n, e_i \rangle = \alpha_i$ for all $i \leq n$, as $n \to \infty$ by the continuity of the inner product we have $\langle x, e_i \rangle = \alpha_i$.

Conversely, if $\{s_n\}$ converges, then it is a Cauchy sequence so $\sum_{i=m+1}^{n} |\alpha_i|^2 \to 0$. Thus $\sum_{i=1}^{\infty} |\alpha_i|^2$ converges.

Because of the Bessel’s inequality, for all $x \in X$, we have $\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 < \infty$. Thus from the above theorem $\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ converges to a vector $\hat{x} \in X$ such that $\langle \hat{x}, e_i \rangle = \langle x, e_i \rangle$, or $x - \hat{x} \perp e_i$ for all $i$. This leads to the following result.

**Theorem 2.** Let $\{e_i\}_{i=1}^{\infty}$ be an orthonormal set in a Hilbert space $X$. Then for every $x$ in $X$, $\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ is the orthogonal projection of $x$ onto the closure of the linear subspace $S$ generated by the $\{e_i\}_{i=1}^{\infty}$.

For finite $n$-dimensional vector spaces, a simple way of testing whether an orthonormal set (which has at most $n$ vectors because of linear independence) is a basis is to check if the number of vectors in the set equal to $n$. For infinite-dimensional vector spaces, we can no longer rely on such a counting test, and thus we need to have other mechanisms to check whether an orthonormal set is a basis. The above discussion prompts us to consider whether Bessel’s inequality becomes an equality or whether the closed subspace $S$ generated by the $e_i$’s is equal to $X$. The following theorem provide a set of equivalent criteria for an orthonormal basis, as well as its fundamental properties.

**Theorem 3 (Orthonormal Basis).** Let $\{e_i\}$ be an orthonormal set in a Hilbert space $X$. The the following statements are equivalent criteria for $\{e_i\}$ to be an orthonormal basis of $X$:

1. (Series expansion) For any $x$ in $X$ one has
   $$x = \sum_{i} \langle x, e_i \rangle e_i. \quad (3)$$

2. (Parseval equality) For any two vectors $x$ and $y$ in $X$ one has
   $$\langle x, y \rangle = \sum_{i} \langle x, e_i \rangle \overline{\langle y, e_i \rangle}. \quad (4)$$

3. For any $x$ in $X$ one has
   $$\|x\|^2 = \sum_{i} |\langle x, e_i \rangle|^2. \quad (5)$$

4. The only vector orthogonal to each of the $e_i$’s is the zero vector. In other words, $\{e_i\}$ is an maximal orthonormal set in the sense that we cannot add to it another unit vector to make a new orthonormal set.

5. The closed subspace $S$ generated by the $\{e_i\}$ is $X$. 

3
Proof. (1) ⇒ (2). Let \( x \sum_i \langle x, e_i \rangle e_i \) and \( y \sum_j \langle y, e_j \rangle e_j \). Because \( \langle e_i, e_j \rangle = \delta[i-j] \) and the continuity of the inner product, we have

\[
\langle x, y \rangle = \sum_i \langle x, e_i \rangle \sum_j \langle y, e_j \rangle \langle e_i, e_j \rangle = \sum_i \langle x, e_i \rangle \langle y, e_i \rangle \]

(2) ⇒ (3). This is obvious by taking \( x = y \).

(3) ⇒ (4). If \( \langle x, e_i \rangle = 0 \) for all \( i \) then \( \| x \|^2 = \sum_i |\langle x, e_i \rangle|^2 = 0 \), which means \( x = 0 \).

(4) ⇒ (5). Suppose that \( S \neq \mathcal{X} \), then it can be shown that there exists \( z \in \mathcal{X}, z \neq 0 \), such that \( z \perp S \), which implies \( z \) orthogonal to all \( e_i \)'s, a contradiction.

(5) ⇒ (1). If \( S = \mathcal{X} \) then the orthogonal projection onto \( S \) is the identity map and thus by Theorem 2 we have

\[
x = \sum_i \langle x, e_i \rangle e_i
\]

for every \( x \) in \( \mathcal{X} \).

The series expansion (3) is sometime referred to as (generalized) Fourier series in Hilbert space since it generalizes the notion of Fourier series (see Example 1). In that case the coefficients \( \langle x, e_i \rangle \) are called the Fourier coefficients of \( x \) with respect to the \( e_i \). The term basis for \( \{e_i\} \) is justified as of the uniqueness of its series expansion is demonstrated in the following result.

**Corollary 1.** The series expansion by an orthonormal basis \( \{e_i\} \) is unique for each \( x \) in \( \mathcal{X} \).

**Proof.** Suppose that \( x = \sum_i \alpha_i e_i = \sum_i \beta_i e_i \). Then \( \sum_i (\alpha_i - \beta_i) e_i = 0 \). Hence \( \sum_i |\alpha_i - \beta_i|^2 = \|0\|^2 = 0 \), which implies \( \alpha_i = \beta_i \) for all \( i \).

**Remark 1.** Previously, we talked about Hamel bases, which involves only finite linear combinations. In infinite dimensional Hilbert space, Hamel bases are rarely used, while orthogonal bases are extremely useful.

**Remark 2.** So far, we have considered countable orthonormal bases. Hilbert spaces pose such an orthonormal basis is called separable. Same results can be extended to uncountable bases and nonseparable Hilbert spaces (see Naylor and Sell [6]).

**Example 1** (Fourier bases). The most famous orthonormal basis is the Fourier basis. The Fourier theory states that any square-integrable function defined on an interval \([0, T]\), i.e. \( f(t) \in L^2[0, T] \), can be expressed as a linear combination of complex exponentials with frequencies \( k\omega_0 \), where \( \omega_0 = 2\pi/T \)

\[
f(t) = \sum_{k=-\infty}^{\infty} F[k] e^{jk\omega_0 t}, \tag{6}
\]

with

\[
F[k] = \frac{1}{T} \int_0^T f(t) e^{-jk\omega_0 t} dt. \tag{7}
\]
For \( f(t) \in L^2[0,T] \), then the series on the left side of (6) converges to \( f(t) \) in the \( L^2 \) sense; that is, if we denote \( \hat{f}_N(t) \) the truncated series with \( k \) going from \(-N\) to \( N \), then the error \( \|f(t) - \hat{f}_N(t)\|_{L^2[0,T]} \) goes to zeros at \( N \to \infty \).

Now consider \( g_k(t) = \frac{1}{\sqrt{T}} e^{j\omega_0 t} \). Then it is easy to verify that \( \langle g_k, g_l \rangle_{L^2[0,T]} = \delta[k - l] \), or \( \{g_k(t)\}_{k \in \mathbb{Z}} \) is an orthonormal set. Combining (6) and (7), for any \( f(t) \in L^2[0,T] \) we have

\[
f = \sum_{k \in \mathbb{Z}} \langle f, g_k \rangle_{L^2[0,T]} g_k.
\]

In other words, \( \{g_k(t)\}_{k \in \mathbb{Z}} \) is an orthonormal basis for \( L^2[0,T] \). From this, using Theorem 3 we immediately have the Parseval relation and best approxim ation property for the Fourier series.

**Example 2** (Sinc bases). Using the language of Hilbert spaces, the Shannon sampling theorem simply states that the normalized sinc function

\[
\tilde{sinc}_T(t) = \frac{1}{\sqrt{T}} \frac{\sin(\pi t/T)}{\pi t/T}
\]

together with its translates

\[
\{\tilde{sinc}_T(t - nT)\}_{n \in \mathbb{Z}}
\]

provides an orthonormal basis for the space of bandlimited functions

\[
BL[-\pi/T,\pi/T] \overset{\text{def}}{=} \{f \in L^2(\mathbb{R}) \mid F(\omega) = 0, \forall |\omega| > \pi/T\}.
\]

This means for each \( f(t) \in BL[-\pi/T,\pi/T] \), we have

\[
f(t) = \sum_{n=\infty}^{\infty} \langle f, \tilde{sinc}_T(\cdot - nT) \rangle \tilde{sinc}_T(t - nT).
\]

Furthermore,

\[
\langle f, \tilde{sinc}_T(\cdot - nT) \rangle = \sqrt{T} f(nT).
\]

The proof for this statement is left as an exercise. Such view of the Shannon sampling theorem allows us to extend the classical sampling theory to more general function spaces and more general sampling kernels [7] (also see Section 4.6)

**Example 3** (Wavelet bases). In 1910 Haar [3] showed that the following function

\[
\psi(t) = \begin{cases} 
1 & \text{if } 0 \leq t < 1/2 \\
-1 & \text{if } 1/2 \leq t < 1 \\
0 & \text{otherwise}
\end{cases} \quad (8)
\]

together with its scaled and translated versions on dyadic scales

\[
\psi_{j,n} = 2^{-j/2} \psi \left( \frac{t - n2^j}{2^j} \right), \quad j, n \in \mathbb{Z} \quad (9)
\]

provide an orthogonal basis for functions in \( L^2(\mathbb{R}) \). This turns out to be the first and simplest wavelet basis. Recently, wavelets had have a growing impact in both theory and practice in signal
processing. The study and construction of wavelet bases are later greatly simplified with the introduction of the multiresolution analysis framework [5]. But for now, using the above framework, we will directly demonstrate that with the Haar wavelet function defined in (8), \( \{ \psi_{j,n} \}_{j,n \in \mathbb{Z}} \) is indeed an orthogonal basis for \( L^2(\mathbb{R}) \).

First, we need to show that \( \langle \psi_{j,n}, \psi_{j',n'} \rangle = \delta[j - j'] \delta[n - n'] \). Indeed, if \( j = j' \) then these two wavelet functions do not overlap unless \( n = n' \). If \( j \neq j' \) and suppose that \( j > j' \), then the support of \( \psi_{j',n'} \) would either have none overlap with \( \psi_{j,n} \) or entirely in a constant half of \( \psi_{j,n} \), and their inner product is equal to 0.

Next, we need to show that \( \{ \psi_{j,n} \}_{j,n \in \mathbb{Z}} \) is complete in \( L^2(\mathbb{R}) \) by showing that if \( f(t) \in L^2(\mathbb{R}) \) and \( \langle f, \psi_{j,n} \rangle = 0 \) for all \( j,n \in \mathbb{Z} \), then \( f(t) = 0 \). Let \( I_{j,n} \) denote the dyadic interval \([2^j n, 2^j (n+1)]\). Then if \( \langle f, \psi_{j,n} \rangle = 0 \) means \( \int_{I_{j,n}} f(t) dt = 0 \). Therefore if \( \langle f, \psi_{j,n} \rangle = 0 \) for all \( j,n \in \mathbb{Z} \), then \( \int_{I_{j,n}} f(t) dt = c \) for all \( j,n \in \mathbb{Z} \). From this we see that \( f(t) = c \) almost everywhere. But \( f(t) \in L^2(\mathbb{R}) \) so we must have \( c = 0 \).

3 Riesz and Biorthogonal Bases

TO DO: Riesz bases, splines, and shift-invariant subspaces.

4 Frames

Frames have recently became a useful tool in signal processing and communications. There are two ways to view frames. First, they can be considered as a generalization of linear inverse problem in \( \mathbb{C}^n \) to general Hilbert spaces. Second, they can be considered as generalization of bases that to overcomplete linear expansions.

4.1 Frames for Linear Inverse Problems in Hilbert Spaces

Consider the following linear inverse problem:

\[
Ax = b
\]

where \( A \in \mathbb{C}^{m \times n} \) is fixed, \( b \in \mathbb{C}^m \) is given, and \( x \in \mathbb{C}^n \) is unknown. The problem is to find \( x \) from a given \( b \). In many cases, instead of \( b \) we only have access to its noisy version \( \hat{b} = b + \eta \). Examples of such problem include deconvolution, tomography, transform coding, and communications. In these cases, the possible noise is due to model mismatch, measurement error, transmission error, quantization, and/or thresholding.

To address this linear inverse problem, we face with the following two basic questions:

1. Can we reconstruct \( x \) in a numerically stable way from \( b \) (or \( \hat{b} \))?

2. Which is the “optimal” reconstruction algorithm in the presence of noise?

For the first question, we know that the answer depends on the condition number of \( A \), which is defined as the ratio between its largest and smallest singular values:

\[
\kappa(A) = \frac{\sigma_1(A)}{\sigma_n(A)}.
\]
Figure 1: System diagram of the common linear inverse problem. The problem is to recover \( \mathbf{x} \) given a noisy \( \hat{\mathbf{b}} \).

Since \( \sigma_1 \) and \( \sigma_n \) are the largest and smallest singular values of \( \mathbf{A} \), we have for all \( \mathbf{x} \):

\[
\sigma_n \| \mathbf{x} \|_2 \leq \| \mathbf{A} \mathbf{x} \|_2 \leq \sigma_1 \| \mathbf{x} \|_2,
\]

and the bounds are tight. Thus, intuitively the ratio \( \kappa(\mathbf{A}) = \frac{\sigma_1(\mathbf{A})}{\sigma_n(\mathbf{A})} \) represents the dynamic range of the gains by \( \mathbf{A} \). Thus the smaller this ratio (\( \kappa(\mathbf{A}) \geq 1 \)), the more stable (or well-conditioned) the problem is.

For the second question, when \( \mathbf{A} \) is a full column rank matrix, among all (typically infinite) linear inverses that reconstruct \( \mathbf{x} \) from \( \mathbf{b} \), the pseudo-inverse \( \mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \) that produces \( \hat{\mathbf{x}} = \mathbf{A}^\dagger \mathbf{b} \) is the “optimal” linear reconstruction in certain senses. It can shown that among all left inverses of \( \mathbf{A} \), \( \mathbf{A}^\dagger \) has the minimum spectral norm, as well as the minimum Frobenius norm.

Let’s denote \( a_i^H \) as the rows of \( \mathbf{A} \), then \( \mathbf{A} \) can be considered as a linear operator, \( \mathbf{A} : \mathbb{C}^n \to \mathbb{C}^m \), with

\[
\mathbf{b} = \mathbf{A} \mathbf{x} \iff \mathbf{b}_i = \langle \mathbf{x}, a_i \rangle_{\mathbb{C}^n}, \quad i = 1, 2, \ldots, m.
\]

Thus generalizing the condition (10) to general Hilbert spaces lead us to the following definition.

**Definition 1 (Frame).** A sequence \( \{ \phi_k \} \in \mathcal{H} \) in a Hilbert space \( \mathcal{H} \) is a frame if there exist two constants \( \alpha > 0 \) and \( \beta < \infty \) such that for any \( \mathbf{x} \in \mathcal{H} \)

\[
\alpha \| \mathbf{x} \|^2 \leq \sum_{k \in \Gamma} |\langle \mathbf{x}, \phi_k \rangle|^2 \leq \beta \| \mathbf{x} \|^2.
\]

where \( \alpha \) and \( \beta \) are called the frame bounds. When \( \alpha = \beta \) the frame is said to be tight. Associated with a frame is the frame operator \( \mathbf{A} \), defined as the linear operator from \( \mathcal{H} \) to \( l^2(\Gamma) \) as

\[
(\mathbf{A} \mathbf{x})_k = \langle \mathbf{x}, \phi_k \rangle, \quad \text{for } k \in \Gamma.
\]

There is a clear connection between (10) and (11). From this connection we the significance of the frame definition in context of a linear inverse problem: \( \{ \phi_k \} \in \mathcal{H} \) is a frame means that one can recover \( \mathbf{x} \in \mathcal{H} \) from \( \{ \langle \mathbf{x}, \phi_k \rangle \} \in l^2(\Gamma) \) in a numerically stable way. The frame bound ratio \( \beta/\alpha \) indicates how tight the frame \( \{ \phi_k \} \in \mathcal{H} \) is: the closer \( \beta/\alpha \) to 1, the tighter the frame is and the more well-conditioned the associated inverse problem is.
4.2 Frames as a Generalization of Bases

Recall that if \( \{\phi_k\} \) is an orthonormal basis of \( \mathcal{H} \), then from the Parseval equality, for any \( x \in \mathcal{H} \)

\[
\|x\|^2 = \sum_k |\langle x, \phi_k \rangle|^2.
\]

Thus any orthonormal basis \( \{\phi_k\} \) is a tight frame. However, a tight frame is not necessarily an orthonormal basis due to the lack of the orthogonal assumption.

Example 4 (Gabor frames). A Gabor frame is a frame for \( L^2(\mathbb{R}) \) of the form

\[
\left\{ e^{2\pi j m t} g(t - na) \right\}_{m,n \in \mathbb{Z}},
\]

where \( a,b > 0 \) and \( g \in L^2(\mathbb{R}) \) is a fixed function. Computing the frame coefficients with the frame (13) amounts to taking the short-time (or windowed) Fourier transform with the sliding window \( g(t - na) \):

\[
c_{m,n} = \langle f(t), e^{2\pi j m t} g(t - na) \rangle = \int_{\mathbb{R}} (f(t) g^*(t - na)) e^{-2\pi j m t} dt.
\]

The frame coefficients \( \{c_{m,n}\}_{m,n \in \mathbb{Z}} \) provide a local time-frequency analysis of the function \( f \). In particular, when

\[
g(t) = \begin{cases} 
  \frac{1}{\sqrt{T}}, & \text{for } 0 \leq t \leq T \\
  0, & \text{otherwise},
\end{cases}
\]

and \( a = T, \ b = 1/T \), then (14) is simply the block Fourier series coefficients. In this case, from Example 1 we see that the set in (13) is a collection of orthonormal bases for \( L^2 \) functions on intervals \( [nT, (n + 1)T] \), \( n \in \mathbb{Z} \), and thus together they form an orthonormal basis for \( L^2(\mathbb{R}) \). More useful Gabor systems are obtained when \( \{g(t - na)\}_{n \in \mathbb{Z}} \) are smooth and overlapping windows to avoid the blocking artifact.

The product \( ab \) represents the sampling density of the Gabor system (13). We already see that for \( ab = 1 \), we can have a basis. For \( ab < 1 \), we would have redundant frame. A motivation for studying (redundant) Gabor frames is due to the Balian-Low Theorem which states that a function \( g \) generating a Gabor basis cannot be well localized in both time and frequency.

The frame questions come up when we want to know under what condition on the sampling intervals \( a \) and \( b \) for time and frequency, respectively, and the window function \( g \), we can reconstruct in a numerically stable way the function \( f \) from its short-time Fourier coefficients \( \{c_{m,n}\}_{m,n \in \mathbb{Z}} \).

4.3 Pseudo Inverse and Dual Frame

Suppose that \( \{\phi_k\}_{k \in \Gamma} \) is a frame in a Hilbert space \( \mathcal{H} \). The associated frame operator \( \mathbf{A} : \mathcal{H} \to l^2(\Gamma) \) is defined as in (12). It is easy to see that the adjoint of \( \mathbf{A} \) is

\[
\mathbf{A}^* \mathbf{c} = \sum_{k \in \Gamma} c_k \phi_k, \quad \text{for } \mathbf{c} \in l^2(\Gamma)
\]

(15)

The self-adjoint operator \(^1\mathbf{A}^* \mathbf{A} : \mathcal{H} \to \mathcal{H} \) play an important role in the frame theory.

\[
\mathbf{A}^* \mathbf{A} \mathbf{x} = \sum_{k \in \Gamma} \langle x, \phi_k \rangle \phi_k.
\]

\(^1\)Some authors use the term frame operator for \( \mathbf{A}^* \mathbf{A} \).
By noting that,
\[
\langle A^* A x, x \rangle = \sum_{k \in \Gamma} |\langle x, \phi_k \rangle|^2,
\]
we can write the frame condition (11) \emph{equivalently} in operator notation as
\[
\alpha I \leq A^* A \leq \beta I,
\]
where \( I \) is the identity operator on \( \mathcal{H} \). The partial ordering of self-adjoint operators on \( \mathcal{H} \) is defined as
\[
U_1 \leq U_2 \iff \langle U_1 x, x \rangle \leq \langle U_2 x, x \rangle, \forall x \in \mathcal{H}.
\]
Using this order, we can work with self-adjoint operators almost as with real numbers. In particular, from (18) we known that \( A^* A \) is invertible and
\[
\beta^{-1} I \leq (A^* A)^{-1} \leq \alpha^{-1} I.
\]
Thus, as in the matrix case, we can define the \emph{pseudo inverse} operator as
\[
A^\dagger = (A^* A)^{-1} A^*,
\]
which is known to be exist and bounded because the frame condition. Reconstruction \( x \) from frame coefficients \( \{\langle x, \phi_k \rangle\}_{k \in \Gamma} \) using the pseudo inverse can be written as
\[
x = A^\dagger A x = (A^* A)^{-1} A^* A x = \sum_{k \in \Gamma} \langle x, \phi_k \rangle (A^* A)^{-1} \phi_k.
\]
This leads us to the following definition of the \emph{dual frame} of \( \{\phi_k\}_{k \in \Gamma} \)
\[
\tilde{\phi}_k = (A^* A)^{-1} \phi_k.
\]
Because of (19), we see that \( \{\tilde{\phi}_k\}_{k \in \Gamma} \) is also a frame of \( \mathcal{H} \) with frame bounds \( \beta^{-1} \) and \( \alpha^{-1} \). Two frames \( \{\phi_k\}_{k \in \Gamma} \) and \( \{\tilde{\phi}_k\}_{k \in \Gamma} \) have the dual role as it can be easy to check that
\[
x = \sum_{k \in \Gamma} \langle x, \phi_k \rangle \tilde{\phi}_k = \sum_{k \in \Gamma} \langle x, \tilde{\phi}_k \rangle \phi_k, \quad \forall x \in \mathcal{H}.
\]
In the case of tight frame (i.e. when \( \alpha = \beta \)), the dual frame is simply \( \tilde{\phi}_k = \alpha^{-1} \phi_k \).

**Example 5.** For a Gabor frame in Example 4 one can show that (exercise) its dual frame also has the form
\[
\left\{ e^{2\pi j m b t} \tilde{g}(t - na) \right\}_{m,n \in \mathbb{Z}}.
\]
Thus we only need to compute a single function \( \tilde{g} \) for the dual frame reconstruction (22).
4.4 Frame Reconstruction Algorithm

We see from the previous subsection that both the pseudo inverse and dual frame computations require the inversion of $A^*A$ given in (16). This is not trivial in general, especially in infinite-dimensional cases. A particular useful result of frame theory provides iterative reconstruction algorithms for general frames.

Consider $R = I - \frac{2}{\alpha + \beta} A^* A$. Because of (18), we have

$$\frac{\alpha - \beta}{\alpha + \beta} I \leq R \leq \frac{\beta - \alpha}{\alpha + \beta} I.$$ 

Therefore,

$$\|R\| \leq \frac{\beta - \alpha}{\beta + \alpha} < 1.$$

Thus, we have

$$(A^*A)^{-1} = \frac{2}{\alpha + \beta}(I - R)^{-1} = \frac{2}{\alpha + \beta} \sum_{i=0}^{\infty} R^i$$

From this, we can derive (exercise) the following iterative reconstruction formula (also known as the frame algorithm).

$$x_n = x_{n-1} + \frac{2}{\alpha + \beta} \sum_{k \in \Gamma} ((x, \phi_k) - (x_{n-1}, \phi_k)) \phi_k$$

This algorithm requires the knowledge of frame bounds $\alpha, \beta$. Faster convergence is obtained by the conjugate gradient method [2], which works without knowledge of frame bounds.

4.5 Frame Reconstruction in Present of Noise

When there is additive noise in the frame coefficients, the pseudo inverse eliminates the influence of errors that are orthogonal to the range of the frame operator. Therefore, if we have access to $\hat{y} = y + \eta$ instead of $y = Ax$, then the pseudo inverse provides the solution $\hat{x} = A^\dagger \hat{y}$ that minimizes the residual $\|A\hat{x} - \hat{y}\|$. This is called the least-squares solution. For a tight frame, the pseudo inverse is simply the scaled transposed matrix of the frame operator, since $A^T A = \alpha \cdot I$.

We will now review results that allow us to quantify the performance of a left inverse. It can be shown [5] that the pseudo inverse has minimum sup norm among all the left inverses of the frame operator. Let $S$ be an arbitrary left inverse of $A$. The norm of an operator $S$ is defined as

$$\|S\| = \sup_{y \neq 0} \frac{\|Sy\|}{\|y\|},$$

and for a matrix, this is the spectral norm can be computed by [4]

$$\|S\| = \max\{ \sqrt{\lambda} : \lambda \text{ is an eigenvalue of } SS^T \}.$$ (24)

The influence of the operator norm in the reconstruction can be seen in the following. With the noise model setup as above, the reconstruction error by $S$ is

$$\epsilon \overset{\text{def}}{=} \hat{x} - x = S(y + \eta) - x = S(Ax + \eta) - x = S\eta.$$ (25)
Therefore, \[ \|\epsilon\| \leq \|S\| \|\eta\|. \] (26)

In other words, when the energy of the noise \( \eta \) is bounded, the sup norm of the inverse matrix provides an upper bound for the reconstruction error, and this bound is tight.

In some cases, we can assume that the additive noise \( \eta \) is white and zero-mean, which means that
\[
E\{\eta[i]\} = 0, \quad \text{and} \quad E\{\eta[i], \eta[j]\} = \delta[i-j]\sigma^2, \quad \text{for all } i, j,
\] (27)
or its auto-correlation matrix \( R_\eta \overset{\text{def}}{=} E\{\eta\eta^T\} = \sigma^2 \cdot I \). This noise model is approximately true when, for instance, \( y \) is uniformly scalar quantized. In this case, the auto-correlation of the reconstruction error by \( S \) is
\[
R_\epsilon \overset{\text{def}}{=} E\{\epsilon\epsilon^T\} = E\{S\eta\eta^T S^T\} = S R_\eta S^T = \sigma^2 SS^T.
\]

Hence, for signals of finite length \( N \), the reconstruction mean squared error (MSE) is \[ MSE \overset{\text{def}}{=} \frac{E\{\|\epsilon\|^2\}}{N} = \frac{\text{tr}(R_\epsilon)}{N} = \frac{\sigma^2 \text{tr}(SS^T)}{N} = \frac{\sigma^2 \|S\|^2_F}{N}. \] (28)

We already proved that among all the left inverses of a matrix \( A \), its pseudo-inverse \( A^\dagger \) has minimum spectral norm as well as Frobenius norm. Using the orthogonal projection property of the pseudo inverse, the minimum operator norm property can also be proved for frame operators in general Hilbert spaces. From this result we see that the pseudo inverse provides the optimal linear reconstruction in the present of noise.

**Example 6.** To get a gist of the aforementioned properties of frames, consider the following illustrative example. Consider a redundant transform that takes a scalar \( x \in \mathbb{R} \) and outputs a vector \( y = (y_1, y_2)^T \in \mathbb{R}^2 \) such that \( y_i = x, \ i = 1, 2 \). There are infinite ways to reconstruct \( x \) from \( y \): one simple way is to assign \( \hat{x}_1 = y_1 \), and another way is to compute \( \hat{x}_2 = (y_1 + y_2)/2 \). Under the white noise model given in (27), the performance by these two reconstruction methods can be quantified as: \( MSE_1 = E\{\|x - \hat{x}_1\|^2\} = E\{\eta_1^2\} = \sigma^2 \), and \( MSE_2 = E\{\|x - \hat{x}_2\|^2\} = \frac{1}{4} E\{(\eta_1 + \eta_2)^2\} = \sigma^2/2 \).

Thus, we reduce the MSE by half by using the second reconstruction method instead of the first one. In fact, the second reconstruction method is the pseudo inverse, which minimizes the MSE in this case.

### 4.6 Application to Generalized Sampling

As we see in Example 2, the classical Shannon sampling theory can be viewed as an orthonormal basis expansion using the \( \text{sinc} \) bases. However, this theory restricts the class of functions that can be sampled to be bandlimited, and it uses the impractical \( \text{sinc} \) kernel.
Let’s consider a generalized sampling framework in which the sampled data of an unknown function \( x(t) \in L^2(\mathbb{R}) \) are given as
\[
s_k = \langle x, \phi_k \rangle, \quad \text{for } k \in \Gamma,
\] (29)
where \( \phi_k \) is the point spreading function (PSF) of the \( k \)-th sensing device (see Figure 4.6). From the Riesz representation theorem, we know that any linear sensing device can be modeled as in (29). This framework allows for nonuniform sampling locations and nonuniform PSF’s.

A general “sampling theorem” can be stated as: any function \( x \in \mathcal{H} \) can be recovered in a numerically stable way from samples \( s_k \) if and only if \( \{\phi_k\}_{k \in \Gamma} \) is a frame of \( \mathcal{H} \). In that case, the frame reconstruction algorithms provide practical methods to recover the sampled function \( x \).

References


