5.1 Introduction

Over the next several chapters, we embark on the construction of various bases for
signal analysis and synthesis. To create representations

\[ x = \sum_k \alpha_k \varphi_k, \]

we need \{\varphi_k\} to “fill the space” (completeness), and sometimes for \{\varphi_k\} to be an
orthonormal set. However, these properties are not enough for a representation to
be useful. For most applications, the utility of a representation is tied to time,
frequency, scale, and resolution properties of the \varphi_k s—the topic of this chapter.

Time and frequency properties are perhaps the most intuitive. Think of a basis
function as a note on a musical score. It has a given frequency (for example, middle
A has the frequency of 440Hz) and start time, and its type (♩, ♪, ♩) indicates its duration. We can think of the musical score as a time-frequency plane\(^{34}\) and notes as rectangles in that plane with horizontal extent determined by start and end times and vertical extent related in some way to frequency. Mapping a musical score to the time-frequency plane is far from the artistic view usually associated with music, but it helps visualize the notion of the time-frequency plane (see Figure 5.1).

Time and frequency views of a signal are intertwined in several ways. The duality of the Fourier transform gives a precise sense of interchangeability, since if \(x(t)\) has the transform \(X(\omega)\), \(X(t)\) has the transform \(2\pi x(-\omega)\). More relevant to this chapter, the uncertainty principle determines the trade-off between concentration in one domain and spread in the other; signals concentrated in time will be spread in frequency, and, by duality, signals concentrated in frequency will be spread in time. It also bounds the product of spreads in time and frequency, with the lower bound reached by Gaussian functions. This cornerstone result of Fourier theory is due to Heisenberg in physics and associated with Gabor in signal theory (see Historical Remarks).

Another natural notion for signals is scale. For example, given a portrait of a person, recognizing that person should not depend on whether they occupy one-tenth or one-half of the image.\(^{35}\) Thus, image recognition should be scale invariant. Signals that are scales of each other are often considered as equivalent. However, this scale invariance is a purely continuous-time property, since discrete-time sequences cannot be rescaled easily. In particular, while continuous-time rescaling can be undone easily, discrete-time rescaling, such as downsampling by a factor of 2, cannot be undone in general.

A fourth important notion is that of resolution. Intuitively, a blurred photograph does not have the resolution of a sharp one, even when the two prints are of the same physical size. Thus, resolution is related to bandwidth of a signal, or, more generally, to the number of degrees of freedom per unit time (or space). Clearly, scale and resolution interact, and most notably in discrete time.

### 5.2 Time and Frequency Localization

#### Time Spread

Consider a function \(x(t)\) or a sequence \(x_n\), where \(t\) or \(n\) is a time index. We now discuss localization of the function or sequence in time. The easiest case is when the support of the signal is finite, that is, \(x(t)\) is nonzero only in \([T_1, T_2]\), or, \(x_n\) is nonzero only for \(\{N_1, N_1 + 1, \ldots, N_2\}\). This is often called compact support. If a function (sequence) is of compact support, then its Fourier transform cannot be of compact

\(^{34}\)Note that the frequency axis is logarithmic rather than linear.

\(^{35}\)This is true within some bounds, which is linked to resolution; see below.
support (it can only have isolated zeros). That is, a function (sequence) cannot be perfectly localized in both time and frequency. This fundamental property of the Fourier transform is explored in Exercise 5.1 for the DFT.

If not of compact support, a signal might decay rapidly as \( t \) or \( n \) go to \( \pm \infty \). Such decay is necessary for working in \( L^2(\mathbb{R}) \) or \( \ell^2(\mathbb{Z}) \); for example, for finite energy \( (L^2(\mathbb{R})) \), a function must decay faster than \( |t|^{-1/2} \) for large \( t \) (see Chapter 3).

A concise way to describe locality (or lack thereof), is to introduce a spreading measure akin to standard deviation, which requires normalization so that \( |x(t)|^2 \) can be interpreted as a probability density function.\(^{36}\) Its mean is then the time center of the function and its standard deviation is the time spread. Denote the energy of \( x(t) \) by

\[
E_x = \| x \|^2 = \int_{-\infty}^{\infty} |x(t)|^2 \, dt.
\]

Then define the time center \( \mu_t \) as

\[
\mu_t = \frac{1}{E_x} \int_{-\infty}^{\infty} t |x(t)|^2 \, dt,
\]

and the time spread \( \Delta_t \) as

\[
\Delta_t^2 = \frac{1}{E_x} \int_{-\infty}^{\infty} (t - \mu_t)^2 |x(t)|^2 \, dt.
\]

Example 5.1 (Time spreads). Consider the following signals and their time spreads:

(i) The box function

\[
b(t) = \begin{cases} 
1, & -1/2 < t \leq 1/2; \\
0, & \text{otherwise};
\end{cases}
\]

has \( \mu_t = 0 \) and \( \Delta_t^2 = 1/12 \).

(ii) For the sinc function from (1.65),

\[
x(t) = \frac{1}{\sqrt{\pi}} (t \sin \pi) = \frac{1}{\sqrt{\pi}} \frac{\sin t}{t},
\]

\( |x(t)|^2 \) decays only as \( 1/|t|^2 \), so \( \Delta_t^2 \) is infinite.

(iii) The Gaussian function as in (1.69) with \( \gamma = (2\alpha/\pi)^{1/4} \) is of unit energy as we have seen in (1.71). Then, the version centered at \( t = 0 \) is of the form:

\[
g(t) = \left( \frac{2\alpha}{\pi} \right)^{1/4} e^{-\alpha t^2},
\]

and has \( \mu_t = 0 \) and \( \Delta_t^2 = 1 \).

\(^{36}\)Note that this normalization is precisely the same as restricting attention to unit-energy functions.
From the above example, we see that the time spread can vary widely. In particular, it can be unbounded, even for functions which are widely used, like the sinc function. Exercise 5.2 explores functions based on the box function and convolutions thereof.

We restrict our definition of the time spread to continuous-time functions for now. While it can be extended to sequences, it will not be done here since its frequency-domain equivalent (where functions are periodic) is not easily generalizable.

Exercise 5.3 explores some of the properties of the time spread $\Delta_t$: (a) it is invariant to time shifts $x(t-t_0)$; (b) it is invariant to modulations by an exponential $e^{j\omega_0 t}x(t)$; (c) energy-conserving scaling by $s^2$ increases $\Delta_t$ by a factor $s$.

**Frequency Spread**

We now discuss the dual concept to time localization of $x(t)$—frequency localization of its DTFT pair $X(\omega)$. In terms of localization, the simplest case is when $X(\omega)$ is of compact support, or, $|X(\omega)| = 0$ for $\omega \notin [-\Omega, \Omega]$. This is the notion of bandlimitidness used in the sampling theorem in Chapter 4. The time domain function has infinite support.

Even if not compactly supported, the Fourier transform can have fast decay, and thus the notion of a frequency spread. Similarly to our discussion of the time spread, we normalize the frequency spread, so as to be able to interpret $|X(\omega)|^2$ as a probability density function. Denote the energy of $X(\omega)$ by

$$E_\omega = ||x||^2 = \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega.$$  

The frequency center $\mu_\omega$ is then

$$\mu_\omega = \frac{1}{2\pi E_\omega} \int_{-\infty}^{\infty} \omega |X(\omega)|^2 d\omega, \quad (5.5)$$

and the frequency spread $\Delta_\omega$

$$\Delta_\omega^2 = \frac{1}{2\pi E_\omega} \int_{-\infty}^{\infty} (\omega - \mu_\omega)^2 |X(\omega)|^2 d\omega. \quad (5.6)$$

Because of the $2\pi$-periodicity of the Fourier transform $X(e^{j\omega})$, there is no natural definition of $\Delta_\omega^2$ for sequences. Thus, we continue to focus on continuous-time functions for now.

**Example 5.2.** Consider a bandlimited Fourier transform given by

$$X(\omega) = \begin{cases} \sqrt{\pi}, & |\omega| < \frac{\pi}{s}, \\ 0, & \text{otherwise}. \end{cases}$$

Then

$$\Delta_\omega^2 = \frac{1}{2\pi} \int_{-\pi/s}^{\pi/s} \omega^2 d\omega = \frac{\pi^2}{3s^2}.$$
5.3. Heisenberg Boxes and the Uncertainty Principle

Figure 5.2: The time-frequency plane and a Heisenberg box. (a) A time-domain function \( x(t) \). (b) Its Fourier transform magnitude \( |X(\omega)| \). (c) A Heisenberg box centered at \((\mu_t, \mu_\omega)\) and of size \((\Delta_t, \Delta_\omega)\).

Figure 5.3: A time and/or frequency shift of a function, simply shifts the Heisenberg box.

The corresponding time-domain function is a sinc, for which we have seen that \( \Delta_t^2 \) is unbounded.

Exercise 5.3 explores some of the properties of the frequency spread \( \Delta_\omega \): (a) it is invariant to time shifts \( x(t - t_0) \); (b) it is invariant to modulations by an exponential \( e^{j\omega_0 t}x(t) \) (but not to modulations by a real function such as a cosine); (c) energy-conserving scaling by \( s \), \( 1/\sqrt{s}x(t/s) \), decreases \( \Delta_\omega \) by a factor \( s \). This is exactly the inverse of the effect on the time spread, as is to be expected from the scaling property of the Fourier transform.

5.3 Heisenberg Boxes and the Uncertainty Principle

Given a function \( x(t) \) and its Fourier transform \( X(\omega) \), we have the 4-tuple \((\mu_t, \Delta_t, \mu_\omega, \Delta_\omega)\) describing the function’s center in time and frequency \((\mu_t, \mu_\omega)\) and the function’s spread in time and frequency \((\Delta_t, \Delta_\omega)\). It is convenient to show this in a diagram, as in Figure 5.2. This gives a conceptual picture, conveying the idea that there is a center of mass \((\mu_t, \mu_\omega)\) and a spread \((\Delta_t, \Delta_\omega)\) shown by a rectangular box with appropriate location and size. The plane on which this is drawn is called the time-frequency plane, and the box is usually called a Heisenberg box, or a time-frequency tile.

From our previous discussion on time shifting and complex modulation, we know that a function \( y(t) \) obtained by shift and modulation of \( x(t) \),

\[
y(t) = e^{j\omega_0 t}x(t - t_0),
\]

will simply have a Heisenberg box shifted by \( t_0 \) and \( \omega_0 \), as depicted in Figure 5.3.

What about scaling? Consider energy-invariant rescaling by a factor \( s \),

\[
y(t) = \frac{1}{\sqrt{s}} x\left(\frac{t}{s}\right).
\]

Figure 5.4: The effect of scaling as in (5.8) on the associated Heisenberg box. The case \( s = 2 \) is shown.
If \( x(t) \) has a Heisenberg box specified by \((\mu_t, \Delta t, \mu_\omega, \Delta_\omega)\), then the box for \( y(t) \) is specified by \((s\mu_t/\sqrt{s}, s\mu_\omega/\sqrt{s}, s\Delta t, s\Delta_\omega)\), as shown in Figure 5.4. The effects of shift, modulation and scaling on the Heisenberg boxes of the resulting functions are summarized in Table 5.1.

<table>
<thead>
<tr>
<th>Function</th>
<th>Time center</th>
<th>Time spread</th>
<th>Freq. center</th>
<th>Freq. spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x(t) )</td>
<td>( \mu_t )</td>
<td>( \Delta t )</td>
<td>( \mu_\omega )</td>
<td>( \Delta_\omega )</td>
</tr>
<tr>
<td>( x(t - t_0) )</td>
<td>( \mu_t + t_0 )</td>
<td>( \Delta t )</td>
<td>( \mu_\omega )</td>
<td>( \Delta_\omega )</td>
</tr>
<tr>
<td>( e^{j\omega_0 t} x(t) )</td>
<td>( \mu_t )</td>
<td>( \Delta t )</td>
<td>( \mu_\omega + \omega_0 )</td>
<td>( \Delta_\omega )</td>
</tr>
<tr>
<td>( \frac{1}{\sqrt{s}} x(t/s) )</td>
<td>( s\mu_t )</td>
<td>( s\Delta t )</td>
<td>( \frac{1}{s}\mu_\omega )</td>
<td>( \frac{1}{s}\Delta_\omega )</td>
</tr>
</tbody>
</table>

Table 5.1: Effect of shift, modulation and scaling on Heisenberg boxes \((\mu_t, \Delta t, \mu_\omega, \Delta_\omega)\).

So far, we have considered the effect on Heisenberg boxes of shifting in time and frequency and rescaling of functions. How about their sizes? The intuition, corroborated by what we saw with scaling, is that one can trade time spread for frequency spread. Moreover, from Examples 5.1 and 5.2, we know that a function that is narrow in one domain will be broad in the other. It is thus intuitive that the size of the Heisenberg box is lower bounded, so that no function can be arbitrarily narrow in both time and frequency. While the result is often called the *Heisenberg uncertainty principle*, it has been shown independently by a number of others, including Gabor.

**Theorem 5.1 (Uncertainty Principle).** Given a function \( x \in L^2(\mathbb{R}) \), the product of its squared time and frequency spreads is lower bounded as

\[
\Delta_t^2 \Delta_\omega^2 \geq \frac{1}{4}.
\]  

(5.9)

The lower bound is attained by Gaussian functions from (1.69):

\[
x(t) = \gamma e^{-\alpha t^2}, \quad \alpha > 0.
\]  

(5.10)

**Proof.** We prove the theorem for real functions; see Exercise 5.4 for the complex case. Without loss of generality, assume that \( x(t) \) is centered at \( t = 0 \), and that \( x(t) \) has unit energy; otherwise we may shift and scale it appropriately. Since \( x(t) \) is real, it is also centered at \( \omega = 0 \), so \( \mu_t = \mu_\omega = 0 \).

Suppose \( x(t) \) has a bounded derivative \( x'(t) \); if not, \( \Delta_\omega^2 = \infty \) so the statement holds trivially. Consider the function \( t x(t) x'(t) \) and its integral. Using the Cauchy-Schwarz inequality (1.16), we can write

\[
\left| \int_{-\infty}^{\infty} t x(t) x'(t) dt \right|^2 \leq \int_{-\infty}^{\infty} |t x(t)|^2 dt \int_{-\infty}^{\infty} |x'(t)|^2 dt
\]  

(\( a \)) \[
\leq \frac{1}{\Delta t} \int_{-\infty}^{\infty} |t x(t)|^2 dt \int_{-\infty}^{\infty} |x'(t)|^2 dt = \Delta_t^2 \Delta_\omega^2, \quad \Delta_t \Delta_\omega^2.
\]  

(5.11)
5.3. Heisenberg Boxes and the Uncertainty Principle

where (a) follows from Parseval’s relation and the fact that \( x'(t) \) has Fourier transform \( j\omega X(\omega) \). We now simplify the left side:

\[
\int_{-\infty}^{\infty} tx(t)x'(t)\,dt = \left( \frac{1}{2} \int_{-\infty}^{\infty} \frac{dx^2(t)}{dt}\,dt \right) - \frac{1}{2} \int_{-\infty}^{\infty} x^2(t)\,dt = -\frac{1}{2},
\]

where (a) follows from \((x^2(t))' = 2x'(t)x(t)\); (b) is due to integration by parts; and (c) holds because \( x(t) \in \mathcal{L}^2(\mathbb{R}) \) implies that it decays faster than \( 1/|t| \) for \( t \to \pm\infty \), and thus \( \lim_{t\to\pm\infty} tx^2(t) = 0 \) (see Chapter 3). Substituting this into (5.11) yields (5.9).

To find functions that meet the bound with equality, recall that Cauchy-Schwarz inequality becomes an equality if and only if the two functions are collinear (scalar multiples of each other). In our case this means \( x'(t) = \beta tx(t) \). Functions satisfying this relation have the form \( x(t) = \gamma e^{\beta t^2/2} = \gamma e^{-\alpha t^2} \)—the Gaussian functions.

The uncertainty principle points to a fundamental limitation in time-frequency analysis. If we desire to analyze a signal with a “probing function” to extract information about the signal around a location \((\mu_t, \mu_\omega)\), the probing function is necessarily blurred. That is, if we desire very precise frequency information about the signal, its time location will be uncertain, and vice versa.

Example 5.3 (Sinusoid plus Dirac impulse). A generic example to illustrate the tension between time and frequency resolution is the analysis of a signal containing a Dirac impulse in time and a Dirac impulse in frequency:

\[
x(t) = \delta(t-\tau) + e^{j\Omega t},
\]

with the Fourier transform

\[
X(\omega) = e^{-j\omega \tau} + 2\pi \delta(\omega - \Omega).
\]

Clearly, to locate the Dirac impulse in time or in frequency, one needs to be as sharp as possible in the that particular domain, thus compromising the sharpness in the other domain. We illustrate this with a numerical example which, while simple, conveys this basic, but fundamental, trade-off.

Consider a periodic sequence with period \( N = 256 \), containing a complex sinusoidal sequence and a Dirac impulse, as shown in Figure 5.5(a). In Figure 5.5(b), we show the magnitude of the DFT of one period, which has 256 frequency bins, and perfectly identifies the exponential component, while missing the time-domain Dirac impulse completely. In order to increase the time resolution, we divide the period into 16 pieces of length 16, taking the DFT of each, as shown in Figure 5.5(c). Now, we can identify approximately where the time-domain Dirac impulse occurs, but in turn, the frequency resolution is reduced, since we now have only 16 frequency bins. Finally, Figure 5.5(d) shows the dual case to Figure 5.5(b), that is, we plot the magnitude of each sample over time. The Dirac impulse is now perfectly visible, while the sinusoid is just a background wave, with its frequency not easily identified.
Example 5.4 (Chirp Signal). As another example of time-frequency analysis and the trade-off between time and frequency sharpness, we consider a signal consisting of a windowed chirp (complex exponential with a rising frequency). Instead of a fixed frequency $\omega_0$, the “local frequency” is linearly growing with time: $\omega_0 t$. The signal is

$$x(t) = w(t) e^{j\omega_0 t^2},$$

where $w(t)$ is an appropriate window function. Such a chirp signal is not as esoteric as it may look; bats use such signals to hunt for bugs. An example is given in Figure 5.6(a), and an “ideal” time-frequency analysis is sketched in Figure 5.6(b).

As analyzing functions, we choose windowed complex exponentials (but with a fixed frequency). The choice we have is the size of the window (assuming the analyzing function will cover all shifts and modulations of interest). A short window allows for sharp frequency analysis, however, no frequency is really present other than at one instant! A short window will do justice to the transient nature of “frequency” in the chirp, but will only give a very approximate frequency analysis due to the uncertainty principle. A compromise between time and frequency sharpness must be sought, and one such possible analysis is shown in Figure 5.7.

While the uncertainty principle uses a spreading measure akin to variance, other measures can be defined. Though they typically lack fundamental bounds of the kind given by the uncertainty principle (5.9), they can be quite useful as well as intuitive. One such measure, easily applicable to functions that are symmetric in both time and frequency, finds the centered intervals containing $\alpha\%$ of the energy in time and frequency (where $\alpha$ is typically 90 or 95). For a function $x(t)$ of unit norm and symmetric around $\mu_t$, the time spread $\hat{\Delta}_t^{(\alpha)}$ is now defined such that

$$\int_{\mu_t - \frac{1}{2} \hat{\Delta}_t^{(\alpha)}}^{\mu_t + \frac{1}{2} \hat{\Delta}_t^{(\alpha)}} |x(t)|^2 dt = \alpha. \quad (5.12)$$

Similarly, the frequency spread $\hat{\Delta}_\omega^{(\alpha)}$ around the point of symmetry $\mu_\omega$ is defined such that

$$\frac{1}{2\pi} \int_{\mu_\omega - \frac{1}{4} \hat{\Delta}_\omega^{(\alpha)}}^{\mu_\omega + \frac{1}{4} \hat{\Delta}_\omega^{(\alpha)}} |X(\omega)|^2 d\omega = \alpha. \quad (5.13)$$

Exercise 5.3(c) shows that $\hat{\Delta}_t^{(\alpha)}$ and $\hat{\Delta}_\omega^{(\alpha)}$ satisfy the same shift, modulation and scaling behavior as $\Delta_t$ and $\Delta_\omega$ in Table 5.1.
Figure 5.6: Chirp signal and time-frequency analysis with a linearly increasing frequency. 
(a) An example of a windowed chirp with a linearly increasing frequency. Real and imaginary parts are shown as well as the raised cosine window. 
(b) Idealized time-frequency analysis.

Figure 5.7: Analysis of a chirp signal. 
(a) One of the analyzing functions, consisting of a (complex) modulated window. 
(b) Magnitude squared of the inner products between the chirp and various shifts and modulates of the analyzing functions, showing a blurred version of the chirp.

Uncertainty Principle for Discrete Time

Thus far, we have restricted our attention to (continuous-time) functions. Analogous results for (discrete-time) sequences are not as elegant, except in the case of strictly lowpass sequences.

For $x \in \ell^2(\mathbb{Z})$ with energy $E_x = \sum_n |x_n|^2$, define the time center $\mu_n$ and time spread $\Delta_n$ as

$$
\mu_n = \frac{1}{E_x} \sum_{n \in \mathbb{Z}} n |x_n|^2 \quad \text{and} \quad \Delta_n^2 = \frac{1}{E_x} \sum_{n \in \mathbb{Z}} (n - \mu_n)^2 |x_n|^2. 
$$

Define the frequency center $\mu_\omega$ and frequency spread $\Delta_\omega$ as

$$
\mu_\omega = \frac{1}{2\pi E_x} \int_{-\pi}^{\pi} \omega |X(e^{j\omega})|^2 d\omega \quad \text{and} \quad \Delta_\omega^2 = \frac{1}{2\pi E_x} \int_{-\pi}^{\pi} (\omega - \mu_\omega)^2 |X(e^{j\omega})|^2 d\omega.
$$

Because of the interval of integration $[-\pi, \pi]$, the value of $\mu_\omega$ may not match the intuitive notion of frequency center of the sequence. For example, while a sequence supported in Fourier domain on $\cup_{k \in \mathbb{Z}} [0.9\pi + k2\pi, 1.1\pi + k2\pi]$ seems to have its frequency center near $\pi$, $\mu_\omega$ may be far from $\pi$. (TBD: Expand)

With these definitions paralleling those for continuous-time functions, we can obtain a result very similar to Theorem 5.1. One could imagine that it follows from combining Theorem 5.1 with Nyquist-rate sampling of a bandlimited function. The proof suggested in Exercise 5.5 uses the Cauchy-Schwarz inequality similarly to our earlier proof.

**Theorem 5.2** (Discrete-Time Uncertainty Principle). Given a sequence $x \in \ell^2(\mathbb{Z})$ with $X(e^{j\pi}) = 0$, the product of its squared time and frequency spreads is lower bounded as

$$
\Delta_n^2 \Delta_\omega^2 > \frac{1}{4}. 
$$

(5.14)

In addition to the above uncertainty principle for infinite sequences, there is a simple and powerful uncertainty principle for finite-dimensional sequences and their DFTs. This is explored in Exercise 5.8 and arises again in Chapter (TBD).
Figure 5.8: Stretching of a sequence by a factor 2, followed by contraction. This is achieved with upsampling by 2, followed by downsampling by 2, recovering the original. (a) Original. (b) Upsampled. (c) Downsampled.

5.4 Scale and Scaling

In the previous section, we introduced the idea of signal analysis as computing an inner product with a "probing function." Along with the Heisenberg box 4-tuple, another key property of a probing function is its scale. Scale is closely related to time spread, but it is inherently a relative (rather than absolute) quantity. Before further describing scale, let us revisit scaling, especially to point out the fundamental difference between continuous- and discrete-time scaling operations.

An energy-conserving rescaling of $x(t)$ by a factor $s \in \mathbb{R}^+$ yields

$$y(t) = \sqrt{s} x(st) \quad \text{FT} \quad Y(\omega) = \frac{1}{\sqrt{s}} X\left(\frac{\omega}{s}\right).$$

Clearly, such continuous-time scaling is reversible, since a rescaling of $y(t)$ by $(1/s)$ gives

$$\frac{1}{\sqrt{s}} y\left(\frac{t}{s}\right) = x(t).$$

The situation is more complicated in discrete time, where we need multirate signal processing tools introduced in Section 2.5. Given a sequence $x_n$, a "stretching" by an integer factor $N$ can be achieved by upsampling by $N$ as in (2.129). This can be undone by downsampling by $N$ as in (2.121) (see Figure 5.8).

If instead of stretching, we want to "contract" a sequence by an integer factor $N$, we can do this by downsampling by $N$. However, such an operation cannot be undone, as $(N - 1)$ samples out of every $N$ have been lost, and are replaced by zeros during upsampling (see an example for $N = 2$ in Figure 5.9).

Thus, scale changes are more complicated in discrete time. In particular, compressing the time axis cannot be undone since samples are lost in the operation. Scale changes by rational factors are possible through combinations of integer upsampling and downsampling, but cannot be undone in general (see Exercise 5.6).

Let us go back to the notion of scale, and think of a familiar case where scale plays a key role—maps. The usual notion of scale in maps is the following: a map of scale 1:100,000 is a representation where an object of length 1 km is represented by a length of $(10^3\text{m})/10^5 = 1\text{cm}$. That is, the scale factor $s = 10^5$ is used as
a contraction factor, to map a reality $x(t)$ into a scaled version $y(t) = \sqrt{s}x(st)$ with the energy normalization factor $\sqrt{s}$ of no realistic significance because reality and a map are not of the same dimension. However, reality does provide us with something important: a baseline scale against which to compare the map.

When we look at functions in $L^2(\mathbb{R})$, a baseline scale does not necessarily exist. When $y(t) = \sqrt{s} x(st)$, we say that $y$ is at a larger scale if $s > 1$, and at a smaller scale if $s \in (0, 1)$. There is no absolute scale for $y$ unless we arbitrarily define a scale for $x$. We indeed do this sometimes, as we will see in Chapter (TBD).

Now consider the use of a probing function $\phi(t)$ to extract some information about $x(t)$. If we compute the inner product between the probing function and a scaled function, we get

$$\langle \sqrt{s}x(st), \phi(t) \rangle = \sqrt{s} \int x(st)\phi(t) \, dt = \frac{1}{\sqrt{s}} \int x(\tau)\phi(\frac{\tau}{s}) \, d\tau$$

$$= \left\langle x(t), \frac{1}{\sqrt{s}} \phi(t/s) \right\rangle .$$

(5.16)

Probing a contracted function is equivalent to stretching the probe, thus emphasizing that scale is relative. If only stretched and contracted versions of a single probe are available, large scale features in $x(t)$ are seen using stretched probing functions, while fine details in $x(t)$ are seen using contracted probing functions.

In summary, large scales $s \gg 1$ correspond to contracted versions of reality, or to widely spread probing functions. This duality is inherent in the inner product (5.16). Figure 5.10 shows an aerial photograph with different scale factors as per our convention, while Figure 5.11 shows the interaction of signals with various size features and probing functions.
Signals with features at different scales require probing functions adapted to the scales of those features. (a) A wide-area feature requires a wide probing function. (b) A sharp feature requires a sharp probing function.

5.5 Resolution, Bandwidth and Degrees of Freedom

The notion of resolution is intuitive for images. If we compare two photographs of the same size depicting the same reality, one sharp and the other blurry, we say that the former has higher resolution than the latter. While this intuition is related to a notion of bandwidth, it is not the only interpretation.

A more universal notion is to define resolution as the number of degrees of freedom per unit time (or unit space for images) for a set of signals. Classical bandwidth is then proportional to resolution. Consider the set of functions \( x(t) \) with spectra \( X(\omega) \) supported on the interval \([-\Omega, \Omega]\). Then, the sampling theorem (see Chapter 4) states that samples taken every \( T = \pi/\Omega \text{sec} \), or \( x_n = x(nT), n \in \mathbb{Z} \), uniquely specify \( x(t) \). In other words, real functions of bandwidth \( 2\Omega \) have \( \Omega/\pi \) real degrees of freedom per unit time.

As an example of a set of functions that, while not bandlimited, do have a finite number of degrees of freedom per unit time, consider piecewise constant functions over unit intervals:

\[
x(t) = x_{[t]} = x_n, \quad n \leq t < n + 1, \quad n \in \mathbb{Z}.
\]

Clearly, \( x(t) \) has 1 degree of freedom per unit time, but an unbounded spectrum since it is discontinuous at every integer. This function is part of a general class of functions belonging to the so-called shift-invariant subspaces we studied in Chapter 4 (see also Exercise 5.7).

Scaling affects resolution, and as in the previous section there is a difference between discrete and continuous time. For sequences, we start with a reference sequence space \( S \) in which there are no fixed relationships between samples and the rate of the samples is taken to be 1 per unit time. In the reference space, the resolution is 1 per unit time because each sample is a degree of freedom. Downsampling a sequence from \( S \) by \( N \) leads to a sequence having a resolution of \( 1/N \) per unit time. Higher-resolution sequences can be obtained by combining sequences appropriately; see Example 5.5 below. For continuous-time functions, scaling is less disruptive to the time axis, so calculating the number of degrees of freedom per unit time is not difficult. If \( x(t) \) has resolution \( \tau \) (per unit time), then \( x(at) \) has resolution \( a\tau \).

Filtering can affect resolution as well. If a function of bandwidth \([-\Omega, \Omega]\) is perfectly lowpass filtered to \([-\beta\Omega, \beta\Omega]\), \( 0 < \beta < 1 \), then its resolution changes from \( \Omega/\pi \) to \( \beta\Omega/\pi \). The same holds for sequences, where an ideal lowpass filter with support \([-\beta\pi, \beta\pi]\), \( 0 < \beta < 1 \), reduces the resolution to \( \beta \) samples per unit time.

Example 5.5 (Scale and Resolution). First consider the continuous-time case. Let \( x(t) \) be a bandlimited function with frequency support \([-\Omega, \Omega]\). Then \( y(t) = \sqrt{2} x(2t) \) has doubled scale and resolution. Inversely, \( y(t) = \frac{1}{\sqrt{2}} x(t/2) \) has half the
5.6 Haar Tiling (Old)

**Figure 5.12:** Interplay of scale and resolution for continuous- and discrete-time signals. We assume the original signal to have scale $s = 1$ and resolution $\tau = 1$, and indicate the resulting scale and resolution of the output by $s'$ and $\tau'$. (a) Filtering of signal. (b) Downsampling by 2. (c) Upsampling by 2. (d) Downsampling followed by upsampling. (e) Filtering and downsampling. (f) Upsampling and filtering.

scale and half the resolution. Finally, $y(t) = (h \ast x)(t)$, where $h(t)$ is an ideal low-pass filter with support $[-\Omega/2, \Omega/2]$, has unchanged scale but half the resolution.

Now let $x_n$ be a discrete-time sequence. The downsampled sequence $y_n$ as in (2.116) has doubled scale and half the resolution with respect to $x_n$. The upsampled sequence $y_n$ as in (2.117) has half the scale and unchanged resolution. A sequence first downsampled by 2 and then upsampled by 2 keeps all the even samples and zeros out the odd ones, with unchanged scale and half the resolution. Finally, filtering with an ideal halfband lowpass filter with frequency support $[-\pi/2, \pi/2]$ leaves the scale unchanged, but halves the resolution. Some of the above relations are depicted in Figure 5.12.

**Example 5.6 (How does Haar tile the time-frequency plane?)**. To illustrate the concepts introduced in this chapter and to give a hint of where we are going, we go back to our favorite example—Haar. Recall the smoothing and differencing operators we covered in Section ???. We want to try to get a slightly better frequency localization than that obtained by the Dirac representation. Thus, we use the Haar filters and first smooth our signal. This is shown in Fig. 5.13. We
Figure 5.14: Differencing operator. (a) Block diagram. (b) An example signal. (c) Differenced signal from part (b).

Figure 5.15: Putting it all together. The two-channel filter bank implementing both projections onto the smooth space as well as the detail space, the sum of which gives back the original signal.
5.6. Haar Tiling (Old)

Figure 5.16: Time-frequency tiling obtained with the Haar basis.

know that

\[ g_n = \frac{1}{\sqrt{2}} (\delta_n + \delta_{n-1}) , \]

and \( \langle g, g_{-2k} \rangle = \delta_k \). We also know that this combination is an orthogonal projection

\[ P_{V_1} = GU_2D_2G^T \]

onto the space of smooth signals we call \( V_1 \). Recall from before \((??)-(??)\) that a basis for that space is \( \Phi_g = \{g_{-2k}\}_{k \in \mathbb{Z}} \). You can see in Fig. 5.13(b) an example of the original signal while in part (c) we see the result of applying operator \( P_{V_1} \) to it, that is,

\[ x_{V_1} = \begin{bmatrix} \cdots & \frac{1}{2}(x_0 + x_1) & \frac{1}{2}(x_0 - x_1) & \frac{1}{2}(x_2 + x_3) & \frac{1}{2}(x_2 - x_3) & \cdots \end{bmatrix} . \]

We can now repeat the process for the differencing operator, shown in Fig. 5.14. Now

\[ h_n = \frac{1}{\sqrt{2}} (\delta_n - \delta_{n-1}) , \]

and \( \langle h, h_{-2k} \rangle = \delta_k \). We know that this combination is an orthogonal projection as well,

\[ P_{W_1} = HU_2D_2H^T , \]

onto the space of detail signals we call \( W_1 \). A basis for that space is \( \Phi_h = \{h_{-2k}\}_{k \in \mathbb{Z}} \). You can see in Fig. 5.14(b) an example of the original signal while in part (c) we see the result of applying operator \( P_{W_1} \) to it, that is,

\[ x_{W_1} = \begin{bmatrix} \cdots & \frac{1}{2}(x_0 - x_1) & -\frac{1}{2}(x_0 - x_1) & \frac{1}{2}(x_2 - x_3) & -\frac{1}{2}(x_2 - x_3) & \cdots \end{bmatrix} . \]

If we look carefully at parts (c) of Figs. 5.13 and 5.14, we see that the original signal can be obtained by summing the two. This is shown in Fig. 5.15. One can verify then that a basis for the original space is \( \Phi = \{g_{-2k}, h_{-2k}\}_{k \in \mathbb{Z}} \). The original signal is the sum of the two projections:

\[ x = x_{V_1} + x_{W_1} = P_{V_1}x + P_{W_1}x , \]
implemented by a two-channel filter bank given in Fig. 5.15(a). Finally, such a decomposition has an associated time-frequency tiling given in Fig. 5.16. Each tile corresponds to one basis function. What we can see is exactly what we expected: The frequency localization of the Haar basis is better than that of the Dirac representation as the frequency has been split in half; the price we paid was a slightly worse time localization as shown by the basis functions now being of length 2.

Can we do better than that? As a preview of coming attractions, consider smoothing the smoothed version some more. What do we get? We start getting better frequency localization at low frequencies while keeping good time localization at high frequencies (see Fig. 5.17). To see what the basis is now, we need to transform this filter bank into an equivalent three-channel filter bank as in Fig. 5.18. To do that, we need to use the identities on how we can interchange filtering and sampling given in Section ??.

We write expressions for equivalent filters in z-transform-domain:

\[
H^{(1)}(z) = H(z) = \frac{1}{\sqrt{2}}(1 - z^{-1}),
\]
\[
H^{(2)}(z) = G(z)H(z^2) = \frac{1}{2}(1 + z^{-1} - z^{-2} - z^{-3}),
\]
\[
G^{(2)}(z) = G(z)G(z^2) = \frac{1}{2}(1 + z^{-1} + z^{-2} + z^{-3}),
\]

giving rise to the following output signals:

\[
x_{W1} = \begin{bmatrix} \ldots & \frac{1}{2}(x_0 - x_1) & -\frac{1}{2}(x_0 - x_1) & \frac{1}{2}(x_2 - x_3) & -\frac{1}{2}(x_2 - x_3) & \ldots \end{bmatrix},
\]
\[
x_{W2} = \begin{bmatrix} \ldots & \frac{1}{4}(x_0 + x_1 - x_2 - x_3) & \frac{1}{4}(x_0 + x_1 - x_2 - x_3) & \frac{1}{4}(x_0 + x_1 - x_2 - x_3) & \frac{1}{4}(x_0 + x_1 - x_2 - x_3) & \ldots \end{bmatrix},
\]
\[
x_{V2} = \begin{bmatrix} \ldots & \frac{1}{4}(x_0 + x_1 + x_2 + x_3) & \frac{1}{4}(x_0 + x_1 + x_2 + x_3) & \frac{1}{4}(x_0 + x_1 + x_2 + x_3) & \frac{1}{4}(x_0 + x_1 + x_2 + x_3) & \ldots \end{bmatrix}.
\]

It is easy to see that summing these three signals, we get the original signal back. Continuing in this way, we will get to the discrete wavelet transform with Haar filters. Fig. 5.19 shows the associated time-frequency tiling.

Fig. 5.20 shows some possibilities of tiling the time-frequency plane and various trade-offs involved. Part (a) is the DWT with 3 levels, in part (b) we kept on splitting the highpass channel, while part (c) shows a more arbitrary tiling.

### 5.7 Case Studies

So far, our discussion has been mostly conceptual, and the examples synthetic. We now look at case studies using real-world signals such as music, images and communication signals. Our discussion is meant to develop intuition rather than be rigorous. We want to excite you to continue studying the rich set of tools responsible for examples below.
5.7. Case Studies

5.7.1 Music and Time-Frequency Analysis
5.7.2 Images and Pyramids
5.7.3 Singularities, Denoising and Superresolution
5.7.4 Channel Equalization and OFDM
## Chapter at a Glance

We now summarize the main concepts and results seen in this chapter in a tabular form.

### Uncertainty Principle

For a function $x(t) \in L^2(\mathbb{R})$ with Fourier transform $X(\omega)$,

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy $E_x$</td>
<td>$E_x = \int</td>
</tr>
<tr>
<td>Time center $\mu_t$</td>
<td>$\mu_t = \frac{1}{E_x} \int t</td>
</tr>
<tr>
<td>Time spread $\Delta_t$</td>
<td>$\Delta_t = \left( \frac{1}{E_x} \int (t - \mu_t)^2</td>
</tr>
<tr>
<td>Frequency center $\mu_\omega$</td>
<td>$\mu_\omega = \frac{1}{2\pi E_x} \int \omega</td>
</tr>
<tr>
<td>Frequency spread $\Delta_\omega$</td>
<td>$\Delta_\omega = \left( \frac{1}{2\pi E_x} \int (\omega - \mu_\omega)^2</td>
</tr>
</tbody>
</table>

$\Rightarrow \Delta_t^2 \Delta_\omega^2 \geq \frac{1}{4}$, with equality achieved by a Gaussian $x(t)$.

### Scale

If $x(t)$ is defined to have scale 1, then $y(t) = \sqrt{s}x(st)$ has scale $s$.

### Resolution

The resolution of a set of functions is the number of degrees of freedom per unit time.
Historical Remarks

Uncertainty principles stemming from the Cauchy-Schwarz inequality have a long and rich history. The best known one is Heisenberg’s uncertainty principle in quantum physics, first developed in a 1927 essay [31]. Werner Karl Heisenberg (1901-1976) was a German physicist, credited as a founder of quantum mechanics, for which he was awarded the Nobel Prize in 1932. He had seven children, one of whom, Martin Heisenberg, was a celebrated geneticist. He collaborated with Bohr, Pauli and Dirac, among others. While he was initially attacked by the Nazi war machine for promoting Einstein’s views, he did head the Nazi nuclear project during the war. His role in the project has been a subject of controversy every since, with differing views on whether he was deliberately stalling Hitler’s efforts or not.

Kennard is credited with the first mathematically exact formulation of the uncertainty principle, and Robertson and Schrödinger provided generalizations. The uncertainty principle presented in Theorem 5.1 was proven by Weyl and Pauli and introduced to signal processing by Dennis Gabor (1900-1979) [25], a Hungarian physicist, and another winner of the Nobel Prize for physics (he is also known as inventor of holography). By finding a lower bound to $\Delta t \Delta \omega$, Gabor was intending to define an information measure or capacity for signals. Shannon’s communication theory [52] proved much more fruitful for this purpose, but Gabor’s proposal of signal analysis by shifted and modulated Gaussian functions has been a cornerstone of time-frequency analysis ever since. Slepian’s survey [53] is enlightening on these topics.

Further Reading

Many of the uncertainty principles for discrete-time signals are considerably more complicated than Theorem 5.2. We have given only a result that follows papers by Ishii and Furukawa [34] and Calvez and Vilbé [7].

Donoho and Stark [20] derived new uncertainty principles in various domains. Particularly influential was an uncertainty principle for finite-dimensional signals and a demonstration of its significance for signal recovery (see Exercises 5.8 and 5.9). More recently, Donoho and Huo [19] introduced performance guarantees for $\ell^1$ minimization-based signal recovery algorithms; this has sparked a large body of work.

Exercises with Solutions

5.1. TBD
Exercises

5.1. Finite Sequences and Their DFTs:
Show that if a sequence has a finite number of terms, then its DFT cannot be zero over an interval (that is, it can only have isolated zeros). Conversely, show that if a discrete Fourier transform is zero over an interval, then the corresponding sequence has an infinite number of nonzero terms.

5.2. Box Function, Its Convolution, and Limits:
Given is the box function from (5.3).
(i) What is the time spread $\Delta t^2$ of the triangle function $(b * b)$?
(ii) What is the time spread $\Delta t^2$ of the function $b$ convolved with itself $N$ times?

5.3. Properties of Time and Frequency Spreads:
Consider the time and frequency spreads as defined in (5.2) and (5.6), respectively.
(i) Show that time shifts and complex modulations of $x(t)$ as in (5.7) leave $\Delta t$ and $\Delta \omega$ unchanged.
(ii) Show that energy conserving scaling of $x(t)$ as in (5.8) increases $\Delta t$ by $s$, while decreasing $\Delta \omega$ by $s$, thus leaving the time-frequency product unchanged.
(iii) Show (i)-(ii) for the time-frequency spreads $\Delta_{t(\alpha)}$ and $\Delta_{\omega(\alpha)}$ as defined in (5.12) and (5.13).

5.4. Uncertainty Principle for Complex Functions:
Prove Theorem 5.1 without assuming that $x(t)$ is a real function.
(Hint: The proof requires more than the Cauchy-Schwarz inequality and integration by parts. Use the product rule of differentiation, $\frac{d}{dt}|x(t)|^2 = x'(t)x^*(t) + x^*(t)x(t)$. Also, use that for any $\alpha \in \mathbb{C}$, $|\alpha| \geq \frac{1}{2}|\alpha + \alpha^*|$.)

5.5. Discrete-Time Uncertainty Principle:
Prove Theorem 5.2 for real sequences. Do not forget to provide an argument for the strictness of inequality (5.14).
(Hint: Use the Cauchy-Schwarz inequality to bound $\int_{-\pi}^{\pi} \omega X(e^{j\omega}) \left| \sum_{n} X(e^{j\omega}) \right|^2 d\omega$.)

5.6. Rational Scale Changes on Sequences:
A scale change by a factor $M/N$ can be achieved by upsampling by $M$ followed by downsampling by $N$.
(i) Consider a scale change by $3/2$, and show that it can be implemented either by upsampling by 3, followed by downsampling by 2, or the converse.
(ii) Show that the scale change by $3/2$ cannot be undone, even though it is a stretching operation.
(iii) Using the fact that when $M$ and $N$ are coprime, upsampling by $M$ and downsampling by $N$ commute, show that a sampling rate change by $M/N$ cannot be undone unless $N = 1$.

5.7. Shift-Invariant Subspaces and Degrees of Freedom:
Define a shift-invariant subspace $S$ as
$$ S = \text{span}(\{\varphi(t - nT)\}_{n \in \mathbb{Z}}), \quad T \in \mathbb{R}^+. $$
(i) Show that the piecewise-constant function defined in (5.17) belongs to such a space when $\varphi(t)$ is the indicator function of the interval $[0, 1]$ and $T = 1$.
(ii) Show that a function in $S$ has exactly $1/T$ degrees of freedom per unit time.

5.8. Uncertainty Principle for the DFT:
Let $x$ and $X$ be a length-$N$ DFT pair, and let $N_t$ and $N_\omega$ denote the number of nonzero components of $x$ and $X$, respectively.
(i) Prove that $X$ cannot have $N_t$ consecutive zeros, where “consecutive” is interpreted mod $N$.

(*Hint:* For an arbitrary selection of $N_t$ consecutive components of $X$, form a linear system relating the nonzero components of $x$ to the selected components of $X$.)

(ii) Using the result of the first part, prove $N_t N_\omega \geq N$. This uncertainty principle is due to Donoho and Stark [20].

5.9. Signal Recovery Based on the Finite-Dimensional Uncertainty Principle:
Suppose the DFT of a length-$N$ signal $x$ is known to have only $N_\omega$ nonzero components. Using the result of Exercise 5.8, show that the limited DFT-domain support makes it possible to uniquely recover $x$ from any $M$ (time-domain) components as long as $2(N - M)N_\omega < N$.

(*Hint:* Show that nonunique recovery leads to a contradiction.)
Figure 5.17: What worked once works again; We smooth the already smoothed version of the signal some more.
Figure 5.18: Equivalent three-channel filter bank obtained from the two-level discrete wavelet transform.

Figure 5.19: Time-frequency tiling obtained with the Haar basis and a two-level split.

Figure 5.20: Various time-frequency tilings.