

ECE-700 Time-Frequency/Wavelet Notes

Phil Schniter

March 12, 2007

1 Time-Frequency Analysis and Continuous Wavelet Transform

- Why Transforms?: In the field of signal processing we frequently encounter the use of “transforms.” Transforms are named such because they take a signal and *transform* it into another signal, hopefully one which is easier to process or analyze than the original. Essentially, transforms are used to manipulate signals such that their most important characteristics are made plainly evident. To isolate a signal’s important characteristics, however, one must employ a transform that is well matched to that signal. For example, the Fourier transform, while well matched to certain classes of signal, does not efficiently extract information about signals in other classes. This latter fact motivates our development of the wavelet transform.
- Limitations of Fourier Analysis: Let’s again consider the Continuous-Time Fourier Transform (CTFT) pair:

$$\begin{aligned}X(\Omega) &= \int_{-\infty}^{\infty} x(t)e^{-j\Omega t} dt \\x(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega)e^{j\Omega t} d\Omega,\end{aligned}$$

where we have abbreviated our earlier notation $B(j\Omega)$ to $B(\Omega)$. The Fourier transform pair supplies us with our notion of “frequency.” In other words, all of our intuitions regarding the relationship between the time domain and the frequency domain can be traced to this particular transform pair.

It will be useful to view the CTFT as a complex-exponential signal expansion. Specifically, the inverse CTFT equation above says that the time-domain signal $x(t)$ can be expressed as a weighted “summation” of basis elements $\{b_{\Omega}(t), -\infty < \Omega < \infty\}$, where $b_{\Omega}(t) := e^{j\Omega t}$ is the basis element corresponding to frequency Ω . Since the number of CTFT basis elements is uncountably infinite, the summation is accomplished with an integral. Notice that $X(\Omega)$, which specifies the contribution of $b_{\Omega}(t)$ to $x(t)$, can be interpreted as a measure of the similarity between $b_{\Omega}^*(t)$ and $x(t)$.

The Fourier Series (FS) can be considered as a special sub-case of the CTFT that applies when the time-domain signal $x(t)$ is periodic. Recall that, if $x(t)$ is periodic with period T , then $x(t)$ can be expressed as a weighted summation of basis elements $\{b_k(t)\}_{k=-\infty}^{\infty}$,

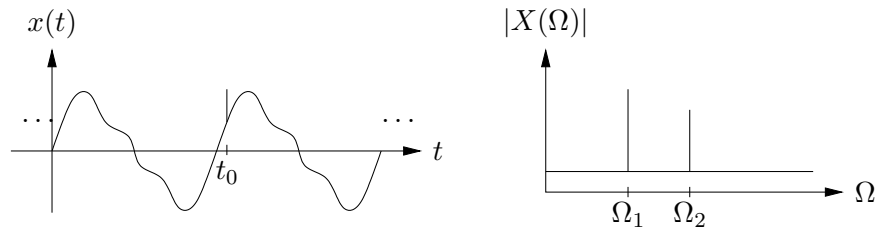
where now $b_k(t) := e^{j\frac{2\pi}{T}tk}$:

$$x(t) = \sum_{k=-\infty}^{\infty} X[k]e^{j\frac{2\pi}{T}tk}$$

$$X[k] = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)e^{-j\frac{2\pi}{T}tk} dt.$$

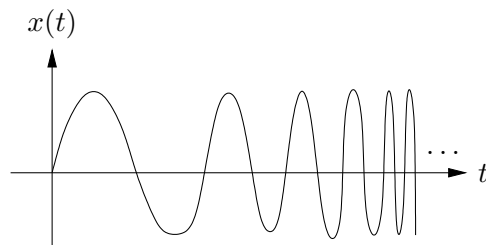
Here, the basis elements come from a countably infinite set and are parameterized by the discrete frequency index $k \in \mathbb{Z}$. The coefficient $X[k]$, which specifies the contribution of $b_k(t)$ to $x(t)$, can be interpreted as a measure of similarity between $b_k^*(t)$ and $x(t)$ on the interval $[-\frac{T}{2}, \frac{T}{2}]$.

Though quite popular, Fourier analysis is not always the best tool to analyze a signal whose characteristics vary with time. For example, consider a signal composed of a periodic component plus a sharp “glitch” at time t_0 , illustrated in time- and frequency-domains below.



Fourier analysis is successful in reducing the complicated-looking periodic component into a few simple parameters: the frequencies $\{\Omega_1, \Omega_2\}$ and their corresponding magnitudes/phases. The glitch component, described compactly in terms of its time-domain location t_0 and amplitude, however, is not described efficiently in the frequency domain since it produces a wide spread of frequency components. Thus, neither time- nor frequency-domain representations alone give an efficient description of the glitched periodic signal: each representation distills only certain aspects of the signal.

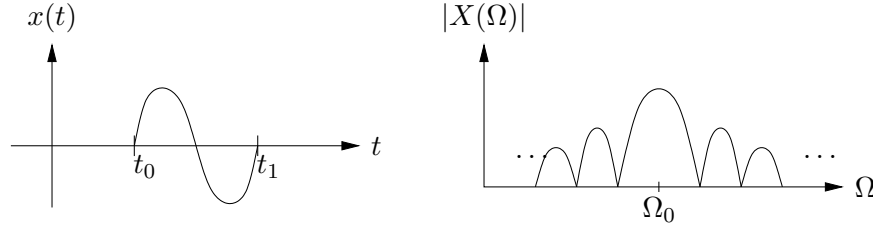
As another example, consider the “linear chirp” $x(t) = \sin(\Omega t^2)$ illustrated below.



Though written using the $\sin(\cdot)$ function, the chirp is not described by a single Fourier frequency. We might try to be clever and write the time-varying phase argument of $\sin(\cdot)$ as $\varphi(t) = \frac{\partial\varphi(t)}{\partial t}t + \varphi(0)$, where $\frac{\partial\varphi(t)}{\partial t}$ can be interpreted as an “instantaneous frequency.” In this case, $\varphi(t) = \Omega t^2$, which gives the instantaneous frequency $\Omega(t) := \frac{\partial\varphi(t)}{\partial t} = 2\Omega t$, which exhibits a simple linear variation in time. But here we must be cautious, since this newly defined notion of “instantaneous frequency” is *not* consistent with the Fourier notion of frequency. Recall that the CTFT says that a signal can be constructed as a superposition

of fixed-frequency basis elements $e^{j\Omega t}$, each spread out uniformly over all time. In other words, there is nothing “instantaneous” about Fourier frequency.

As a third example, consider a Ω_0 -frequency sinusoid that is rectangularly windowed to extract only one period.



Here, our instantaneous-frequency argument might claim that

$$\Omega(t) = \begin{cases} \Omega_0 & t \in \text{window} \\ 0 & t \notin \text{window} \end{cases},$$

where $\Omega(t)$ takes on exactly two values. In contrast, Fourier theory says that rectangular windowing induces frequency-domain spreading with a $\frac{\sin(\Omega)}{\Omega}$ profile, resulting in a continuum of Fourier frequency components. Here again, we see that our notions of “instantaneous” frequency are not compatible with the standard Fourier notion of frequency.

- **Time-Frequency Uncertainty Principle:** Recall that Fourier basis $b_\Omega(t) = e^{j\Omega t}$ exhibits poor time resolution—a consequence of the fact that $|b_\Omega(t)|$ for every fixed Ω is uniformly spread over all $t \in (-\infty, \infty)$. By “time resolution,” we mean the ability to easily identify the timing of signal events like the “glitch” described earlier in an example.

At the opposite extreme, a basis composed of shifted Dirac deltas $b_\tau(t) = \delta(t - \tau)$ would have excellent time resolution but terrible “frequency resolution,” since every Dirac basis element is evenly spread over all Fourier frequencies $\Omega \in (-\infty, \infty)$. This can be understood from the fact that $|B_\tau(\Omega)| = |\int_{-\infty}^{\infty} b_\tau(t)e^{-j\Omega t} dt| = 1 \forall \Omega$, for every fixed τ . By “frequency localization,” we mean the ability to easily identify the spectral properties of signal components that are sinusoidal in nature.

These observations motivate the question: Does there exist a basis that provides simultaneously excellent time *and* frequency resolutions? The answer is “not really”; there is a fundamental tradeoff between the time and frequency resolutions of any basis. This can be understood from the fact that no basis element can be simultaneously well concentrated in the time and frequency domains. This idea is made concrete below.

Consider an arbitrary basis element, or waveform, $b(t)$, whose CTFT will be denoted by $B(\Omega) = \int_{-\infty}^{\infty} b(t)e^{-j\Omega t}$. In some cases it will be convenient to express frequency in Hertz rather than radians, for which we write $\Omega = 2\pi f$. We define the temporal and spectral centers¹ as

$$t_c = \frac{1}{E} \int_{-\infty}^{\infty} t |b(t)|^2 dt$$

$$f_c = \frac{1}{E} \int_{-\infty}^{\infty} f |B(2\pi f)|^2 df,$$

¹It may be interesting to note that both $\frac{1}{E}|b(t)|^2$ and $\frac{1}{E}|B(2\pi f)|^2$ are non-negative and integrate to one, thereby satisfying the requirements of probability density functions. The temporal/spectral centers can then be interpreted as the *means* (i.e., centers of mass) in the time/frequency domains, respectively.

and the temporal and spectral widths² as

$$\Delta_t = \sqrt{\frac{1}{E} \int_{-\infty}^{\infty} (t - t_c)^2 |b(t)|^2 dt}$$

$$\Delta_f = \sqrt{\frac{1}{E} \int_{-\infty}^{\infty} (f - f_c)^2 |B(2\pi f)|^2 df},$$

where E denotes the energy of the waveform, i.e.,

$$E = \int_{-\infty}^{\infty} |b(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |B(\Omega)|^2 d\Omega = \int_{-\infty}^{\infty} |B(2\pi f)|^2 df$$

via Parseval's theorem. If the waveform is well-localized in time, then $b(t)$ will be concentrated at $t = t_c$ and Δ_t will be small. If the waveform is well-localized in frequency, then $B(2\pi f)$ will be concentrated at $f = f_c$ and Δ_f will be small. If the waveform is well-localized in both time *and* frequency, then $\Delta_t \Delta_f$ will be small. The quantity $\Delta_t \Delta_f$ is known as the “time-bandwidth product.”

From the definitions above one can derive the fundamental properties below. When interpreting the properties, it helps to think of the waveform $b(t)$ as a prototype that can be used to generate an entire basis set. For example, the Fourier basis $\{b_\Omega(t), -\infty < \Omega < \infty\}$ can be generated by frequency shifts of $b(t) = 1$, while the Dirac basis $\{b_\tau(t), -\infty < \tau < \infty\}$ can be generated by time shifts of $b(t) = \delta(t)$.

1. Δ_t and Δ_f are invariant to time and frequency³ shifts.

$$\begin{aligned} \Delta_t(b(t)) &= \Delta_t(b(t - t_0)) \quad \forall t_0 \in \mathbb{R} \\ \Delta_f(B(2\pi f)) &= \Delta_f(B(2\pi(f - f_0))) \quad \forall f_0 \in \mathbb{R} \end{aligned}$$

This implies that all basis elements constructed from time and/or frequency shifts of a prototype waveform $b(t)$ will inherit the temporal and spectral widths of $b(t)$.

2. The “time-bandwidth product” $\Delta_t \Delta_f$ is invariant to time-scaling.⁴

$$\left. \begin{aligned} \Delta_t(b(t/a)) &= |a| \Delta_t(b(t)) \\ \Delta_f(b(t/a)) &= \frac{1}{|a|} \Delta_f(b(t)) \end{aligned} \right\} \Rightarrow \Delta_t \Delta_f(b(t/a)) = \Delta_t \Delta_f(b(t)) \quad \forall a \in \mathbb{R}$$

Observe that time-domain expansion (i.e., $|a| > 1$) increases the temporal width but decreases the spectral width, while time-domain contraction (i.e., $|a| < 1$) does the opposite. This suggests that time-scaling might be a useful tool for the design of a basis element with a particular tradeoff between time and frequency resolution. On the other hand, scaling cannot simultaneously increase both time *and* frequency resolution.

3. The time-bandwidth product is lower bounded as follows:

$$\Delta_t \Delta_f \geq \frac{1}{4\pi}$$

This is known as the “time-frequency uncertainty principle.”

²The quantities Δ_t^2 and Δ_f^2 are analogous to variances.

³Keep in mind the fact that $b(t)$ and $B(2\pi f) = \int_{-\infty}^{\infty} b(t) e^{-j2\pi f t} dt$ are alternate descriptions of the same waveform; we could have written $\Delta_f(b(t) e^{j2\pi f_0 t})$ in place of $\Delta_f(B(2\pi(f - f_0)))$ above.

⁴The invariance property holds also for frequency scaling, as implied by the Fourier transform property $b(t/a) \leftrightarrow |a| B(2\pi a f)$.

4. The Gaussian pulse $g(t)$ achieves the minimum time-bandwidth product $\Delta_t \Delta_f = \frac{1}{4\pi}$.

$$\begin{aligned} g(t) &= \sqrt{\frac{\alpha}{\pi}} e^{-\alpha t^2}, \quad \alpha \in \mathbb{R} \\ G(\Omega) &= e^{-\frac{\Omega^2}{4\alpha}} \end{aligned}$$

Note that this waveform is neither bandlimited nor time-limited, but well concentrated in both domains (around the points $t_c = 0$ and $f_c = 0$).

Properties 1 and 2 can be easily verified using the definitions above. Properties 3 and 4 follow from the Cauchy-Schwarz inequality (see the proof below).

Since the Gaussian pulse $g(t)$ achieves the minimum time-bandwidth product, it makes for a good prototype waveform, at least in theory. In other words, we might consider constructing a basis from time shifted, frequency shifted, time scaled, or frequency scaled versions of $g(t)$ to give a range of spectral/temporal centers and spectral/temporal resolutions. Since the Gaussian pulse has doubly-infinite time-support, though, it is not very practical. Basis construction from a prototype waveform is the main concept behind “Short-Time Fourier Analysis” and the “Continuous Wavelet Transform” discussed next.

Proof. Say that $q(t)$ is a unit-energy and time/frequency-centered version of $b(t)$, i.e.,

$$\begin{aligned} q(t) &= \frac{1}{\sqrt{E}} e^{-j2\pi f_c(t+t_c)} b(t+t_c) \\ Q(2\pi f) &= \frac{1}{\sqrt{E}} e^{j2\pi f t_c} B(2\pi(f+f_c)). \end{aligned}$$

so that

$$\begin{aligned} \int_{-\infty}^{\infty} t |q(t)|^2 dt &= \frac{1}{E} \int_{-\infty}^{\infty} (t-t_c) |b(t)|^2 dt = 0 \\ \int_{-\infty}^{\infty} t^2 |q(t)|^2 dt &= \frac{1}{E} \int_{-\infty}^{\infty} (t-t_c)^2 |b(t)|^2 dt = \Delta_t^2 \\ \int_{-\infty}^{\infty} f |Q(2\pi f)|^2 df &= \frac{1}{E} \int_{-\infty}^{\infty} (f-f_c) |B(2\pi f)|^2 df = 0 \\ \int_{-\infty}^{\infty} f^2 |Q(2\pi f)|^2 df &= \frac{1}{E} \int_{-\infty}^{\infty} (f-f_c)^2 |B(2\pi f)|^2 df = \Delta_f^2. \end{aligned}$$

The Cauchy-Schwarz inequality says that

$$\left| \int_{-\infty}^{\infty} x(t)y(t)dt \right|^2 \leq \int_{-\infty}^{\infty} |x(t)|^2 dt \int_{-\infty}^{\infty} |y(t)|^2 dt$$

with equality when $x(t) = ky(t)$ for scalar k . Choosing $x(t) = tq(t)$ and $y(t) = q'(t)$, where $q'(t)$ denotes the derivative of $q(t)$, we find that

$$\begin{aligned} \int_{-\infty}^{\infty} |x(t)|^2 dt &= \int_{-\infty}^{\infty} t^2 |q(t)|^2 dt = \Delta_t^2 \\ \int_{-\infty}^{\infty} |y(t)|^2 dt &= \int_{-\infty}^{\infty} |Y(2\pi f)|^2 df = (2\pi)^2 \int_{-\infty}^{\infty} f^2 |Q(2\pi f)|^2 df = (2\pi)^2 \Delta_f^2, \end{aligned}$$

where we used Parseval's equality and the fact that the Fourier transform of $q'(t)$ equals $j\Omega Q(\Omega)$. Thus, the right side of the Cauchy-Schwarz inequality equals $(2\pi)^2 \Delta_f^2 \Delta_t^2$. Since $q(t)q'(t) = \frac{1}{2} \frac{\partial^2 q(t)}{\partial t^2}$, we find that

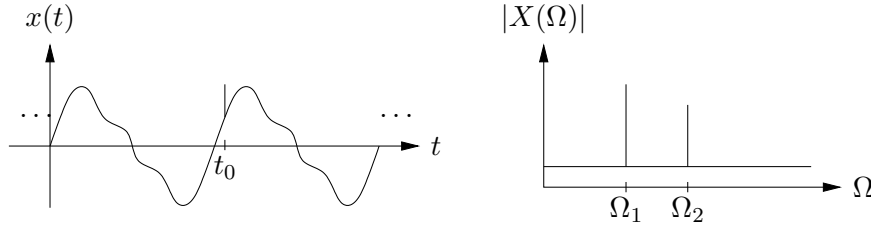
$$\left| \int_{-\infty}^{\infty} x(t)y(t)dt \right|^2 = \left| \frac{1}{2} \int_{-\infty}^{\infty} t \frac{\partial^2 q(t)}{\partial t^2} dt \right|^2 = \frac{1}{4} \left| tq^2(t) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} q^2(t)dt \right|^2 = \frac{1}{4},$$

assuming that $q(t)$ vanishes faster than $\frac{1}{\sqrt{|t|}}$ as $t \rightarrow \pm\infty$. Combining with our earlier result, we see that

$$\Delta_t \Delta_f \geq \frac{1}{4\pi},$$

with equality when $q'(t) = ktq(t)$ for scalar constant k . This is satisfied by $q(t)$ of the form $q(t) = \sqrt{\alpha/\pi} e^{-\alpha t^2}$ for $\alpha \in \mathbb{R}$. \square

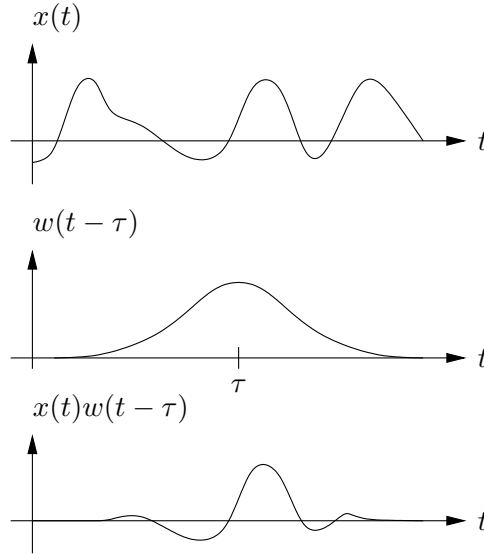
- *Short-Time Fourier Analysis:* We saw earlier that Fourier analysis is not well suited to describing local changes in “frequency content” because the frequency components defined by the Fourier transform have infinite (i.e., global) time support. For example, if we have a signal with periodic components plus a glitch at time t_0 , we might want accurate knowledge of both the periodic component frequencies *and* the glitch time.



The Short-Time Fourier Transform (STFT) provides a means of joint time-frequency analysis. The STFT pair can be written

$$\begin{aligned} X_{\text{STFT}}(\Omega, \tau) &= \int_{-\infty}^{\infty} x(t)w(t - \tau)e^{-j\Omega t} dt \\ x(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X_{\text{STFT}}(\Omega, \tau)w(t - \tau)e^{j\Omega t} d\Omega d\tau \end{aligned}$$

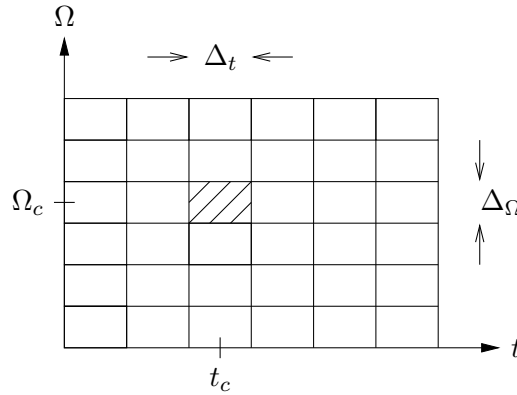
assuming real-valued $w(t)$ for which $\int |w(t)|^2 dt = 1$. The STFT can be interpreted as a “sliding window CTFT”: to calculate $X_{\text{STFT}}(\Omega, \tau)$, slide the center of window $w(t)$ to time τ , window the input signal, and compute the CTFT of the result. This is illustrated in the following figure. The idea is to isolate the signal in the vicinity of time τ , and then perform a CTFT analysis in order to estimate the “local” frequency content at time τ .



Essentially, the STFT uses the basis elements

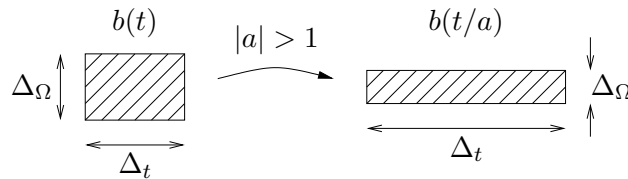
$$b_{\Omega, \tau}(t) = w(t - \tau)e^{j\Omega t}$$

over the range $t \in (-\infty, \infty)$ and $\Omega \in (-\infty, \infty)$. This can be understood as time and frequency shifts of the window function $w(t)$. The STFT basis is often illustrated by a tiling of the time-frequency plane, where each tile represents a particular basis element:



The height and width of a tile represent the spectral and temporal widths of the basis element, respectively, and the position of a tile represents the spectral and temporal centers of the basis element. Note that, while the tiling diagram above suggests that the STFT uses a discrete set of time/frequency shifts, the STFT basis is really constructed from a continuum of time/frequency shifts.

Note that we can decrease spectral width Δ_Ω at the cost of increased temporal width Δ_t by stretching basis waveforms in time, although the time-bandwidth product $\Delta_t \Delta_\Omega$ (i.e., the area of each tile) will remain constant.

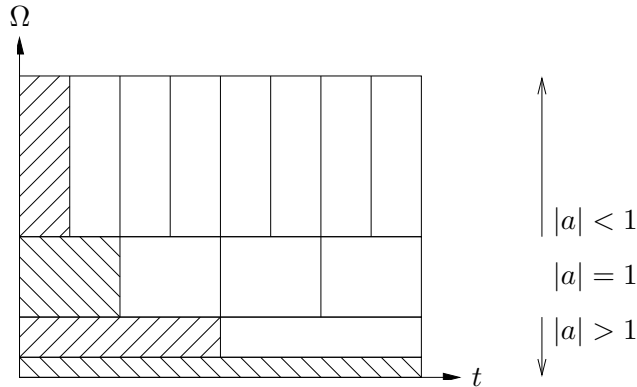


Our observations can be summarized as follows:

- The time resolutions and frequency resolutions of every STFT basis element will equal those of the window $w(t)$. (All STFT tiles have the same shape.)
 - The use of a wide window will give good frequency resolution but poor time resolution, while the use of a narrow window will give good time resolution but poor frequency resolution. (When tiles are stretched in one direction they shrink in the other.)
 - The combined time-frequency resolution of the basis, proportional to $\frac{1}{\Delta_t \Delta_\Omega}$, is determined not by window width but by window shape. The shape defined by the Gaussian window⁵ $w(t) = \sqrt{\alpha/\pi} e^{-\alpha t^2}$ (for $\alpha \in \mathbb{R}$) gives the highest time-frequency resolution, although its infinite time-support makes it impossible to implement. (The Gaussian window results in tiles with minimum area.)
- Continuous Wavelet Transform: The STFT provided a means of (joint) time-frequency analysis with the property that spectral/temporal widths (or resolutions) were the same for all basis elements. Let’s now take a closer look at the implications of uniform resolution.

Consider two signals composed of sinusoids with frequency 1 Hz and 1.001 Hz, respectively. It may be difficult to distinguish between these two signals in the presence of background noise unless many cycles are observed, implying the need for a many-second observation. Now consider two signals with pure frequencies of 1000 Hz and 1001 Hz—again, a 0.1% difference. Here it should be possible to distinguish the two signals in an interval of much less than one second. In other words, good frequency resolution requires longer observation times as frequency decreases. Thus, it might be more convenient to construct a basis whose elements have larger temporal widths at low frequencies.

The previous example motivates a multi-resolution time-frequency tiling of the form:



The Continuous Wavelet Transform (CWT) accomplishes the above multi-resolution tiling by time-shifting and time-scaling a prototype function $\psi(t)$, often called the “mother wavelet.” The a -scaled and τ -shifted basis element is given by [1]

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-\tau}{a}\right) \quad \text{where} \quad \begin{cases} a, \tau \in \mathbb{R} \\ \int_{-\infty}^{\infty} \psi(t) dt = 0 \\ C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\Omega)|^2}{|\Omega|} d\Omega < \infty \end{cases}$$

⁵The STFT using a Gaussian window is known as the Gabor Transform (1946).

The conditions above imply that $\Psi(\Omega)$ is bandpass and that $\psi(t)$ is sufficiently smooth. We assume that $\|\psi(t)\| = 1$, so that the definition above ensures $\|\psi_{a,\tau}(t)\| = 1$ for all a and τ . In this case, the CWT is defined by the transform pair

$$X_{\text{CWT}}(a, \tau) = \int_{-\infty}^{\infty} x(t)\psi_{a,\tau}^*(t) dt$$

$$x(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X_{\text{CWT}}(a, \tau)\psi_{a,\tau}(t) \frac{d\tau da}{a^2}$$

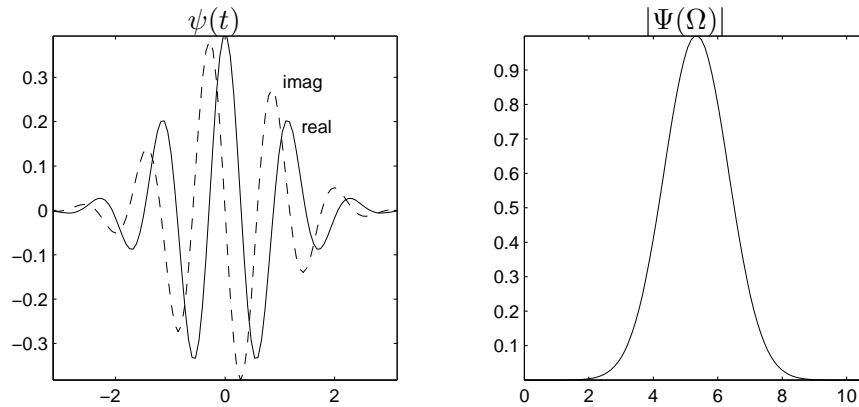
In basic terms, the CWT says that a waveform can be decomposed into a collection of shifted and stretched versions of the mother wavelet $\psi(t)$. As such, it is usually said that wavelets perform a “time-scale” analysis rather than a time-frequency analysis.

The Morlet wavelet is a classic example of the CWT. It employs a windowed complex exponential as the mother wavelet [1]:

$$\psi(t) = \frac{1}{\sqrt{2\pi}} e^{j\Omega_0 t} e^{-t^2/2}$$

$$\Psi(\Omega) = e^{-(\Omega+\Omega_0)^2/2}$$

where it is typical to select $\Omega_0 = \pi\sqrt{\frac{2}{\log 2}}$. (See illustration below.) While this wavelet does not exactly satisfy the conditions established earlier, it nearly satisfies them since $\Psi(0) \approx 7 \times 10^{-7} \neq 0$. Since the correction is minor, in practice it is often ignored.



While the CWT discussed above is an interesting theoretical and pedagogical tool, the discrete wavelet transform (DWT) is much more practical. Before shifting our focus to the DWT, we take a step back and review some of the basic concepts from the branch of mathematics known as Hilbert Space theory. These concepts will be essential in our development of the DWT. In fact, Hilbert Space theory is the mathematical foundation for nearly all of signal processing, so it may be interesting to see what this foundation provides for us.

2 An Introduction to Hilbert Space Theory

Hilbert spaces provide the mathematical foundation for signal processing theory. In this section we attempt to clearly define some key Hilbert space concepts like *vectors*, *norms*, *inner products*, *subspaces*, *orthogonality*, *orthonormal bases*, and *projections*. The intent is not to bury you in mathematics, but to familiarize you with the terminology, provide intuition, and leave you with a “lookup table” for future reference.

- Vector Space:

- A vector space consists of the following four components:

1. A set of vectors V ,
2. A field of scalars \mathbb{F} (where, for our purposes, \mathbb{F} is either \mathbb{R} or \mathbb{C}),
3. The operation of vector addition “+” (i.e., $+ : V \times V \rightarrow V$)
4. The operation of scalar multiplication “•” (i.e., $\bullet : \mathbb{F} \times V \rightarrow V$)

for which the following properties hold. (Assume $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\alpha, \beta \in \mathbb{F}$.)

- a) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (commutativity)
- b) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (associativity)
 $(\alpha\beta) \bullet \mathbf{x} = \alpha \bullet (\beta \bullet \mathbf{x})$
- c) $\alpha \bullet (\mathbf{x} + \mathbf{y}) = \alpha \bullet \mathbf{x} + \alpha \bullet \mathbf{y}$ (distributivity)
 $(\alpha + \beta) \bullet \mathbf{x} = \alpha \bullet \mathbf{x} + \beta \bullet \mathbf{x}$
- d) $\forall \mathbf{x} \in V, \exists \mathbf{0} \in V$ s.t. $\mathbf{x} + \mathbf{0} = \mathbf{x}$ (additive identity)
- e) $\forall \mathbf{x} \in V, \exists (-\mathbf{x}) \in V$ s.t. $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$ (additive inverse)
- f) $\forall \mathbf{x} \in V, 1 \bullet \mathbf{x} = \mathbf{x}$ (multiplicative identity)

Important examples of vector spaces include

- i) $V = \mathbb{R}^N, \mathbb{F} = \mathbb{R}$ (real N -vectors)
- ii) $V = \mathbb{C}^N, \mathbb{F} = \mathbb{C}$ (complex N -vectors)
- iii) $V = \{\{x[n]\} \text{ s.t. } \sum_{n=-\infty}^{\infty} |x[n]|^p < \infty\}, \mathbb{F} = \mathbb{C}$ (sequences in “ ℓ_p ”)
- iv) $V = \{f(t) : \int_{-\infty}^{\infty} |f(t)|^p dt < \infty\}, \mathbb{F} = \mathbb{C}$ (functions in “ \mathcal{L}_p ”)

where we have assumed the usual definitions of addition and multiplication. From now on, we will denote the arbitrary vector space $(V, \mathbb{F}, +, \bullet)$ by the shorthand V and assume the usual selection of $(\mathbb{F}, +, \bullet)$. We will also suppress the “•” in scalar multiplication, so that $\alpha \bullet \mathbf{x}$ becomes $\alpha\mathbf{x}$.

- A subspace of V is a vector space defined from the subset $M \subset V$ for which

1. $\forall \mathbf{x}, \mathbf{y} \in M, \mathbf{x} + \mathbf{y} \in M$
2. $\forall \mathbf{x} \in M$ and $\forall \alpha \in \mathbb{F}, \alpha\mathbf{x} \in M$

(Note that every subspace must contain $\mathbf{0}$, and that V is a subspace of itself.)

- The span of a set $S \subset V$ is the subspace of V containing all linear combinations of vectors in S . When $S = \{\mathbf{x}_0, \dots, \mathbf{x}_{N-1}\}$,

$$\text{span}(S) := \left\{ \sum_{i=0}^{N-1} \alpha_i \mathbf{x}_i : \alpha_i \in \mathbb{F} \right\}$$

- A subset of linearly-independent vectors $\{\mathbf{x}_0, \dots, \mathbf{x}_{N-1}\} \subset V$ is called a basis for V when its span equals V . In such a case, we say that V has dimension N . We say that V is infinite-dimensional⁶ if it contains an infinite number of linearly independent vectors.
- V is a direct sum of two subspaces M and N , written $V = M \oplus N$, iff every $\mathbf{x} \in V$ has a unique representation $\mathbf{x} = \mathbf{m} + \mathbf{n}$ for $\mathbf{m} \in M$ and $\mathbf{n} \in N$. (Note that this requires $M \cap N = \{\mathbf{0}\}$.)

• Normed Vector Space: Now we equip a vector space V with a notion of “size.”

- A norm is a function ($\|\cdot\| : V \rightarrow \mathbb{R}$) such that the following properties hold ($\forall \mathbf{x}, \mathbf{y} \in V$ and $\forall \alpha \in \mathbb{F}$):
 1. $\|\mathbf{x}\| \geq 0$ with equality iff $\mathbf{x} = \mathbf{0}$,
 2. $\|\alpha\mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$,
 3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, (the triangle inequality).

In simple terms, the norm measures the size of a vector. Adding the norm operation to a vector space yields a normed vector space. Important examples include:

- i) $V = \mathbb{R}^N$, $\|[x_0, \dots, x_{N-1}]^t\| := \sqrt{\sum_{i=0}^{N-1} x_i^2} = \sqrt{\mathbf{x}^t \mathbf{x}}$.
- ii) $V = \mathbb{C}^N$, $\|[x_0, \dots, x_{N-1}]^t\| := \sqrt{\sum_{i=0}^{N-1} |x_i|^2} = \sqrt{\mathbf{x}^H \mathbf{x}}$.
- ii) $V = \ell_p$, $\|\{x[n]\}\| := \left(\sum_{n=-\infty}^{\infty} |x[n]|^p\right)^{\frac{1}{p}}$.
- iv) $V = \mathcal{L}_p$, $\|f(t)\| := \left(\int_{-\infty}^{\infty} |f(t)|^p dt\right)^{\frac{1}{p}}$.

• Inner-Product Space: Next we equip a normed vector space V with a notion of “direction.”

- An inner product is a function ($\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$) such that the following properties hold ($\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\forall \alpha \in \mathbb{F}$):
 1. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^*$,
 2. $\langle \mathbf{x}, \alpha \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$... implying that $\langle \alpha \mathbf{x}, \mathbf{y} \rangle = \alpha^* \langle \mathbf{x}, \mathbf{y} \rangle$,
 3. $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$,
 4. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ with equality iff $\mathbf{x} = \mathbf{0}$.

In simple terms, the inner product measures the relative alignment between two vectors. Adding an inner product operation to a vector space yields an inner product space. Important examples include:

- i) $V = \mathbb{R}^N$, $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^t \mathbf{y}$.
- ii) $V = \mathbb{C}^N$, $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^H \mathbf{y}$.
- iii) $V = \ell_2$, $\langle \{x[n]\}, \{y[n]\} \rangle := \sum_{n=-\infty}^{\infty} x^*[n]y[n]$
- iv) $V = \mathcal{L}_2$, $\langle f(t), g(t) \rangle := \int_{-\infty}^{\infty} f^*(t)g(t)dt$.

The inner products above are the “usual” choices for those spaces.

The inner product naturally defines a norm:

$$\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle},$$

though not every norm can be defined from an inner product.⁷ Thus, an inner product

⁶The definition of an infinite-dimensional basis would be complicated by issues relating to the convergence of infinite series. Hence we postpone discussion of infinite-dimensional bases until the Hilbert Space section.

⁷An example for inner product space \mathcal{L}_2 would be any norm $\|f\| := \sqrt[p]{\int_{-\infty}^{\infty} |f(t)|^p dt}$ such that $p > 2$.

space can be considered as a normed vector space with additional structure. Assume, from now on, that we adopt the inner-product norm when given a choice.

- The Cauchy-Schwarz inequality says

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\| \quad \text{with equality iff } \exists \alpha \in \mathbb{F} \text{ s.t. } \mathbf{x} = \alpha \mathbf{y}$$

When $\langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{R}$, the inner product can be used to define an “angle” between vectors:

$$\cos(\theta) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}$$

- Vectors \mathbf{x} and \mathbf{y} are said to be orthogonal, denoted as $\mathbf{x} \perp \mathbf{y}$, when $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. The Pythagorean theorem says:

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \quad \text{when } \mathbf{x} \perp \mathbf{y}.$$

Vectors \mathbf{x} and \mathbf{y} are said to be orthonormal when $\mathbf{x} \perp \mathbf{y}$ and $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$.

- $\mathbf{x} \perp S$ means $\mathbf{x} \perp \mathbf{y}$ for all $\mathbf{y} \in S$. S is an orthogonal set if $\mathbf{x} \perp \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in S$ s.t. $\mathbf{x} \neq \mathbf{y}$. An orthogonal set S is an orthonormal set if $\|\mathbf{x}\| = 1$ for all $\mathbf{x} \in S$. Some examples of orthonormal sets are

i) $\mathbb{R}^3 : S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$,

ii) \mathbb{C}^N : Subsets of columns from unitary matrices,

iii) ℓ_2 : Subsets of shifted Kronecker delta sequences $S \subset \{\{\delta[n-k]\} : k \in \mathbb{Z}\}$,

iv) \mathcal{L}_2 : $S = \left\{ \frac{1}{\sqrt{T}} p_T(t-nT) : n \in \mathbb{Z} \right\}$ for T -wide pulse $p_T(t) = (u(t) - u(t-T))$ and unit step $u(t)$.

where in each case we assume the usual inner product.

- Hilbert Space: Now we consider inner product spaces with nice convergence properties that allow us to define countably-infinite orthonormal bases.

- A Hilbert space is a complete inner product space. A complete space⁸ is one where all Cauchy sequences converge to some vector within the space. For sequence $\{\mathbf{x}_n\}$ to be Cauchy, the distance between its elements must eventually become arbitrarily small:

$$\forall \epsilon > 0, \exists N_\epsilon \text{ s.t. } \forall n, m \geq N_\epsilon, \|\mathbf{x}_n - \mathbf{x}_m\| < \epsilon$$

For a sequence $\{\mathbf{x}_n\}$ to be convergent to \mathbf{x} , the distance between its elements and \mathbf{x} must eventually become arbitrarily small:

$$\forall \epsilon > 0, \exists N_\epsilon \text{ s.t. } \forall n \geq N_\epsilon, \|\mathbf{x}_n - \mathbf{x}\| < \epsilon$$

Examples are listed below (assuming the usual inner products):

i) $V = \mathbb{R}^N$,

ii) $V = \mathbb{C}^N$,

⁸The rational numbers provide an example of an incomplete set. We know that it is possible to construct a sequence of rational numbers which approximate an irrational number arbitrarily closely. It is easy to see that such a sequence will be Cauchy. However, the sequence will not converge to any *rational* number, and so the rationals cannot be complete.

- iii) $V = \ell_2$ (i.e., square summable sequences),
 - iv) $V = \mathcal{L}_2$ (i.e., square integrable functions).
- We will always deal with separable Hilbert spaces, which are those that have a countable⁹ orthonormal (ON) basis. A countable orthonormal basis for V is a countable orthonormal set $S = \{\mathbf{x}_k\}$ such that every vector in V can be represented as a linear combination of elements in S :

$$\forall \mathbf{y} \in V, \exists \{\alpha_k\} \text{ s.t. } \mathbf{y} = \sum_k \alpha_k \mathbf{x}_k.$$

Due to the orthonormality of S , the basis coefficients are given by

$$\alpha_k = \langle \mathbf{x}_k, \mathbf{y} \rangle.$$

We can see this via:

$$\langle \mathbf{x}_k, \mathbf{y} \rangle = \left\langle \mathbf{x}_k, \lim_{n \rightarrow \infty} \sum_{i=0}^n \alpha_i \mathbf{x}_i \right\rangle = \lim_{n \rightarrow \infty} \left\langle \mathbf{x}_k, \sum_{i=0}^n \alpha_i \mathbf{x}_i \right\rangle = \lim_{n \rightarrow \infty} \sum_{i=0}^n \alpha_i \underbrace{\langle \mathbf{x}_k, \mathbf{x}_i \rangle}_{\delta[k-i]} = \alpha_k$$

(where the second equality invokes the continuity of the inner product).

In finite n -dimensional spaces (e.g., \mathbb{R}^n or \mathbb{C}^n), any n -element ON set constitutes an ON basis. In infinite-dimensional spaces, we have the following *equivalences* concerning an orthonormal set $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots\} \in V$:

- a) $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots\}$ is an ON basis for V .
- b) If $\langle \mathbf{x}_i, \mathbf{y} \rangle = 0$ for all i , then $\mathbf{y} = \mathbf{0}$.
- c) $\forall \mathbf{y} \in V, \|\mathbf{y}\|^2 = \sum_i |\langle \mathbf{x}_i, \mathbf{y} \rangle|^2$. (Parseval's Equality)
- d) Every $\mathbf{y} \in V$ is a limit of some sequence of vectors in $\text{span}(\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots\})$.

Examples of countable ON bases for various Hilbert spaces include:

- i) \mathbb{R}^N : $\{\mathbf{e}_0, \dots, \mathbf{e}_{N-1}\}$ for $\mathbf{e}_i = [0, \dots, 0, 1, 0, \dots, 0]^t$ with “1” in the i^{th} position,
 - ii) \mathbb{C}^N : the columns of any unitary $N \times N$ matrix.
 - iii) ℓ_2 : $\{\{\delta_i[n]\} : i \in \mathbb{Z}\}$, for $\delta_i[n] := \delta[n-i]$ (all shifts of the Kronecker sequence)
 - iv) \mathcal{L}_2 : to be constructed using the DWT...
- Say S is a subspace of Hilbert space V . The orthogonal complement of S in V , denoted S^\perp , is the subspace defined by the set $\{\mathbf{x} \in V : \mathbf{x} \perp S\}$. Assuming that S is closed, we can write $V = S \oplus S^\perp$.
- The orthogonal projection of \mathbf{y} onto S , where S is a closed subspace of V , is

$$\hat{\mathbf{y}} = \sum_i \langle \mathbf{x}_i, \mathbf{y} \rangle \mathbf{x}_i \text{ s.t. } \{\mathbf{x}_i\} \text{ is an ON basis for } S.$$

Orthogonal projection yields the best approximation of \mathbf{y} in S :

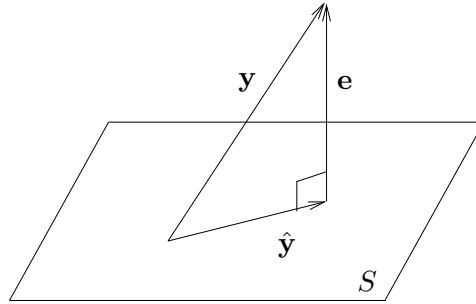
$$\hat{\mathbf{y}} = \arg \min_{\mathbf{x} \in S} \|\mathbf{y} - \mathbf{x}\|.$$

⁹A countable set is a set with at most a countably-infinite number of elements. Finite sets are countable, as are any sets whose elements can be organized into an infinite list. Continuum (e.g., intervals of \mathbb{R}) are uncountably infinite.

The approximation error $\mathbf{e} := \mathbf{y} - \hat{\mathbf{y}}$ obeys the orthogonality principle:

$$\mathbf{e} \perp S.$$

We illustrate this concept using $V = \mathbb{R}^3$ below but stress that the same geometrical interpretation applies to any Hilbert space.



A proof of the orthogonality principle is:

$$\begin{aligned}
 & \mathbf{e} \perp S \\
 \Leftrightarrow & \langle \mathbf{e}, \mathbf{x}_i \rangle = 0 \quad \forall i \\
 \Leftrightarrow & \langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{x}_i \rangle = 0 \quad \forall i \\
 \Leftrightarrow & \langle \mathbf{y}, \mathbf{x}_i \rangle = \langle \hat{\mathbf{y}}, \mathbf{x}_i \rangle \quad \forall i \\
 & = \left\langle \sum_j \langle \mathbf{x}_j, \mathbf{y} \rangle \mathbf{x}_j, \mathbf{x}_i \right\rangle \quad \forall i \\
 & = \sum_j \langle \mathbf{x}_j, \mathbf{y} \rangle^* \langle \mathbf{x}_j, \mathbf{x}_i \rangle \quad \forall i \\
 & = \sum_j \langle \mathbf{y}, \mathbf{x}_j \rangle \delta_{j-i} \quad \forall i \\
 & = \langle \mathbf{y}, \mathbf{x}_i \rangle \quad \forall i \quad \square
 \end{aligned}$$

3 Discrete Wavelet Transform

- Main Concepts: The *discrete wavelet transform* (DWT) is a representation of a signal $x(t) \in \mathcal{L}_2$ using an orthonormal basis consisting of a countably-infinite set of *wavelets*. Denoting the wavelet basis as $\{\psi_{k,n}(t) : k, n \in \mathbb{Z}\}$, the DWT transform pair is

$$x(t) = \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} d_{k,n} \psi_{k,n}(t)$$

$$d_{k,n} = \langle \psi_{k,n}(t), x(t) \rangle = \int_{-\infty}^{\infty} \psi_{k,n}^*(t) x(t) dt$$

where $\{d_{k,n}\}$ are the wavelet coefficients. Note the relationship to Fourier series and to the sampling theorem: in both cases we can perfectly describe a continuous-time signal $x(t)$ using a countably-infinite (i.e., discrete) set of coefficients. Specifically, Fourier series enabled us to describe *periodic* signals using Fourier coefficients $\{X[k] : k \in \mathbb{Z}\}$, while the sampling theorem enabled us to describe *bandlimited* signals using signal samples $\{x[n] : n \in \mathbb{Z}\}$. In both cases, signals within a limited class are represented using a coefficient set with a single countable index. The DWT can describe *any* signal in \mathcal{L}_2 using a coefficient set parameterized by two countable indices: $\{d_{k,n} : k, n \in \mathbb{Z}\}$.

Wavelets are orthonormal functions in \mathcal{L}_2 obtained by shifting and stretching a “mother wavelet” $\psi(t) \in \mathcal{L}_2$. For example,

$$\psi_{k,n}(t) = 2^{-k/2} \psi(2^{-k}t - n), \quad k, n \in \mathbb{Z}$$

defines a family of wavelets $\{\psi_{k,n}(t) : k, n \in \mathbb{Z}\}$ related by power-of-two (or “dyadic”) stretches. As k increases, the wavelet stretches by a factor two; as n increases, the wavelet shifts right. Note that when $\|\psi(t)\| = 1$, the normalization ensures that $\|\psi_{k,n}(t)\| = 1$ for all $k, n \in \mathbb{Z}$. Power-of-two stretching is a convenient, though somewhat arbitrary, choice. In our treatment of the discrete wavelet transform, however, we will focus on this choice. Even with power-of-two stretches, there are various possibilities for $\psi(t)$, each giving a different flavor of DWT.

Wavelets are constructed so that $\{\psi_{k,n}(t) : n \in \mathbb{Z}\}$ (i.e., the set of all shifted wavelets at fixed scale k) describes a particular level of “detail” in the signal. As k becomes smaller (i.e., closer to $-\infty$), the wavelets become more “fine grained” and the level of detail increases. In this way, the DWT can give a *multi-resolution* description of a signal, very useful in analyzing “real-world” signals. Essentially, the DWT gives us a *discrete multi-resolution description of a continuous-time signal in \mathcal{L}_2* .

In the modules that follow, these DWT concepts will be developed “from scratch” using Hilbert space principles. To aid the development, we make use of the so-called “scaling function” $\phi(t) \in \mathcal{L}_2$, which will be used to approximate the signal *up to a particular level of detail*. Like with wavelets, a family of scaling functions can be constructed via shifts and power-of-two stretches

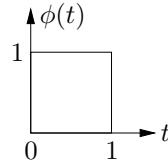
$$\phi_{k,n}(t) = 2^{-k/2} \phi(2^{-k}t - n), \quad k, n \in \mathbb{Z}$$

given mother scaling function $\phi(t)$. The relationships between wavelets and scaling functions will be elaborated upon below via theory and example. See [1, Ch. 4] for a slightly different development.

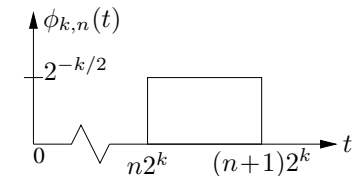
Note that the inner-product expression for $d_{k,n}$ above is written for the general complex-valued case. In our treatment of the discrete wavelet transform, however, we will assume real-valued signals and wavelets, for simplicity. For this reason, we omit the complex conjugations in the remainder of our DWT discussions.

- Example: The Haar System: The Haar basis is perhaps the simplest example of a DWT basis, and we will frequently refer to it in our DWT development. Keep in mind, however, that *the Haar basis is only an example*; there are many other ways of constructing a DWT decomposition.

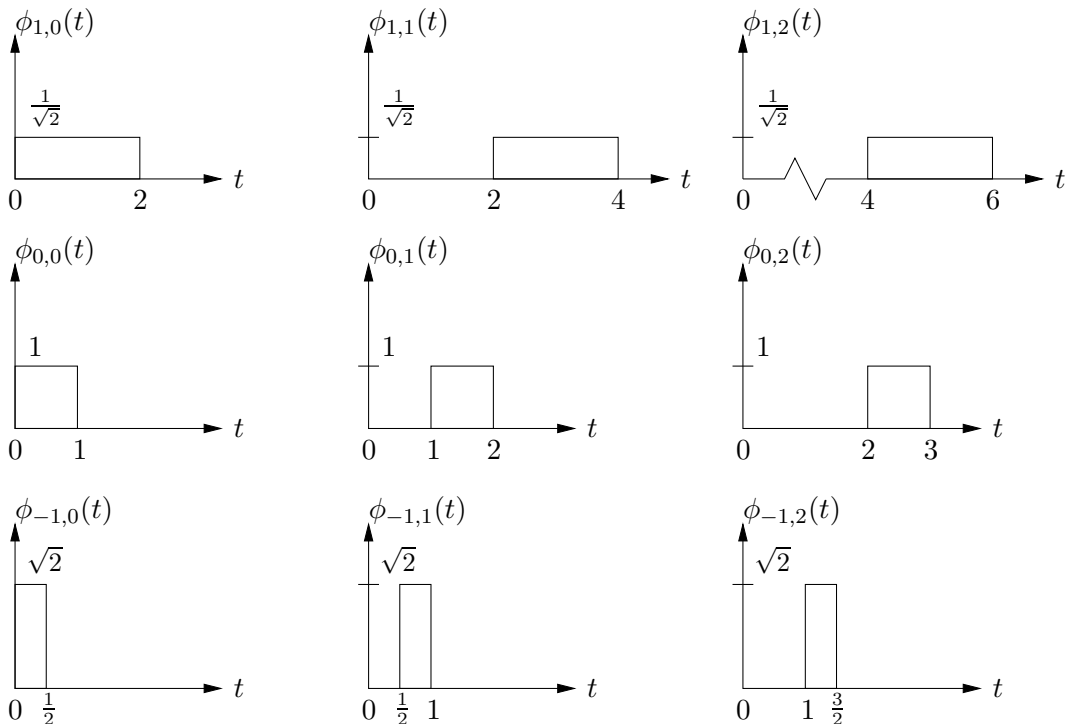
For the Haar case, the mother scaling function is defined by

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{else} \end{cases} .$$


From the mother scaling function, we define a family of shifted and stretched scaling functions $\{\phi_{k,n}(t)\}$ according to

$$\begin{aligned} \phi_{k,n}(t) &= 2^{-k/2} \phi(2^{-k}t - n), \quad k, n \in \mathbb{Z}, \\ &= 2^{-k/2} \phi\left(\frac{1}{2^k}(t - n2^k)\right), \end{aligned}$$


which are illustrated below for various k and n . The second equation above makes clear the principle that incrementing n by one shifts the pulse one place to the right. Observe from the figures that $\{\phi_{k,n}(t) : n \in \mathbb{Z}\}$ is orthonormal for each k (i.e., along each row of figures).



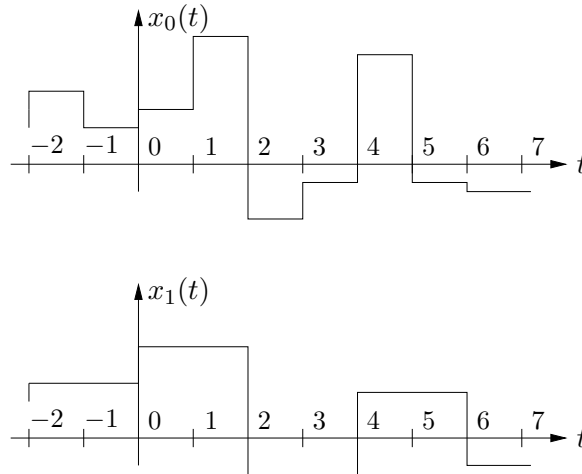
- A Hierarchy of Detail: Given a mother scaling function $\phi(t) \in \mathcal{L}_2$ —the choice of which will be discussed later—we construct scaling functions at “coarseness-level k ” and “shift n ” as follows:

$$\phi_{k,n}(t) := 2^{-k/2}\phi(2^{-k}t - n).$$

We then use V_k to denote the subspace defined by linear combinations of scaling functions at the k^{th} level:

$$V_k := \text{span}\{\phi_{k,n}(t) : n \in \mathbb{Z}\}.$$

In the Haar system, for example, V_0 and V_1 consist of signals with the characteristics of $x_0(t)$ and $x_1(t)$ illustrated below, respectively.

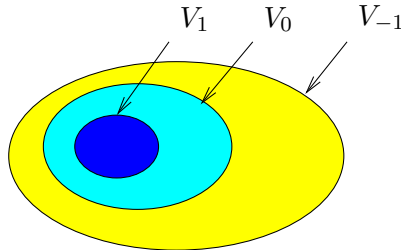


We assume that $\phi(t)$ ensures the following nesting property is satisfied:

$$\begin{array}{ccccccc} \cdots & V_2 & \subset & V_1 & \subset & V_0 & \subset & V_{-1} & \subset & V_{-2} & \cdots \\ & \text{coarse} & \longleftarrow & & & & & & & \longrightarrow & \text{detailed} \end{array}$$

In other words, any signal in V_k can be constructed as a linear combination of “more detailed” signals in V_{k-1} . (The Haar system gives proof that at least one such $\phi(t)$ exists.)

The nesting property can be depicted using the set-theoretic diagram below, where V_{-1} is represented by the contents of the largest egg (which includes the smaller two eggs), V_0 is represented by the contents of the medium-sized egg (which includes the smallest egg), and V_1 is represented by the contents of the smallest egg.



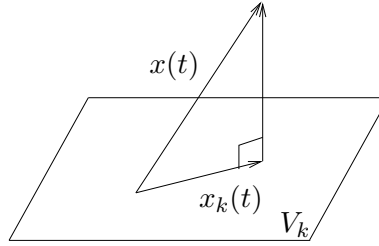
Furthermore, we will assume that

$$\begin{aligned} \lim_{k \rightarrow -\infty} V_k &= \mathcal{L}_2 \text{ (i.e., contains all signals)} \\ \lim_{k \rightarrow +\infty} V_k &= \{\mathbf{0}\} \text{ (i.e., contains only the zero signal)} \\ \{\phi(t - n) : n \in \mathbb{Z}\} &= \text{an orthonormal set.} \end{aligned}$$

Essentially, $\lim_{k \rightarrow -\infty} V_k = \mathcal{L}_2$ means that, for any $x(t) \in \mathcal{L}_2$ and any maximally tolerable error $\epsilon > 0$, we can find an approximation level k such that $\int_{-\infty}^{\infty} |x(t) - x_k(t)|^2 dt < \epsilon$. Note that, while it might at first seem that $\lim_{k \rightarrow -\infty} V_k$ should contain all non-zero constant signals (e.g., $x(t) = a$ for $a \in \mathbb{R}$), the only constant signal in \mathcal{L}_2 , the space of square-integrable signals, is the zero signal.

Note that orthonormal $\{\phi_{0,n}(t) : n \in \mathbb{Z}\}$ implies orthonormal $\{\phi_{k,n}(t) : n \in \mathbb{Z}\}$ for all $k \in \mathbb{Z}$. Since $\{\phi_{k,n}(t) : n \in \mathbb{Z}\}$ is an orthonormal basis for V_k , the best (in \mathcal{L}_2 norm) approximation of $x(t) \in \mathcal{L}_2$ at coarseness-level k (i.e., in V_k) is given by the orthogonal projection

$$\begin{aligned} x_k(t) &= \sum_{n=-\infty}^{\infty} c_{k,n} \phi_{k,n}(t) \\ c_{k,n} &= \langle \phi_{k,n}(t), x(t) \rangle \end{aligned}$$



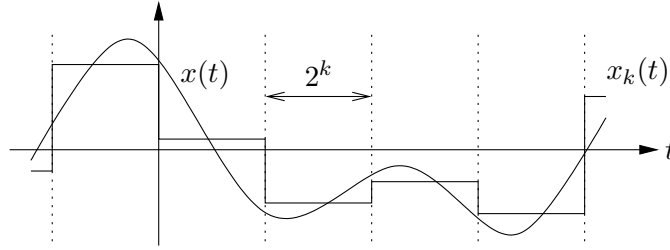
- Haar Approximation in V_k : It is instructive to consider the approximation of signal $x(t) \in \mathcal{L}_2$ at coarseness-level k of the Haar system. For the Haar case, projection of $x(t) \in \mathcal{L}_2$ onto V_k is accomplished using the basis coefficients

$$\begin{aligned} c_{k,n} &= \int_{-\infty}^{\infty} \phi_{k,n}(t) x(t) dt \\ &= \int_{n2^k}^{(n+1)2^k} 2^{-k/2} x(t) dt \end{aligned}$$

giving the approximation

$$\begin{aligned} x_k(t) &= \sum_{n=-\infty}^{\infty} c_{k,n} \phi_{k,n}(t) \\ &= \sum_{n=-\infty}^{\infty} \left(\int_{n2^k}^{(n+1)2^k} 2^{-k/2} x(t) dt \right) \phi_{k,n}(t) \\ &= \sum_{n=-\infty}^{\infty} \underbrace{\left(\frac{1}{2^k} \int_{n2^k}^{(n+1)2^k} x(t) dt \right)}_{\text{average value of } x(t) \text{ in interval}} \underbrace{2^{k/2} \phi_{k,n}(t)}_{\text{height} = 1 \ \forall k}. \end{aligned}$$

This corresponds to taking the average value of the signal in each interval of width 2^k and approximating the function by a constant over that interval. (See below.)



- The Scaling Equation: Consider the level-1 subspace and its orthonormal basis:

$$V_1 = \text{span}\{\phi_{1,n}(t) : n \in \mathbb{Z}\}$$

$$\phi_{1,n}(t) = \frac{1}{\sqrt{2}}\phi\left(\frac{1}{2}t - n\right)$$

Since $V_1 \subset V_0$ (i.e., V_0 is more detailed than V_1) and since $\phi_{1,0}(t) \in V_0$, there must exist coefficients $\{h[n] : n \in \mathbb{Z}\}$ such that

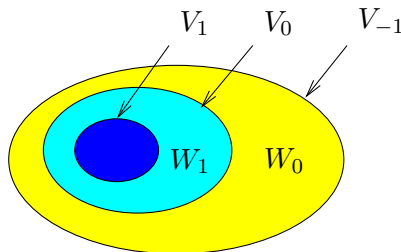
$$\begin{aligned} \phi_{1,0}(t) &= \sum_{n=-\infty}^{\infty} h[n]\phi_{0,n}(t) \\ \Leftrightarrow \frac{1}{\sqrt{2}}\phi\left(\frac{1}{2}t\right) &= \sum_{n=-\infty}^{\infty} h[n]\phi(t-n) \\ \Leftrightarrow \boxed{\phi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} h[n]\phi(2t-n)} \end{aligned}$$

This is known as the “scaling equation.” To be a valid scaling function, $\phi(t)$ must obey the scaling equation for some coefficient set $\{h[n]\}$.

- The Wavelet Scaling Equation: The *difference* in detail between V_k and V_{k-1} will be described using W_k , defined as the orthogonal complement of V_k in V_{k-1} :

$$V_{k-1} = V_k \oplus W_k.$$

At times it will be convenient to write $W_k = V_k^\perp$. This concept is illustrated in the set-theoretic diagram below.



Suppose now, that there exists¹⁰ a “mother wavelet” $\psi(t) \in \mathcal{L}_2$ such that $\{\psi(t-n) : n \in \mathbb{Z}\}$ constitutes an orthonormal basis for W_0 . Because every V_k has an orthonormal basis

¹⁰Unfortunately, proving the existence of this $\psi(t)$, in the case of general $\phi(t)$, is outside the scope of this course; the interested reader is referred to [2, pp. 134-135] for these details. It should be noted that, while such $\psi(t)$ do exist, they are not unique.

$\{\phi_{k,n}(t) : n \in \mathbb{Z}\}$ constructed from shifts and stretches of the mother scaling function $\phi(t)$, it is easily shown that every W_k will have an orthonormal basis $\{\psi_{k,n}(t) : n \in \mathbb{Z}\}$ whose elements are constructed from shifts and stretches of the mother wavelet:

$$\psi_{k,n}(t) = 2^{-k/2}\psi(2^{-k}t - n).$$

The Haar system will soon provide us with a concrete example.

Let's focus, for the moment, on the specific case $k = 1$. Since $W_1 \subset V_0$, there must exist $\{g[n] : n \in \mathbb{Z}\}$ such that

$$\begin{aligned} \psi_{1,0}(t) &= \sum_{n=-\infty}^{\infty} g[n]\phi_{0,n}(t) \\ \Leftrightarrow \frac{1}{\sqrt{2}}\psi\left(\frac{1}{2}t\right) &= \sum_{n=-\infty}^{\infty} g[n]\phi(t - n) \\ \Leftrightarrow \boxed{\psi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} g[n]\phi(2t - n)} \end{aligned}$$

This is known as the “wavelet scaling equation.” To be a valid scaling-function/wavelet pair, $\phi(t)$ and $\psi(t)$ must obey the wavelet scaling equation for some coefficient set $\{g[n]\}$.

- Conditions on $\{h[n]\}$ and $\{g[n]\}$: Here we derive sufficient conditions on the coefficients $\{h[n]\}$ and $\{g[n]\}$ used in the scaling equation and wavelet scaling equation to ensure, for every $k \in \mathbb{Z}$, that the sets $\{\phi_{k,n}(t) : n \in \mathbb{Z}\}$ and $\{\psi_{k,n}(t) : n \in \mathbb{Z}\}$ have the desired orthonormality properties previously discussed. We will make repeated use of our requirement that $\{\phi(t - n) : n \in \mathbb{Z}\}$ is an orthonormal set.

For $\{\phi_{k,n}(t) : n \in \mathbb{Z}\}$ to be orthonormal at all k , we clearly need orthonormality when $k = 1$. This is equivalent to

$$\begin{aligned} \delta[m] &= \langle \phi_{1,0}(t), \phi_{1,m}(t) \rangle \\ &= \left\langle \sum_n h[n]\phi(t - n), \sum_\ell h[\ell]\phi(t - \ell - 2m) \right\rangle \\ &= \sum_n h[n] \sum_\ell h[\ell] \underbrace{\langle \phi(t - n), \phi(t - \ell - 2m) \rangle}_{\delta[n - (\ell + 2m)]} \\ \Leftrightarrow \boxed{\delta[m] = \sum_{n=-\infty}^{\infty} h[n]h[n - 2m]} \end{aligned}$$

Above we used the fact that $\phi_{1,m}(t) = \frac{1}{\sqrt{2}}\phi\left(\frac{1}{2}t - m\right) = \frac{1}{\sqrt{2}}\phi\left(\frac{1}{2}(t - 2m)\right) = \phi_{1,0}(t - 2m)$. So, given orthonormality at level $k = 0$, we have just derived a condition on $\{h[n]\}$ which is necessary and sufficient for orthonormality at level $k = 1$. Yet the same condition is

necessary and sufficient for orthonormality at level $k = 2$:

$$\begin{aligned}
\delta[m] &= \langle \phi_{2,0}(t), \phi_{2,m}(t) \rangle \\
&= \left\langle \sum_n h[n] \phi_{1,n}(t), \sum_\ell h[\ell] \phi_{1,\ell+2m}(t) \right\rangle \\
&= \sum_n h[n] \sum_\ell h[\ell] \underbrace{\langle \phi_{1,n}(t), \phi_{1,\ell+2m}(t) \rangle}_{\delta[n-(\ell+2m)]} \\
&= \sum_{n=-\infty}^{\infty} h[n] h[n-2m].
\end{aligned}$$

Using induction, we conclude that the previous condition will be necessary and sufficient for orthonormality of $\{\phi_{k,n}(t) : n \in \mathbb{Z}\}$ for all $k \in \mathbb{Z}$.

There is an interesting frequency-domain interpretation of this condition. If we define

$$p[m] = h[m] * h[-m] = \sum_n h[n] h[n-m]$$

then we see that our condition is equivalent to $p[2m] = \delta[m]$. In the z -domain, this yields the pair of conditions

$$\begin{cases} P(z) = H(z)H(z^{-1}) \\ 1 = \frac{1}{2} \sum_{i=0}^1 P(z^{1/2} e^{j\frac{2\pi}{2}i}) = \frac{1}{2}P(z^{1/2}) + \frac{1}{2}P(-z^{1/2}) \end{cases}$$

Putting these together,

$$\begin{aligned}
2 &= H(z^{1/2})H(z^{-1/2}) + H(-z^{1/2})H(-z^{-1/2}) \\
\Leftrightarrow &\boxed{2 = H(z)H(z^{-1}) + H(-z)H(-z^{-1})} \\
\Leftrightarrow &2 = |H(e^{j\omega})|^2 + |H(e^{j(\pi-\omega)})|^2
\end{aligned}$$

where the last property invokes the fact that $h[n] \in \mathbb{R}$ and that real-valued impulse responses yield conjugate-symmetric DTFs. Thus we find that $\{h[n]\}$ is the impulse response of a power-symmetric filter. Recall that this property was also shared by the analysis filters in an orthogonal perfect-reconstruction FIR filterbank.

To find conditions on $\{g[n]\}$ ensuring that the set $\{\psi_{k,n}(t) : n \in \mathbb{Z}\}$ is orthonormal at every k , we can repeat the steps above but with $g[n]$ replacing $h[n]$, $\psi_{1,n}(t)$ replacing $\phi_{1,n}(t)$, and the wavelet-scaling equation replacing the scaling equation. This yields

$$\boxed{
\begin{aligned}
\delta[m] &= \sum_{n=-\infty}^{\infty} g[n]g[n-2m] \\
\Leftrightarrow 2 &= G(z)G(z^{-1}) + G(-z)G(-z^{-1})
\end{aligned}
}$$

Next we derive a condition which guarantees that $W_k \perp V_k$, as required by our definition $W_k = V_k^\perp$, for all $k \in \mathbb{Z}$. Note that, for any $k \in \mathbb{Z}$, $W_k \perp V_k$ is guaranteed by $\{\psi_{k,n} : n \in \mathbb{Z}\}$

$\mathbb{Z}\} \perp \{\phi_{k,n} : n \in \mathbb{Z}\}$ which is equivalent to

$$\begin{aligned}
\forall m, 0 &= \langle \psi_{k+1,0}(t), \phi_{k+1,m}(t) \rangle \\
&= \left\langle \sum_n g[n] \phi_{k,n}(t), \sum_\ell h[\ell] \phi_{k,\ell+2m}(t) \right\rangle \\
&= \sum_n g[n] \sum_\ell h[\ell] \underbrace{\langle \phi_{k,n}(t), \phi_{k,\ell+2m}(t) \rangle}_{\delta[n-(\ell+2m)]} \\
&= \sum_n g[n] h[n-2m].
\end{aligned}$$

In other words, a 2-downsampled version of $g[n] * h[-n]$ must consist only of zeros. This necessary and sufficient condition can be restated in the frequency domain as

$$\begin{aligned}
0 &= \downarrow_2 \{G(z)H(z^{-1})\} \\
\Leftrightarrow 0 &= \frac{1}{2} \sum_{p=0}^1 G(z^{1/2} e^{-j\frac{2\pi}{2}p}) H(z^{-1/2} e^{j\frac{2\pi}{2}p}) \\
\Leftrightarrow 0 &= G(z^{1/2})H(z^{-1/2}) + G(-z^{1/2})H(-z^{-1/2}) \\
\Leftrightarrow 0 &= G(z)H(z^{-1}) + G(-z)H(-z^{-1}).
\end{aligned}$$

The choice

$$\boxed{G(z) = \pm z^{-P} H(-z^{-1}) \text{ for odd } P}$$

is sufficient (but perhaps not necessary) to satisfy our condition, since

$$\begin{aligned}
G(z)H(z^{-1}) + G(-z)H(-z^{-1}) &= \pm z^{-P} H(-z^{-1})H(z^{-1}) \mp z^{-P} H(z^{-1})H(-z^{-1}) \\
&= 0.
\end{aligned}$$

In the time domain, the condition on $G(z)$ and $H(z)$ can be expressed

$$\boxed{g[n] = \pm (-1)^n h[P-n] \text{ for odd } P}.$$

Recall that this property was satisfied by the analysis filters in an orthogonal perfect reconstruction FIR filterbank.

Finally, note that the two conditions $2 = H(z)H(z^{-1}) + H(-z)H(-z^{-1})$ and $G(z) = \pm z^{-P} H(-z^{-1})$ for odd P imply $2 = G(z)G(z^{-1}) + G(-z)G(-z^{-1})$, and thus the latter condition on $G(z)$ is redundant. To conclude,

$$\begin{aligned}
G(z) &= \pm z^{-P} H(-z^{-1}) \text{ for odd } P \\
2 &= H(z)H(z^{-1}) + H(-z)H(-z^{-1})
\end{aligned}$$

are *sufficient* to ensure that both $\{\phi_{k,n}(t) : n \in \mathbb{Z}\}$ and $\{\psi_{k,n}(t) : n \in \mathbb{Z}\}$ are orthonormal for all k and that $W_k \perp V_k$ for all k .

- Values of $\{g[n]\}$ and $\{h[n]\}$ for the Haar System: The scaling equation was originally written as

$$\phi_{1,0}(t) = \sum_n h[n] \phi_{0,n}(t).$$

The previous equation leads to a clever trick:

$$\begin{aligned}
 \langle \phi_{0,m}(t), \phi_{1,0}(t) \rangle &= \left\langle \phi_{0,m}(t), \sum_n h[n] \phi_{0,n}(t) \right\rangle \\
 &= \sum_n h[n] \underbrace{\langle \phi_{0,m}(t), \phi_{0,n}(t) \rangle}_{\delta[n-m]} \\
 &= h[m].
 \end{aligned}$$

In other words, we have a way to calculate the coefficients $\{h[m]\}$ if we know $\phi(t)$.

In the Haar case

$$\begin{aligned}
 h[m] &= \int_{-\infty}^{\infty} \phi_{0,m}(t) \phi_{1,0}(t) dt \\
 &= \int_m^{m+1} \phi_{1,0}(t) dt \\
 &= \begin{cases} \frac{1}{\sqrt{2}} & m \in \{0, 1\} \\ 0 & \text{else} \end{cases}
 \end{aligned}$$

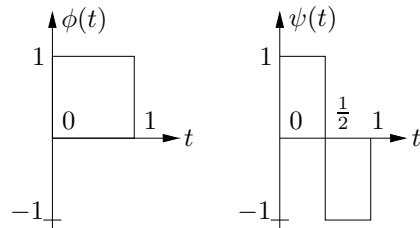
since $\phi_{1,0}(t) = \frac{1}{\sqrt{2}}$ in the interval $[0, 2)$ and zero otherwise. Then, choosing $P = 1$ in $g[n] = (-1)^n h[P - n]$, we find that

$$g[n] = \begin{cases} \frac{1}{\sqrt{2}} & n = 0 \\ -\frac{1}{\sqrt{2}} & n = 1 \\ 0 & \text{else} \end{cases}$$

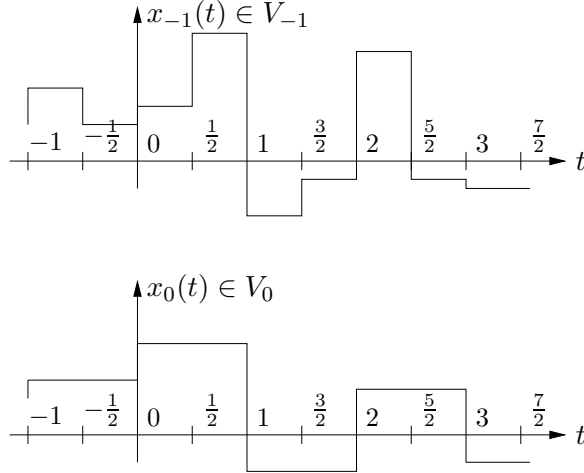
for the Haar system. From the wavelet scaling equation

$$\psi(t) = \sqrt{2} \sum_n g[n] \phi(2t - n) = \phi(2t) - \phi(2t - 1)$$

we can see that the Haar mother wavelet and scaling function look like:



In the Haar case, it is now easy to see, since $V_{-1} = V_0 \oplus W_0$, how integer shifts of the mother wavelet describe the differences between signals in V_{-1} and V_0 :



Note also that, as k decreases towards $-\infty$, the Haar system can be used to describe more and more detailed signals. In fact, any signal in \mathcal{L}_2 can be represented arbitrarily closely (in terms of the \mathcal{L}_2 norm) by choosing $-k$ large enough. In the limit $k \rightarrow -\infty$, the Haar system can be used to represent any signal in \mathcal{L}_2 .

- Wavelets: A Countable Orthonormal Basis for \mathcal{L}_2 : Recall that $V_k = W_{k+1} \oplus V_{k+1}$ and that $V_{k+1} = W_{k+2} \oplus V_{k+2}$. Putting these together and extending the idea yields

$$\begin{aligned}
 V_k &= W_{k+1} \oplus W_{k+2} \oplus V_{k+2} \\
 &= W_{k+1} \oplus W_{k+2} \oplus \cdots \oplus W_\ell \oplus V_\ell, \quad \ell > k+2 \\
 &= W_{k+1} \oplus W_{k+2} \oplus W_{k+3} \oplus \cdots \\
 &= \bigoplus_{i=k+1}^{\infty} W_i
 \end{aligned}$$

If we take the limit as $k \rightarrow -\infty$, we find that

$$V_{-\infty} := \lim_{k \rightarrow -\infty} V_k = \bigoplus_{i=-\infty}^{\infty} W_i$$

Moreover,

$$\begin{array}{lll}
 W_1 \perp V_1 & \text{and} & W_{k \geq 2} \subset V_1 \Rightarrow W_1 \perp W_{k \geq 2} \\
 W_2 \perp V_2 & \text{and} & W_{k \geq 3} \subset V_2 \Rightarrow W_2 \perp W_{k \geq 3} \\
 \vdots & & \vdots
 \end{array}$$

from which it follows that

$$W_k \perp W_{j \neq k},$$

or, in other words, all subspaces W_k are orthogonal to one another. Since the functions $\{\psi_{k,n}(t) : n \in \mathbb{Z}\}$ form an orthonormal basis for W_k , the results above imply that

$$\{\psi_{k,n}(t) : n, k \in \mathbb{Z}\} \text{ constitutes an orthonormal basis for } V_{-\infty}.$$

Given our assumptions about $\phi(t)$, we will have $V_{-\infty} = \mathcal{L}_2$, in which case

$$\boxed{\{\psi_{k,n}(t) : n, k \in \mathbb{Z}\} \text{ constitutes an orthonormal basis for } \mathcal{L}_2}$$

so that, for any $x(t) \in \mathcal{L}_2$, we can write

$$\begin{aligned} x(t) &= \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} d_k[n] \psi_{k,n}(t) \\ d_k[n] &= \langle \psi_{k,n}(t), x(t) \rangle. \end{aligned}$$

This is the key idea behind the orthogonal DWT system that we have been developing!

- Relationship to Filterbanks: Assume that we start with a signal $x(t) \in \mathcal{L}_2$. Denote the best approximation at the 0^{th} level of coarseness by $x_0(t)$. (Recall that $x_0(t)$ is the orthogonal projection of $x(t)$ onto V_0 .) Our goal, for the moment, is to decompose $x_0(t)$ into scaling coefficients and wavelet coefficients at higher levels. Since $x_0(t) \in V_0$ and $V_0 = V_1 \oplus W_1$, there exist coefficients $\{c_0[n]\}$, $\{c_1[n]\}$, and $\{d_1[n]\}$ such that

$$\begin{aligned} x_0(t) &= \sum_n c_0[n] \phi_{0,n}(t) \\ &= \sum_n c_1[n] \phi_{1,n}(t) + \sum_n d_1[n] \psi_{1,n}(t). \end{aligned}$$

Using the fact that $\{\phi_{1,n}(t) : n \in \mathbb{Z}\}$ is an orthonormal basis for V_1 , in conjunction with the scaling equation,

$$\begin{aligned} c_1[n] &= \langle x_0(t), \phi_{1,n}(t) \rangle \\ &= \left\langle \sum_m c_0[m] \phi_{0,m}(t), \phi_{1,n}(t) \right\rangle \\ &= \sum_m c_0[m] \langle \phi_{0,m}(t), \phi_{1,n}(t) \rangle \\ &= \sum_m c_0[m] \left\langle \phi(t-m), \sum_{\ell} h[\ell] \phi(t-\ell-2n) \right\rangle \\ &= \sum_m c_0[m] \sum_{\ell} h[\ell] \underbrace{\langle \phi(t-m), \phi(t-\ell-2n) \rangle}_{\delta[m-\ell-2n]} \\ &= \sum_m c_0[m] h[m-2n]. \end{aligned}$$

The previous expression indicates that $\{c_1[n]\}$ results from convolving $\{c_0[m]\}$ with a time-reversed version of $h[m]$ then downsampling by factor two.

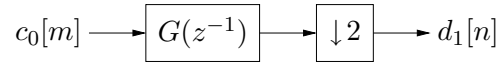
$$c_0[m] \longrightarrow \boxed{H(z^{-1})} \longrightarrow \boxed{\downarrow 2} \longrightarrow c_1[n]$$

Using the fact that $\{\psi_{1,n}(t) : n \in \mathbb{Z}\}$ is an orthonormal basis for W_1 , in conjunction with

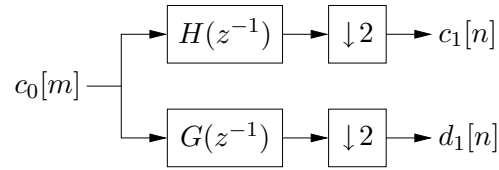
the wavelet scaling equation,

$$\begin{aligned}
d_1[n] &= \langle x_0(t), \psi_{1,n}(t) \rangle \\
&= \left\langle \sum_m c_0[m] \phi_{0,m}(t), \psi_{1,n}(t) \right\rangle \\
&= \sum_m c_0[m] \langle \phi_{0,m}(t), \psi_{1,n}(t) \rangle \\
&= \sum_m c_0[m] \left\langle \phi(t-m), \sum_\ell g[\ell] \phi(t-\ell-2n) \right\rangle \\
&= \sum_m c_0[m] \sum_\ell g[\ell] \underbrace{\langle \phi(t-m), \phi(t-\ell-2n) \rangle}_{\delta[m-\ell-2n]} \\
&= \sum_m c_0[m] g[m-2n].
\end{aligned}$$

The previous expression indicates that $\{d_1[n]\}$ results from convolving $\{c_0[m]\}$ with a time-reversed version of $g[m]$ then downsampling by factor two.



Putting these two operations together, we arrive at what looks like the analysis portion of an FIR filterbank:



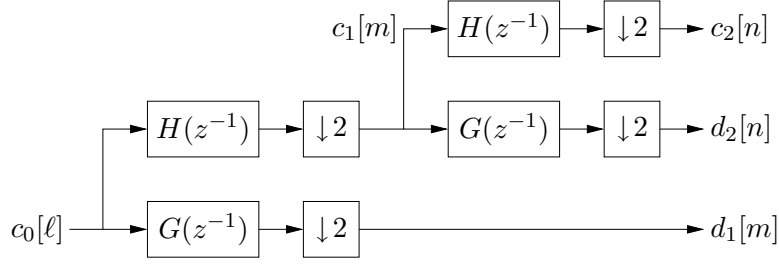
We can repeat this process at the next higher level of coarseness. Since $V_1 = W_2 \oplus V_2$, there exist coefficients $\{c_2[n]\}$ and $\{d_2[n]\}$ such that

$$\begin{aligned}
x_1(t) &= \sum_n c_1[n] \phi_{1,n}(t) \\
&= \sum_n d_2[n] \psi_{2,n}(t) + \sum_n c_2[n] \phi_{2,n}(t)
\end{aligned}$$

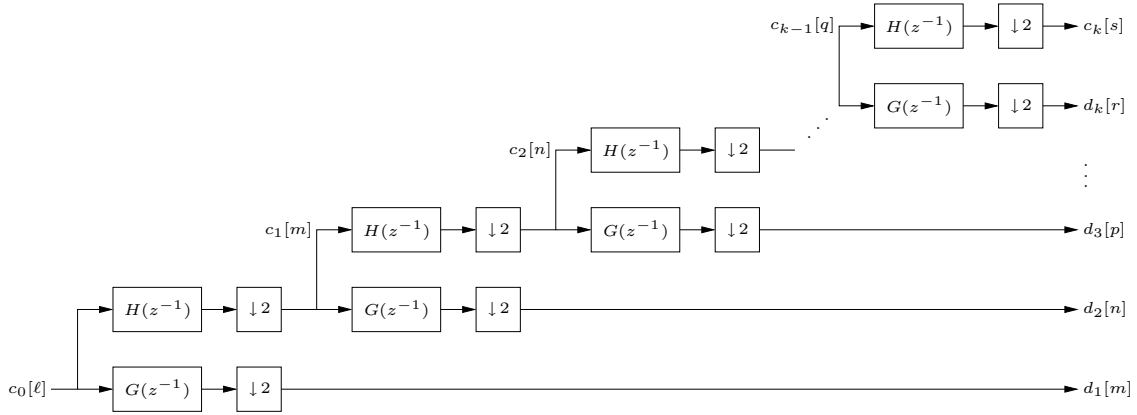
Using the same steps as before, we find that

$$\begin{aligned}
c_2[n] &= \sum_m c_1[m] h[m-2n] \\
d_2[n] &= \sum_m c_1[m] g[m-2n]
\end{aligned}$$

which gives a cascaded analysis filterbank:



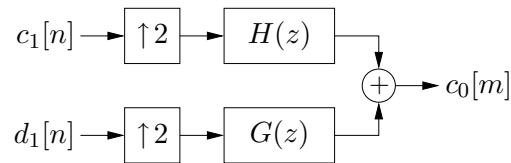
If we use $V_0 = W_1 \oplus W_2 \oplus W_3 \oplus \dots \oplus W_k \oplus V_k$ to repeat this process up to the k^{th} level, we get the iterated analysis filterbank below.



As we might expect, signal reconstruction can be accomplished using cascaded two-channel synthesis filterbanks. Using the same assumptions as before, we have:

$$\begin{aligned}
 c_0[m] &= \langle x_0(t), \phi_{0,m}(t) \rangle \\
 &= \left\langle \sum_n c_1[n] \phi_{1,n}(t) + \sum_n d_1[n] \psi_{1,n}(t), \phi_{0,m}(t) \right\rangle \\
 &= \sum_n c_1[n] \underbrace{\langle \phi_{1,n}(t), \phi_{0,m}(t) \rangle}_{h[m-2n]} + \sum_n d_1[n] \underbrace{\langle \psi_{1,n}(t), \phi_{0,m}(t) \rangle}_{g[m-2n]} \\
 &= \sum_n c_1[n] h[m-2n] + \sum_n d_1[n] g[m-2n]
 \end{aligned}$$

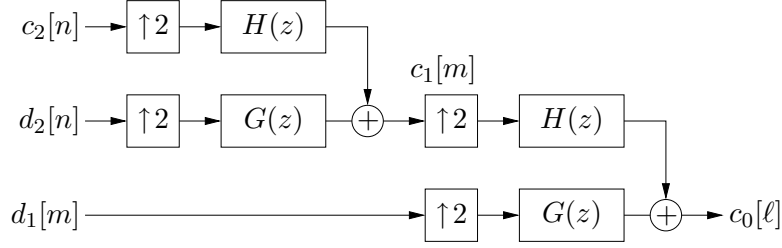
which can be implemented using the block diagram below.



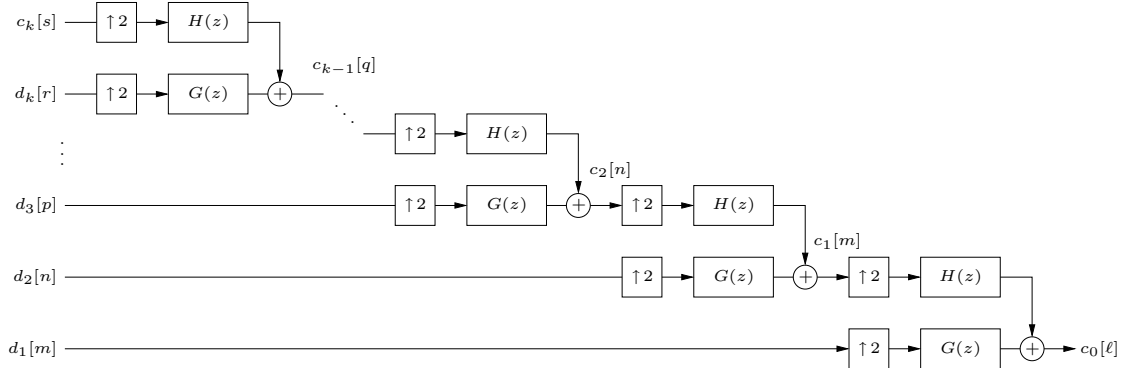
The same procedure can be used to derive

$$c_1[m] = \sum_n c_2[n] h[m-2n] + \sum_n d_2[n] g[m-2n],$$

from which we get the diagram:



To reconstruct from the k^{th} level, we can use the iterated synthesis filterbank:



The table below makes a direct comparison between wavelets and the two-channel orthogonal PR-FIR filterbanks.

	Discrete Wavelet Transform	2-Channel Orthogonal PR-FIR Filterbank
Analysis LPF:	$H(z^{-1})$	$H_0(z)$
Power Symmetry:	$H(z)H(z^{-1}) + H(-z)H(-z^{-1}) = 2$	$H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1}) = 1$
Analysis HPF:	$G(z^{-1})$	$H_1(z)$
Spectral Reverse:	$G(z) = \pm z^{-P}H(-z^{-1})$, P odd	$H_1(z) = \pm z^{-(N-1)}H_0(-z^{-1})$, N even
Synthesis LPF:	$H(z)$	$G_0(z) = 2z^{-(N-1)}H_0(z^{-1})$
Synthesis HPF:	$G(z)$	$G_1(z) = 2z^{-(N-1)}H_1(z^{-1})$

From the table above, we see that the discrete wavelet transform that we have been developing is identical to two-channel orthogonal PR-FIR filterbanks in all but a couple details.

1. Orthogonal PR-FIR filterbanks employ synthesis filters with twice the gain of the analysis filters, whereas in the DWT the gains are equal.
2. Orthogonal PR-FIR filterbanks employ causal filters of length N , whereas the DWT filters are not constrained to be causal.

For convenience, however, the wavelet filters $H(z)$ and $G(z)$ are usually chosen to be causal. For both to have even impulse response length N , we require that $P = N - 1$.

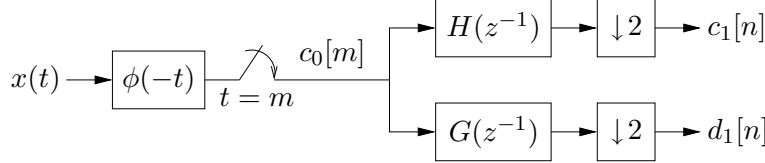
- **Initialization of the Wavelet Transform:** The filterbanks developed in the last section start with the signal representation $\{c_0[n] : n \in \mathbb{Z}\}$ and break the representation down into wavelet coefficients and scaling coefficients at lower resolutions (i.e., higher levels k). Of course, we could easily start at a non-zero level K and break the representation down from

there; $K = 0$ just leads to simple notation. The question remains, though: how do we get the initial coefficients $\{c_0[n]\}$?

From their definition, we see that the scaling coefficients can be written using a convolution:

$$c_0[n] = \langle \phi(t - n), x(t) \rangle = \int_{-\infty}^{\infty} \phi(t - n)x(t)dt = \phi(-t) * x(t)|_{t=n},$$

which suggests that the proper initialization of the wavelet transform is accomplished by passing the continuous-time input $x(t)$ through an analog filter with impulse response $\phi(-t)$ and sampling its output at integer times.

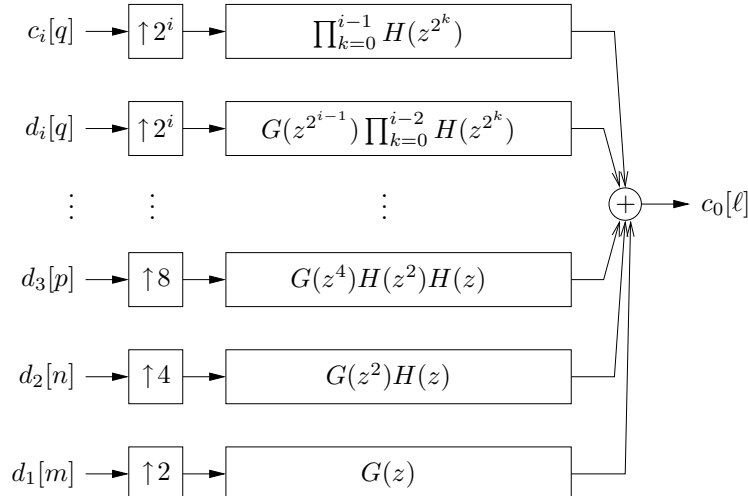


Practically speaking, however, it is very difficult to build an analog filter with impulse response $\phi(-t)$ for typical choices of scaling function.

The most often-used approximation is to set $c_0[n] = x[n]$. This initialization would be exact if $x(t)$ was bandlimited to $\frac{1}{2}$ Hz and CTFT $\Phi(\Omega) = 1$ for $\Omega \in (-\pi, \pi]$ rad/s, but it is not exact in general. (A similar property could be stated if we started the DWT at a non-zero level K .) However, if we regard the wavelet transform primarily as a *discrete-time multi-resolution analysis/synthesis tool*, then the initialization (or continuous-to-discrete conversion) stage is not so important.

- Regularity Conditions, Compact Support, and Daubechies' Wavelets: Up until now we have been somewhat vague about the requirements for the conditions $V_{-\infty} = \mathcal{L}_2$ and $\langle \phi(t - n), \phi(t) \rangle = \delta[n]$. As we shall soon see, the conditions on $\{h[n]\}$ and $\{g[n]\}$ developed thus far are not enough to yield $\phi(t)$ satisfying these conditions! In this section, therefore, we will take a closer look at what happens when $k \rightarrow -\infty$ and arrive at a popular family of filter coefficients $h[n]$ and $g[n]$ which give continuous $\phi(t)$ and $\psi(t)$ that satisfy our desired wavelet system properties. These are known as “Daubechies’ Wavelets.”

Recall the iterated synthesis filterbank. Applying the Noble identities, we can move the up-samplers before the filters, as illustrated below.



The properties of the i -stage cascaded lowpass filter

$$H^{(i)}(z) = \prod_{k=0}^{i-1} H(z^{2^k}), \quad i \geq 1$$

in the limit $i \rightarrow \infty$ give an important characterization of the wavelet system. But how do we know that $\lim_{i \rightarrow \infty} H^{(i)}(e^{j\omega})$ converges to a response in \mathcal{L}_2 ? In fact, there are some rather strict conditions on $H(e^{j\omega})$ that must be satisfied for this convergence to occur. Without such convergence, we might have a finite-stage perfect reconstruction filterbank, but we will *not* have a countable wavelet basis for \mathcal{L}_2 . Below we present some “regularity conditions” on $H(e^{j\omega})$ that ensure convergence of the iterated synthesis lowpass filter. (Note that convergence of the lowpass filter implies convergence of all other filters in the bank.)

Let us denote the impulse response of $H^{(i)}(z)$ by $h^{(i)}[n]$. Writing

$$H^{(i)}(z) = H(z^{2^{i-1}})H^{(i-1)}(z)$$

in the time domain, we have

$$h^{(i)}[n] = \sum_k h[k] h^{(i-1)}[n - 2^{i-1}k].$$

Now define the function

$$\phi^{(i)}(t) = 2^{i/2} \sum_n h^{(i)}[n] p_{[\frac{n}{2^i}, \frac{n+1}{2^i})}(t),$$

where $p_{[a,b)}(t)$ denotes the height-1 pulse with support on the interval $[a, b)$:

$$p_{[a,b)}(t) = \begin{cases} 1 & t \in [a, b) \\ 0 & t \notin [a, b) \end{cases}$$

The definition of $\phi^{(i)}(t)$ implies

$$\begin{aligned} h^{(i)}[n] &= 2^{-i/2} \phi^{(i)}(t), & t \in [\frac{n}{2^i}, \frac{n+1}{2^i}) \\ h^{(i-1)}[n - 2^{i-1}k] &= 2^{-(i-1)/2} \phi^{(i-1)}(2t - k), & t \in [\frac{n}{2^i}, \frac{n+1}{2^i}) \end{aligned}$$

and plugging the two previous expressions into the equation for $h^{(i)}[n]$ yields

$$\phi^{(i)}(t) = \sqrt{2} \sum_k h[k] \phi^{(i-1)}(2t - k).$$

Thus, if $\phi^{(i)}(t)$ converges pointwise to a continuous function, then the limiting function must satisfy the scaling equation, so that $\lim_{i \rightarrow \infty} \phi^{(i)}(t) = \phi(t)$. Daubechies [2] showed that, for pointwise convergence of $\phi^{(i)}(t)$ to a continuous function in \mathcal{L}_2 , it is sufficient that $H(e^{j\omega})$ can be factored as

$$H(e^{j\omega}) = \sqrt{2} \left(\frac{1 + e^{j\omega}}{2} \right)^P R(e^{j\omega}), \quad P \geq 1,$$

for $R(e^{j\omega})$ such that

$$\sup_{\omega} |R(e^{j\omega})| < 2^{P-1}.$$

Here P denotes the number of zeros that $H(e^{j\omega})$ has at $\omega = \pi$. Such conditions are called “regularity” conditions because they ensure the regularity, or “smoothness,” of $\phi(t)$. In fact, if we make the previous condition stronger:

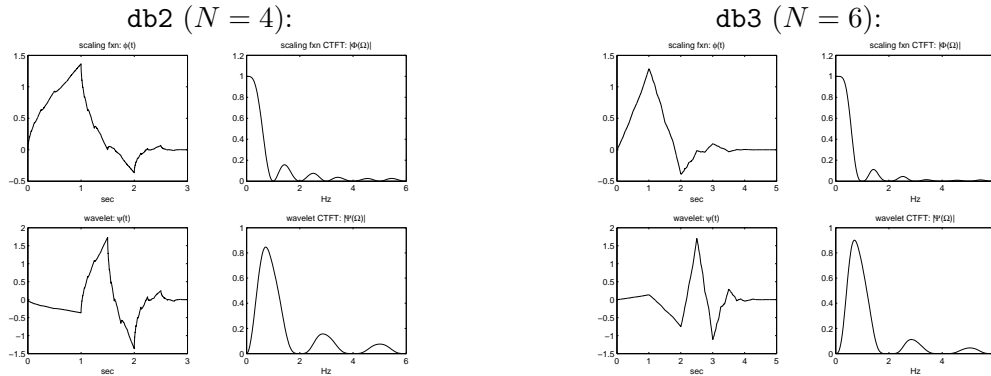
$$\sup_{\omega} |R(e^{j\omega})| < 2^{P-1-\ell}, \quad \ell \geq 1,$$

then $\lim_{i \rightarrow \infty} \phi^{(i)}(t) = \phi(t)$ for $\phi(t)$ that is ℓ -times continuously differentiable.

There is an interesting and important by-product of the preceding analysis. If $h[n]$ is a causal length- N filter, it can be shown that $h^{(i)}[n]$ is causal with length $N^{(i)} = (2^i - 1)(N - 1) + 1$. By construction, then, $\phi^{(i)}(t)$ will be zero outside the interval $[0, \frac{(2^i - 1)(N - 1) + 1}{2^i}]$. Assuming that the regularity conditions are satisfied so that $\lim_{i \rightarrow \infty} \phi^{(i)}(t) = \phi(t)$, it follows that $\phi(t)$ must be zero outside the interval $[0, N - 1]$. In this case we say that $\phi(t)$ has “compact support.” Finally, the wavelet scaling equation implies that, when $\phi(t)$ is compactly supported on $[0, N - 1]$ and $g[n]$ is length N , $\psi(t)$ will also be compactly supported on the interval $[0, N - 1]$.

Daubechies constructed a family of $H(z)$ with impulse response lengths $N = 4, 6, 8, 10, \dots$ which satisfy the regularity conditions. Moreover, her filters have the maximum possible number of zeros at $\omega = \pi$, and thus are maximally regular (i.e., they yield the smoothest possible $\phi(t)$ for a given support interval). It turns out that these filters are the “maximally flat” filters derived by Herrmann [3] long before filterbanks and wavelets were in vogue. Below we show $\phi(t)$, $\Phi(\Omega)$, $\psi(t)$, and $\Psi(\Omega)$ for various members of the Daubechies’ wavelet system.

See [1] for a more complete discussion of these matters.



- Computation of $\phi(t)$ and $\psi(t)$ —The Cascade Algorithm: Given coefficients $\{h[n]\}$ that satisfy the regularity conditions, we can iteratively calculate samples of $\phi(t)$ on a fine grid of points $\{t\}$ using the “cascade algorithm.” Once we have obtained $\phi(t)$, the wavelet scaling equation can be used to construct $\psi(t)$.

In this module we assume that $H(z)$ is causal with impulse response length N . Recall, from our discussion of the regularity conditions, that this implies $\phi(t)$ will have compact support on the interval $[0, N - 1]$. The cascade algorithm is described below.

0. Consider the scaling function at integer times $t = m \in \{0 \dots N - 1\}$:

$$\phi(m) = \sqrt{2} \sum_{n=0}^{N-1} h[n] \phi(2m - n).$$

Knowing that $\phi(t) = 0$ for $t \notin [0, N - 1]$, the previous equation can be written using an $N \times N$ matrix. In the case $N = 4$, we have

$$\begin{bmatrix} \phi(0) \\ \phi(1) \\ \phi(2) \\ \phi(3) \end{bmatrix} = \sqrt{2} \underbrace{\begin{bmatrix} h[0] & 0 & 0 & 0 \\ h[2] & h[1] & h[0] & 0 \\ 0 & h[3] & h[2] & h[1] \\ 0 & 0 & 0 & h[3] \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} \phi(0) \\ \phi(1) \\ \phi(2) \\ \phi(3) \end{bmatrix}.$$

The matrix \mathbf{H} is structured as a “row-decimated convolution matrix.” From the matrix equation above, we see that $[\phi(0), \phi(1), \phi(2), \phi(3)]^t$ must be (some scaled version of) the eigenvector of \mathbf{H} corresponding to eigenvalue $(\sqrt{2})^{-1}$.

In general, the nonzero values of $\{\phi(n) : n \in \mathbb{Z}\}$, i.e., $[\phi(0), \phi(1), \dots, \phi(N - 1)]$, can be calculated by appropriately scaling the eigenvector of the $N \times N$ row-decimated convolution matrix \mathbf{H} corresponding to the eigenvalue $(\sqrt{2})^{-1}$. It can be shown that this eigenvector must be scaled so that $\sum_{n=0}^{N-1} \phi(n) = 1$.

1. Given $\{\phi(n) : n \in \mathbb{Z}\}$, we can use the scaling equation to determine $\{\phi(\frac{n}{2}) : n \in \mathbb{Z}\}$:

$$\phi(\frac{m}{2}) = \sqrt{2} \sum_{n=0}^{N-1} h[n] \phi(m - n).$$

This produces the $2N - 1$ non-zero samples $\{\phi(0), \phi(\frac{1}{2}), \phi(1), \phi(\frac{3}{2}), \dots, \phi(N - 1)\}$.

2. Given $\{\phi(\frac{n}{2}) : n \in \mathbb{Z}\}$, the scaling equation can be used to find $\{\phi(\frac{n}{4}) : n \in \mathbb{Z}\}$:

$$\begin{aligned} \phi(\frac{m}{4}) &= \sqrt{2} \sum_{n=0}^{N-1} h[n] \phi(\frac{m}{2} - n) \\ &= \sqrt{2} \sum_{p \text{ even}} h[\frac{p}{2}] \phi(\frac{m-p}{2}) \\ &= \sqrt{2} \sum_p h_{\uparrow 2}[p] \phi_{\frac{1}{2}}[m - p] \end{aligned}$$

where $h_{\uparrow 2}[n]$ denotes the impulse response of $H(z^2)$, i.e., a 2-upsampled version of $h[n]$, and where $\phi_{\frac{1}{2}}[m] = \phi(\frac{m}{2})$. Note that $\{\phi(\frac{n}{4}) : n \in \mathbb{Z}\}$ is the result of convolving $h_{\uparrow 2}[n]$ with $\{\phi_{\frac{1}{2}}[n]\}$.

3. Given $\{\phi(\frac{n}{4}) : n \in \mathbb{Z}\}$, another convolution yields $\{\phi(\frac{n}{8}) : n \in \mathbb{Z}\}$:

$$\begin{aligned}\phi(\frac{m}{8}) &= \sqrt{2} \sum_{n=0}^{N-1} h[n] \phi(\frac{m}{4} - n) \\ &= \sqrt{2} \sum_p h_{\uparrow 4}[p] \phi_{\frac{1}{4}}[m-p]\end{aligned}$$

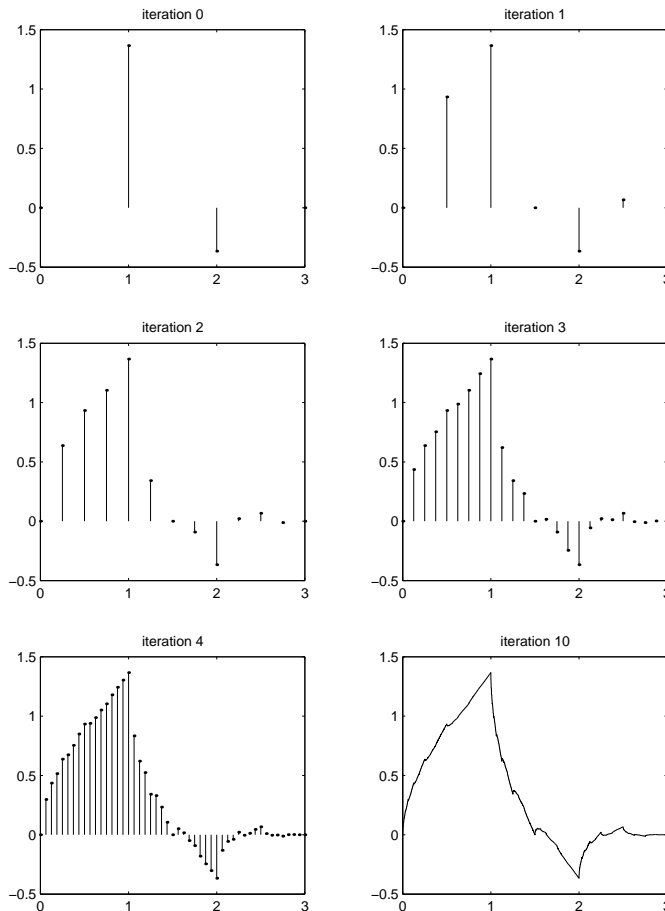
where $h_{\uparrow 4}[n]$ is a 4-upsampled version of $h[n]$ and where $\phi_{\frac{1}{4}}[m] = \phi(\frac{m}{4})$.

ℓ . At the ℓ^{th} stage, $\{\phi(\frac{n}{2^\ell})\}$ is calculated by convolving the result of the $(\ell-1)^{th}$ stage with a $2^{\ell-1}$ -upsampled version of $h[n]$:

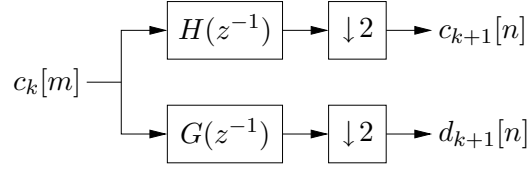
$$\phi_{\frac{1}{2^\ell}}[m] = \sqrt{2} \sum_p h_{\uparrow 2^{\ell-1}}[p] \phi_{\frac{1}{2^{\ell-1}}}[m-p]$$

For $\ell \approx 10$, this gives a very good approximation of $\phi(t)$. At this point, you could verify the key properties of $\phi(t)$, such as orthonormality and the satisfaction of the scaling equation.

Below we show steps 0 through 4 of the cascade algorithm, as well as step 10, using Daubechies' db2 coefficients (for which $N = 4$).



- Finite-length Sequences and the DWT Matrix: The wavelet transform, viewed from a filterbank perspective, consists of iterated 2-channel analysis stages like the one below.



First consider a very long (i.e., practically infinite-length) sequence $\{c_k[m] : m \in \mathbb{Z}\}$. For every pair of input samples $\{c_k[2n], c_k[2n - 1]\}$ that enter the k^{th} filterbank stage, exactly one pair of output samples $\{c_{k+1}[n], d_{k+1}[n]\}$ are generated. In other words, the number of outputs equal the number of inputs during a fixed time interval. This property is convenient from a real-time processing perspective.

For a short sequence $\{c_k[m], m = 0 \dots M - 1\}$, however, linear convolution requires that we make an assumption about the values of $c_k[m]$ for $m \notin \{0 \dots M - 1\}$. One assumption could be

$$c_k[m] = 0 \text{ for } m \notin \{0 \dots M - 1\}.$$

If we assume that both $H(z^{-1})$ and $G(z^{-1})$ have impulse response lengths of N , and that both M and N are even, then linear convolution implies that M nonzero inputs yield $\frac{M+N}{2}$ outputs from each branch, for a total of $M+N > M$ outputs. The fact that each filterbank stage produces more outputs than inputs could be inconvenient for many applications.

A more convenient assumption regarding the tails of $\{c_k[m], m = 0 \dots M - 1\}$ is that the data outside of the time window $\{0 \dots M - 1\}$ is a cyclic extension of data inside the time window:

$$c_k[m] = c_k[\langle m \rangle_M] \text{ for } m \notin \{0 \dots M - 1\}.$$

Recall that a linear convolution with an M -cyclic signal is equivalent to a circular convolution with one M -sample segment of the signal. Furthermore, the output of this circular convolution is itself M -cyclic, implying our 2-downsampled branch outputs will be cyclic with period $M/2$. Thus, given an M -length input sequence, the filterbank output consists of exactly M unique values.

It is instructive to write the circular-convolution analysis filterbank operation in matrix form. Below we give an example for filter length $N = 4$, sequence length $M = 8$, and causal synthesis filters $H(z)$ and $G(z)$.

$$\underbrace{\begin{bmatrix} c_{k+1}[0] \\ c_{k+1}[1] \\ c_{k+1}[2] \\ c_{k+1}[3] \\ d_{k+1}[0] \\ d_{k+1}[1] \\ d_{k+1}[2] \\ d_{k+1}[3] \end{bmatrix}}_{\begin{bmatrix} \mathbf{c}_{k+1} \\ \mathbf{d}_{k+1} \end{bmatrix}} = \underbrace{\begin{bmatrix} h[0] & h[1] & h[2] & h[3] & 0 & 0 & 0 & 0 \\ 0 & 0 & h[0] & h[1] & h[2] & h[3] & 0 & 0 \\ 0 & 0 & 0 & 0 & h[0] & h[1] & h[2] & h[3] \\ h[2] & h[3] & 0 & 0 & 0 & 0 & h[0] & h[1] \\ g[0] & g[1] & g[2] & g[3] & 0 & 0 & 0 & 0 \\ 0 & 0 & g[0] & g[1] & g[2] & g[3] & 0 & 0 \\ 0 & 0 & 0 & 0 & g[0] & g[1] & g[2] & g[3] \\ g[2] & g[3] & 0 & 0 & 0 & 0 & g[0] & g[1] \end{bmatrix}}_{\begin{bmatrix} \mathbf{H}_M \\ \mathbf{G}_M \end{bmatrix}} \underbrace{\begin{bmatrix} c_k[0] \\ c_k[1] \\ c_k[2] \\ c_k[3] \\ c_k[4] \\ c_k[5] \\ c_k[6] \\ c_k[7] \end{bmatrix}}_{\mathbf{c}_k}$$

The matrices \mathbf{H}_M and \mathbf{G}_M have interesting properties. For example, the conditions

$$\begin{aligned}\delta[m] &= \sum_n h[n]h[n-2m] \\ g[n] &= (-1)^n h[N-1-n]\end{aligned}$$

imply that

$$\begin{bmatrix} \mathbf{H}_M \\ \mathbf{G}_M \end{bmatrix}^t \begin{bmatrix} \mathbf{H}_M \\ \mathbf{G}_M \end{bmatrix} = \begin{bmatrix} \mathbf{H}_M \\ \mathbf{G}_M \end{bmatrix} \begin{bmatrix} \mathbf{H}_M \\ \mathbf{G}_M \end{bmatrix}^t = \mathbf{I}_M,$$

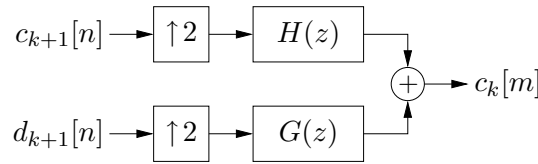
where \mathbf{I}_M denotes the $M \times M$ identity matrix. Thus, it makes sense to define the $M \times M$ ‘‘DWT matrix’’ as

$$\mathbf{T}_M = \begin{bmatrix} \mathbf{H}_M \\ \mathbf{G}_M \end{bmatrix}$$

whose transpose constitutes the $M \times M$ ‘‘inverse DWT matrix’’:

$$\mathbf{T}_M^{-1} = \mathbf{T}_M^t.$$

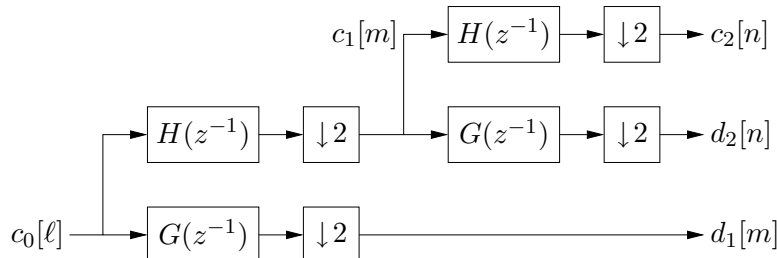
Since the synthesis filterbank



gives perfect reconstruction, and since the cascade of matrix operations $\mathbf{T}_M^t \mathbf{T}_M$ also corresponds to perfect reconstruction, we expect that the matrix operation \mathbf{T}_M^t describes the action of the synthesis filterbank. This is readily confirmed by writing the upsampled circular convolutions in matrix form:

$$\begin{bmatrix} c_k[0] \\ c_k[1] \\ c_k[2] \\ c_k[3] \\ c_k[4] \\ c_k[5] \\ c_k[6] \\ c_k[7] \end{bmatrix} = \underbrace{\begin{bmatrix} h[0] & 0 & 0 & h[2] & g[0] & 0 & 0 & g[2] \\ h[1] & 0 & 0 & h[3] & g[1] & 0 & 0 & g[3] \\ h[2] & h[0] & 0 & 0 & g[2] & g[0] & 0 & 0 \\ h[3] & h[1] & 0 & 0 & g[3] & g[1] & 0 & 0 \\ 0 & h[2] & h[0] & 0 & 0 & g[2] & g[0] & 0 \\ 0 & h[3] & h[1] & 0 & 0 & g[3] & g[1] & 0 \\ 0 & 0 & h[2] & h[0] & 0 & 0 & g[2] & g[0] \\ 0 & 0 & h[3] & h[1] & 0 & 0 & g[3] & g[1] \end{bmatrix}}_{\begin{bmatrix} \mathbf{H}_M^t & \mathbf{G}_M^t \end{bmatrix} = \mathbf{T}_M^t} \begin{bmatrix} c_{k+1}[0] \\ c_{k+1}[1] \\ c_{k+1}[2] \\ c_{k+1}[3] \\ d_{k+1}[0] \\ d_{k+1}[1] \\ d_{k+1}[2] \\ d_{k+1}[3] \end{bmatrix}.$$

So far we have concentrated on one stage in the wavelet decomposition; a two-stage decomposition is illustrated below.



The two-stage analysis operation (assuming circular convolution) can be expressed in matrix form as

$$\begin{bmatrix} \mathbf{c}_{k+2} \\ \mathbf{d}_{k+2} \\ \mathbf{d}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{\frac{M}{2}} & \\ & \\ & \mathbf{I}_{\frac{M}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{k+1} \\ \mathbf{d}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{\frac{M}{2}} & \\ & \mathbf{I}_{\frac{M}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{T}_M \\ \end{bmatrix} \begin{bmatrix} \mathbf{c}_k \end{bmatrix}.$$

Similarly, a three-stage analysis could be implemented via

$$\begin{bmatrix} \mathbf{c}_{k+3} \\ \mathbf{d}_{k+3} \\ \mathbf{d}_{k+2} \\ \mathbf{d}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{\frac{M}{4}} & & \\ & \mathbf{I}_{\frac{M}{4}} & \\ & & \mathbf{I}_{\frac{M}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{T}_{\frac{M}{2}} & \\ & \mathbf{I}_{\frac{M}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{T}_M \\ \end{bmatrix} \begin{bmatrix} \mathbf{c}_k \end{bmatrix}.$$

It should now be evident how to extend this procedure to > 3 stages. As noted earlier, the corresponding synthesis operations are accomplished by transposing the matrix products used in the analysis.

- DWT Implementation using FFTs: Finally, we say a few words about DWT implementation. Here we focus on a single DWT stage and assume circular convolution, yielding an $M \times M$ DWT matrix \mathbf{T}_M . In the general case, $M \times M$ matrix multiplication requires M^2 multiplications. The DWT matrices, however, have a circular-convolution structure which allows us to implement them using significantly less multiplies. Below we present some simple and reasonably efficient approaches for the implementation of \mathbf{T}_M and \mathbf{T}_M^t . We treat the inverse DWT first. Recall that in the lowpass synthesis branch, we upsample the input before circularly convolving with $H(z)$. Denoting the upsampled coefficient sequence by $a[n]$, fast circular convolution $a[n] \otimes h[n]$ can be described as follows (using Matlab notation)

$$\text{ifft}(\text{fft}(\mathbf{a}) .* \text{fft}(\mathbf{h}, \text{length}(\mathbf{a})))$$

where we have assumed¹¹ that $\text{length}(\mathbf{a}) \geq \text{length}(\mathbf{h})$. The highpass branch is handled similarly using $G(z)$, after which the two branch outputs are summed.

Next we treat the forward DWT. Recall that in the lowpass analysis branch, we circularly convolve the input with $H(z^{-1})$ and then downsample the result. The fast circular convolution $a[n] \otimes h[-n]$ can be implemented using

$$\text{wshift}('1', \text{ifft}(\text{fft}(\mathbf{a}) .* \text{fft}(\text{flipud}(\mathbf{h}), \text{length}(\mathbf{a}))), \text{length}(\mathbf{h}) - 1)$$

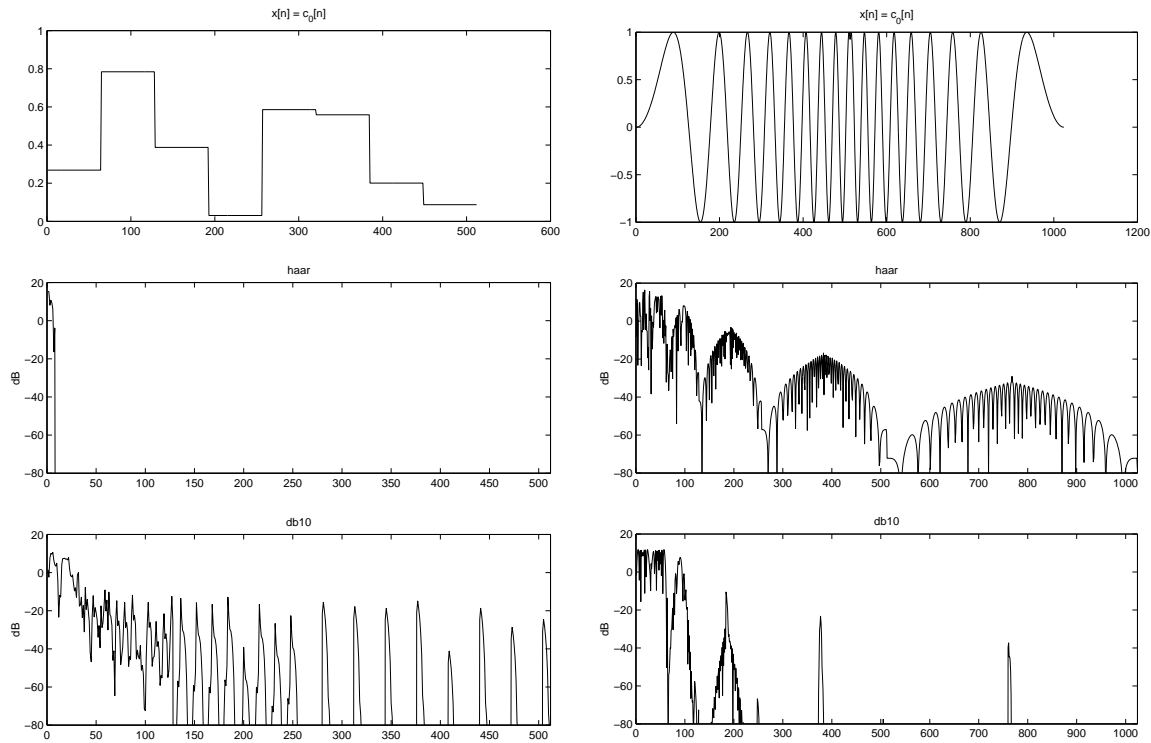
where `wshift` accomplishes a circular shift of the `ifft` output that makes up for the unwanted delay of $\text{length}(\mathbf{h}) - 1$ samples imposed by the `flipud` operation. The highpass branch is handled similarly but with filter $G(z^{-1})$. Finally, each branch is downsampled by factor two.

¹¹When implementing the multi-level wavelet transform, you must ensure that the data length does not become shorter than the filter length!

We note that the proposed approach is not totally efficient because downsampling is performed after circular convolution (and upsampling before circular convolution). Still, we have outlined this approach because it is easy to understand and still results in major savings when M is large: it converts the $\mathcal{O}(M^2)$ matrix multiply into an $\mathcal{O}(M \log_2 M)$ operation.

- DWT Applications—Choice of $\phi(t)$: Transforms are signal processing tools that are used to give a clear view of essential signal characteristics. Fourier transforms are ideal for infinite-duration signals that contain a relatively small number of sinusoids: one can completely describe the signal using only a few coefficients. Fourier transforms, however, are not well-suited to signals of a non-sinusoidal nature (as discussed earlier in the context of time-frequency analysis). The multi-resolution DWT is a more general transform that is well-suited to a larger class of signals. For the DWT to give an efficient description of the signal, however, we must choose a wavelet $\psi(t)$ from which the signal can be constructed (to a good approximation) using only a few stretched and shifted copies.

We illustrate this concept below using two examples. On the left, we analyze a step-like waveform, while on the right we analyze a chirp-like waveform. In both cases, we try DWTs based on the Haar and Daubechies db10 wavelets and plot the log magnitudes of the transform coefficients $[\mathbf{c}_k^t, \mathbf{d}_k^t, \mathbf{d}_{k-1}^t, \mathbf{d}_{k-2}^t, \dots, \mathbf{d}_1^t]$.



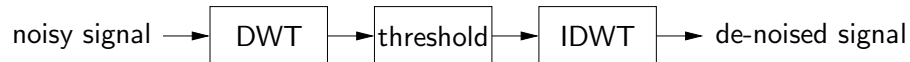
Observe that the Haar DWT yields an extremely efficient representation of the step-waveform: only a few of the transform coefficients are nonzero. The db10 DWT does not give an efficient representation: many coefficients are sizable. This makes sense because the Haar scaling function is well matched to the step-like nature of the time-domain signal. In contrast, the Haar DWT does not give an efficient representation of the chirp-like waveform, while the db10 DWT does better. This makes sense because the sharp edges of the Haar scaling function do not match the smooth chirp signal, while the smoothness of

the `db10` wavelet yields a better match.

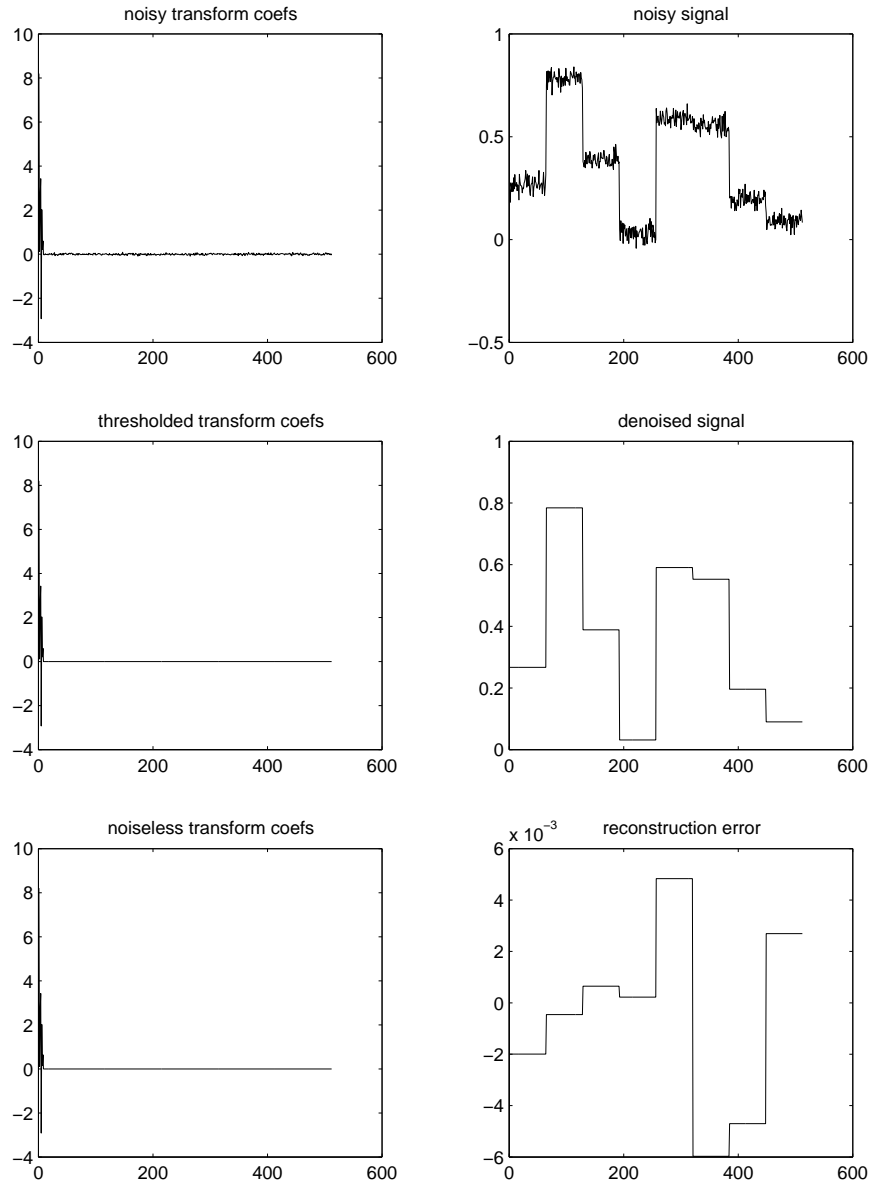
- *DWT Application—De-noising*: Say that the DWT for a particular choice of wavelet yields an efficient representation of a particular signal class. In other words, signals in the class are well-described using a few large transform coefficients.

Now consider unstructured “noise”, which cannot be efficiently represented by any transform, including the DWT. Due to the orthonormality of the DWT basis, such noise sequences make, on average, equal contributions to all transform coefficients. Any given noise sequence is expected to yield many small-valued transform coefficients.

Together, these two ideas suggest a means of “de-noising” a signal. Say that we perform a DWT on a signal from our “well-matched” signal class that has been corrupted by additive noise. We expect that large transform coefficients are composed mostly of signal content, while small transform coefficients should be composed mostly of noise content. Hence, throwing away the transform coefficients whose magnitude is less than some small threshold should improve the signal-to-noise ratio. The de-noising procedure is illustrated below.



Now we give an example of denoising a step-like waveform using the Haar DWT. In the figure below, the top right subplot shows the noisy signal and the top left shows its DWT coefficients. Note the presence of a few large DWT coefficients, expected to contain mostly signal components, as well as the presence of many small-valued coefficients, expected to contain noise. (The bottom left subplot shows the DWT for the original signal before any noise was added, which confirms that all signal energy is contained within a few large coefficients.) If we throw away all DWT coefficients whose magnitude is less than 0.1, we are left with only the large coefficients (shown in the middle left plot) which correspond to the de-noised time-domain signal shown in the middle right plot. The difference between the de-noised signal and the original noiseless signal is shown in the bottom right. Non-zero error results from noise contributions to the large wavelet coefficients; there is no way of distinguishing these noise components from signal components.



References

- [1] M. Vetterli and J. Kovacević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice Hall, 1995.
- [2] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” *Commun. on Pure and Applied Math*, vol. 41, pp. 909–996, Nov. 1988.
- [3] O. Herrmann, “On the approximation problem in nonrecursive digital filter design,” *IEEE Trans. on Circuit Theory*, vol. 18, pp. 411–413, 1971.