Chapter 1

Background

This thesis combines work in several fields. We provide a background for the most important and relevant fields in this chapter. We first introduce the definition and properties of the Wavelet Transforms, focussing on the DT$\mathbb{C}$WT, then we describe the fundamentals of Convolutional Neural Networks and how they learn, and finally we introduce the Scattering Transform, the original inspiration for this thesis.

1.1 The Fourier and Wavelet Transforms

Computer vision is an extremely difficult task. Pixel intensities in an image are typically not very informative in understanding what is in that image. Indeed, these values are sensitive to lighting conditions and camera configurations. It would be easy to take two photos of the same scene and get two vectors x_1 and x_2 that have a very large Euclidean distance, but to a human, would represent the same objects. What is most important in defining an image is difficult to define, however some things are notably more important than others. In particular, the location or phase of the waves that make up an image is much more important than the magnitude of these waves, something that is not necessarily true for audio processing. A simple experiment to demonstrate this is shown in Figure 1.1.

1.1.1 The Fourier Transform

For a signal $f(t) \in L_2(\mathbb{R})$ (square summable signals), the *Fourier transform* is defined as:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \tag{1.1.1}$$

This can be extended to two dimensions for signals $f(\mathbf{u}) \in L_2(\mathbb{R}^2)$:

$$F(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\mathbf{u}) e^{-j\boldsymbol{\omega}^t \mathbf{u}} d\mathbf{u} = \langle f(\mathbf{u}),\ e^{j\boldsymbol{\omega}^t \mathbf{u}} \rangle \tag{1.1.2}$$

Figure 1.1: **Importance of phase over magnitude for images.** The phase of the Fourier transform of the first image is combined with the magnitude of the Fourier transform of the second image and reconstructed. Note that the first image has entirely won out and nothing is left visible of the cameraman.

The Fourier transform is an invaluable signal expansion, as viewing a signal in the frequency space offers many insights, as well as affording many very useful properties (most notably the efficiency of convolution as a product of Fourier transforms). While it is a mainstay in signal processing, it can be a poor feature descriptor due to the infinite support of its basis functions - the complex sinusoids $e^{j\boldsymbol{\omega}^t u}$. If a single pixel changes in the input it can change all of the Fourier coefficients. As natural images are generally non-stationary, we need to be able to isolate frequency components in local regions of an image, and not have this property of global dependence.

1.1.2 The Continuous Wavelet Transform

The *Continuous Wavelet Transform* (CWT), like the Fourier Transform, can be used to decompose a signal into its frequency components. Unlike the Fourier transform, these frequency components can be localized in space. To achieve this, we need a bandpass filter, or *mother wavelet* ψ such that:

$$\int_{-\infty}^{\infty} \psi(\mathbf{u}) d\mathbf{u} = \Psi(0) = 0 \tag{1.1.3}$$

Any function that sufficient decay in frequency and satisfies (1.1.3) satisfies the *admissibility condition*.

As we are working in 2-D for image processing, consider dilations, shifts and rotations of this function by $a, \mathbf{b}, \theta$ where

$$\text{Translation: } T_{\mathbf{b}}x(\mathbf{u}) = x(\mathbf{u}-\mathbf{b}) \tag{1.1.4}$$

$$\text{Dilation: } D_a x(\mathbf{u}) = \frac{1}{a}x\left(\frac{\mathbf{u}}{a}\right), \quad a>0 \tag{1.1.5}$$

$$\text{Rotation: } R_\theta x(\mathbf{u}) = x(r_{-\theta}\mathbf{u}) \tag{1.1.6}$$

where r_θ is the 2-D rotation matrix. Now consider shifts, scales and rotations of our bandpass filter

$$\psi_{\mathbf{b},a,\theta}(\mathbf{u}) = \frac{1}{a}\psi\left(\frac{r_{-\theta}(\mathbf{u}-\mathbf{b})}{a}\right) \tag{1.1.7}$$

which are called the *daughter wavelets*. The 2D CWT of a signal $x(\mathbf{u})$ is defined as

$$CWT_x(\mathbf{b},a,\theta) = \int_{-\infty}^{\infty} \psi^*_{\mathbf{b},a,\theta}(\mathbf{u})x(\mathbf{u})d\mathbf{u} = \langle \psi_{\mathbf{b},a,\theta}(\mathbf{u}),\ x(\mathbf{u})\rangle \tag{1.1.8}$$

1.1.2.1 Properties

The CWT has some particularly nice properties. In particular, it has *covariance* under the three transformations (1.1.4)-(1.1.6):

$$T_{\mathbf{b}_0}x \to CWT_x(\mathbf{b}-\mathbf{b}_0,a,\theta) \tag{1.1.9}$$

$$D_{a_0}x \to CWT_x(\mathbf{b}/a_0, a/a_0, \theta) \tag{1.1.10}$$

$$R_{\theta_0}x \to CWT_x(r_{-\theta_0}\mathbf{b}, a, \theta+\theta_0) \tag{1.1.11}$$

Most importantly, the CWT is now localized in space, which distinguishes it from the Fourier transform. This means that changes in one part of the image will not affect the wavelet coefficients in another part of the image, so long as the distance between the two parts is much larger than the wavelength of the wavelets you are examining.

1.1.2.2 Inverse

The CWT can be inverted by using a *dual* function $\tilde{\psi}$. There are restrictions on what dual function we can use, namely the dual-wavelet pair must have an admissible constant C_ψ that satisfies the cross-admissibility constraint **holschneider_pointwise_1991**. Without going into too much detail about these constraints, we know that we can recover x from CWT_x by:

$$x(\mathbf{u}) = \frac{1}{C_\psi}\int\int\int \frac{1}{a^3}CWT_x(\mathbf{b},a,\theta)\tilde{\psi}_{\mathbf{b},a,\theta}\ d\mathbf{b}dad\theta \tag{1.1.12}$$

1.1.2.3 Interpretation

As the CWT is a convolution with a zero mean function, the wavelet coefficients are only large in the regions of the parameter space $(\mathbf{b}, a, \theta)$ where $\psi_{\mathbf{b},a,\theta}$ 'match' the features of the signal. As the wavelet ψ is well localized, the energy of the coefficients CWT_x will be concentrated on the significant parts of the signal.

For an excellent description of the properties of the CWT in 1-D we recommend [1] and in 2-D we recommend [2].

1.1.3 Discretization and Frames

The CWT is highly redundant. We have taken a 2-D signal and expressed it in 4 dimensions (2 offset, 1 scale and 1 rotation). In reality, we would like to sample the space of the CWT. We would ideally like to fully retain all information in x (be able to reconstruct x from the samples) while sampling over $(\mathbf{b}, a, \theta)$ as little as possible to avoid redundancy.

A set of vectors $\Psi = \{\psi_i\}_{i \in I}$ in a hilbert space $\mathbb{H}$ is a *frame* if there exist two constants $0 < A \leq B < \infty$ such that for all $x \in \mathbb{H}$:

$$A||x||^2 \leq \sum_{i \in I} |\langle x, \psi_i \rangle|^2 \leq B||x||^2 \tag{1.1.13}$$

with A, B called the *frame bounds* [3]. The frame bounds relate to the issue of stable reconstruction. In particular, no vector x with $||x|| > 0$ should be mapped to 0, as this would violate the bound on A from below. This can be interpreted as ensuring our Ψ cover the entire frequency space. The upper bound simply ensures that the transform coefficients are bounded.

Any finite set of vectors that spans the space is frame. An orthogonal basis is a commonly known frame where $A = B = 1$ and $|\psi_i| = 1$ (e.g. the Discrete Wavelet Transform or the Fourier Transform). Tight frames are frames where $A = B$ and Parseval tight frames have the special case $A = B = 1$. It is possible to have frames that have more vectors than dimensions, and this will be the case with many expansions we explore in this thesis.

If $A = B$ and $|\psi_i| = 1$, then A is the measure of the redundancy of the frame. Of course, for the orthogonal basis, $A = 1$ when $|\psi_i| = 1$ so there is no redundancy. For the 2-D DT$\mathbb{C}$WT which we will see shortly, the redundancy is 4.

1.1.3.1 Inversion

(1.1.13) specify the constraints that make a frame representation invertible. The tighter the frame bounds, the more easily it is to invert the signal. This gives us some guide to choosing the sampling grid for the CWT.

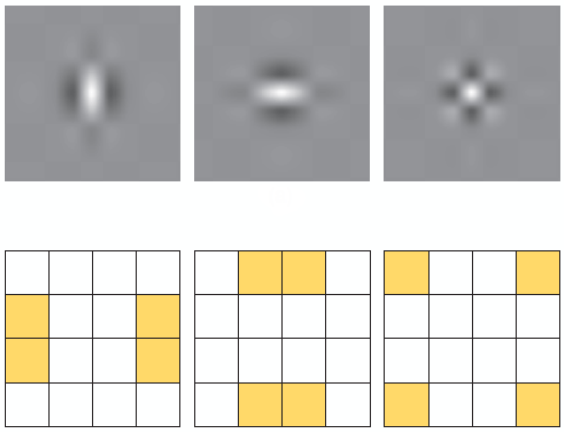


Figure 1.2: **Typical wavelets from the 2D separable DWT.** Top: Wavelet point spread functions for ψ^v (low-high), ψ^h (high-low), and ψ^d (high-high) wavelets. High-high wavelets are in a checkerboard pattern, with no favoured orientation. Bottom: Idealized support of the spectra of each of the wavelets. Image taken from [4].

One particular inverse operator is the *canonical dual frame*. If we define the frame operator $S = \Psi\Psi^*$ then the canonical dual of Ψ is defined as $\tilde{\Psi} = \tilde{\psi}_{i\in I}$ where:

$$\tilde{\psi}_i = S^{-1}\psi_i \tag{1.1.14}$$

then[3]

$$x = \sum_{i\in I}\langle x, \psi_i\rangle\tilde{\psi}_i = \sum_{i\in I}\langle x, \tilde{\psi}_i\rangle\psi_i \tag{1.1.15}$$

If a frame is tight, then so is its dual.

1.1.4 Discrete Wavelet Transform

The 2-D DWT has one scaling fucntion and three wavelet functions, composed of the product of 1-D wavelets in the x and y directions:

$$\phi(x,y) = \phi(x)\phi(y) \tag{1.1.16}$$
$$\psi^h(x,y) = \phi(x)\psi(y) \tag{1.1.17}$$
$$\psi^v(x,y) = \psi(x)\phi(y) \tag{1.1.18}$$
$$\psi^d(x,y) = \psi(x)\psi(y) \tag{1.1.19}$$

with h, v, d indicating the sensitivity to horizontal, vertical and diagonal edges. The point spread functions for the wavelet functions are shown in Figure 1.2.

(1.1.7) gave the equation for the daughter wavelets in 2-D, in 1-D at scales $a = 2^j, j \geq 0$, this is simply:

$$\psi_{b,j}(x) = 2^{-j/2}\psi\left(\frac{x-b}{2^j}\right) \tag{1.1.20}$$

For the four equations above (1.1.16) – (1.1.19), define the daughter wavelets as:

$$\psi_{kl}^{\alpha,j}(x,y) = \psi_{j,k}^{\alpha}(x)\psi_{j,l}^{\alpha} \tag{1.1.21}$$

for $\alpha = h, v, d,\ k, l \in \mathbb{Z}$. We can then get an orthonormal basis with the set $\{\phi_{kl}^{j}, \psi_{kl}^{\alpha,j}\}_{\alpha}$. The wavelet coefficients at chosen scale and location can then be found by taking the inner product of the signal x with the daughter wavelets.

1.1.4.1 Shortcomings

The Discrete Wavelet Transform (DWT) is an orthogonal basis. It is a natural first signal expansion to consider when frustrated with the limitations of the Fourier Transform. It is also a good example of the limitations of non-redundant transforms, as it suffers from several drawbacks:

- The DWT is sensitive to the zero crossings of its wavelets. We would like singularities in the input to yield large wavelet coefficients, but if they fall at a zero crossing of a wavelet, the output can be small. See Figure 1.3.
- They have poor directional selectivity. As the wavelets are purely real, they have passbands in all four quadrants of the frequency plane. While they can pick out edges aligned with the frequency axis, they do not have admissibility for other orientations. See Figure 1.2.
- They are not shift invariant. In particular, small shifts greatly perturb the wavelet coefficients. Figure 1.3 shows this for the centre-left and centre-right images. Figure 1.10 (right) also shows this.

The lack of shift invariance and the possibility of low outputs at singularities is a price to pay for the critically sampled property of the transform. This shortcoming can be overcome with the undecimated DWT [5], [6], but it comes with a heavy computational and memory cost.

1.1.5 Complex Wavelets

Fortunately, we can improve on the DWT with complex wavelets, as they can solve these new shortcomings while maintaining the desired localization properties.

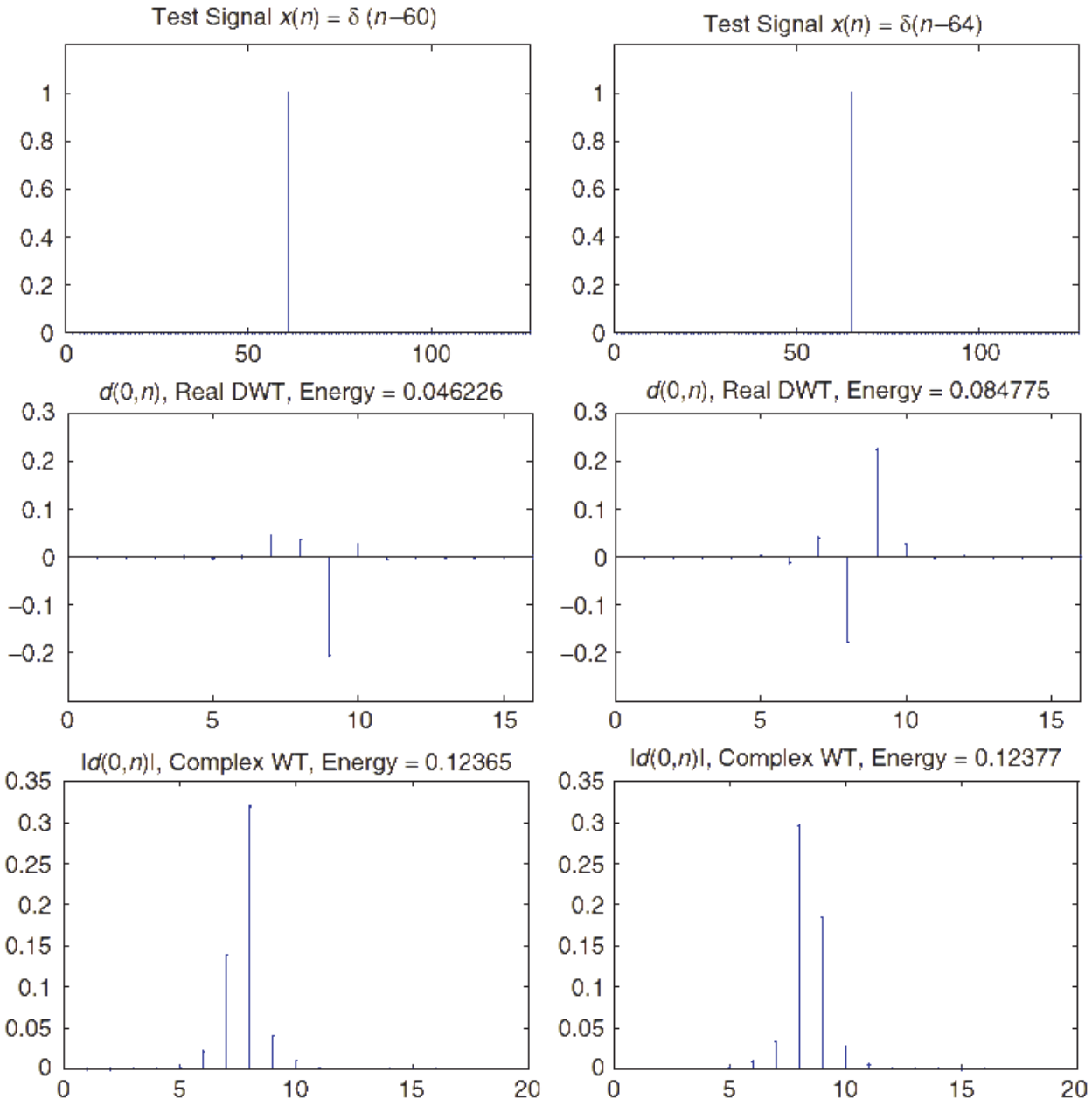


Figure 1.3: **Sensitivity of DWT coefficients to zero crossings and small shifts.** Two impulse signals $\delta(n-60)$ and $\delta(n-64)$ are shown (top), as well as the wavelet coefficients for scale $j=1$ for the DWT (middle) and for the DT$\mathbb{C}$WT (bottom). In the middle row, not only are the coefficients very different from a shifted input, but the energy has almost doubled. As the DWT is an orthonormal transform, this means that this extra energy has come from other scales. In comparison, the energy of the magnitude of the DT$\mathbb{C}$WT coefficients has remained far more constant, as has the shape of the envelope of the output. Image taken from [4].

The Fourier transform does not suffer from a lack of directional selectivity and shift variance, because its basis functions are based on the complex sinusoid:

$$e^{j\omega t} = \cos(\omega t) + j\sin(\omega t) \tag{1.1.22}$$

whereas the DWT's basis functions are based on only the real sinusoid $\cos(\omega t)$.[1] As t moves along the real line, the phase of the Fourier coefficients change linearly, while their magnitude remains constant. In contrast, as t moves along the real line, the sign of the real coefficient flips between -1 and 1, and its magnitude is a rectified sinusoid.

These nice properties come from the fact that the cosine and sine functions of the Fourier transform form a Hilbert Pair and together constitute an analytic signal.

We can achieve these nice properties if the mother wavelet for our wavelet transform is analytic:

$$\psi_c(t) = \psi_r(t) + j\psi_c(t) \tag{1.1.23}$$

where $\psi_r(t)$ and $\psi_c(t)$ form a Hilbert Pair (i.e., they are 90° out of phase with each other).

There are a number of possible ways to do a wavelet transform with complex wavelets. We examine two in particular, a Fourier-based, sampled CWT using Morlet wavelets, and the Dual-Tree Complex Wavelet Transform (DT$\mathbb{C}$WT) developed by Kinsbury [4], [7]–[13].

The Morlet wavelet transform we look at as it is used by Mallat et. al. in their scattering transform [14]–[22], which has been a large inspiration for our work, and we will introduce it shortly. The DT$\mathbb{C}$WT we believe offers several advantages over the Morlet based implementation, and has been the basis for most of our work.

1.1.6 Sampled Morlet Wavelets

The wavelet transform used by Mallat et. al. in their scattering transform are an efficient implementation of the Gabor Transform. While the Gabor wavelets have the best theoretical trade-off between spatial and frequency localization, they have a non-zero mean. This violates (1.1.3) making them inadmissible as wavelets. Instead, the Morlet wavelet has the same shape, but with an extra degree of freedom chosen to set $\int \psi(\mathbf{u})d\mathbf{u} = 0$. This wavelet has equation (in 2D):

$$\psi(\mathbf{u}) = \frac{1}{2\pi\sigma^2}(e^{i\mathbf{u}\xi} - \beta)e^{-\frac{|\mathbf{u}|^2}{2\sigma^2}} \tag{1.1.24}$$

where β is usually $<< 1$ and is this extra degree of freedom, σ is the size of the gaussian window, and ξ is the location of the peak frequency response — i.e., for an octave based transform, $\xi = 3\pi/4$.

Bruna and Mallat add an extra degree of freedom in their original design [15] by allowing for a non-circular Gaussian window over the complex sinusoid, which gives control over the

[1] we have temporarily switched to 1D notation here as it is clearer and easier to use, but the results still hold for 2D

Figure 1.4: **Single Morlet filter with varying slants and window sizes.** Top left — 45° plane wave (real part only). Top right — plane wave with $\sigma = 3, \gamma = 1$. Bottom left — plane wave with $\sigma = 3, \gamma = 0.5$. Bottom right — plane wave with $\sigma = 2, \gamma = 0.5$.

angular resolution of the final wavelet. (1.1.24) now becomes:

$$\psi(\mathbf{u}) = \frac{\gamma}{2\pi\sigma^2}(e^{i\mathbf{u}\xi} - \beta)e^{-\mathbf{u}^\mathbf{t}\Sigma^{-1}\mathbf{u}} \tag{1.1.25}$$

Where

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{2\sigma^2} & 0 \\ 0 & \frac{\gamma^2}{2\sigma^2} \end{bmatrix}$$

The effects of modifying the eccentricity parameter γ and the window size σ are shown in Figure 1.4. A full family of Morlet wavelets at varying scales and orientations is shown in Figure 1.5.

1.1.6.1 Invertibility and Energy Conservation

We can write the wavelet transform of an input x as

$$\mathcal{W}x = \{x * \phi_J, x * \psi_\lambda\}_\lambda \tag{1.1.26}$$

The ℓ^2 norm of the wavelet transform is then defined by

$$\|\mathcal{W}x\|^2 = \|x * \phi_J\|^2 + \sum_\lambda \|x * \psi_\lambda\|^2 \tag{1.1.27}$$

An energy preserving transform will satisfy Plancherel's equality, so that

$$\|\mathcal{W}x\| = \|x\| \tag{1.1.28}$$

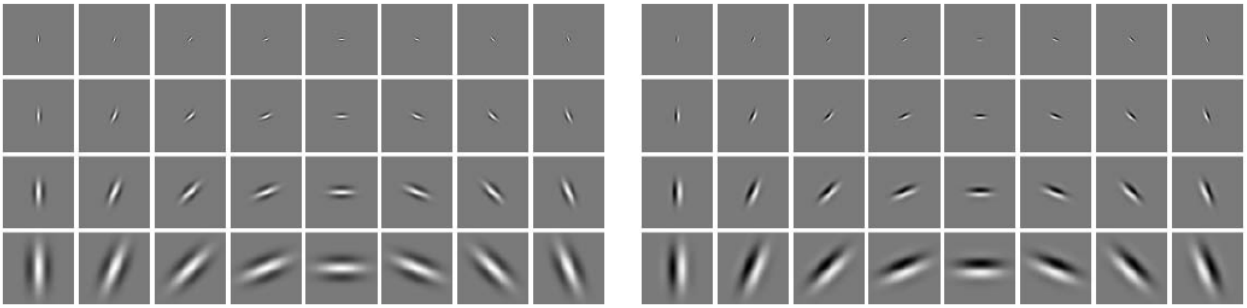


Figure 1.5: **The full dictionary of Morlet wavelets used by Mallat.** The real filters are on the left and the imaginary on the right. The first row correspond to scale $j = 1$, increasing up to $j = 4$. The first column corresponding to $\theta = 0$, rotating through $\pi/8$ up to the eighth column of $7\pi/8$, $\gamma = 1/2$.

which is a nice property to have for invertibility, as well as for analysing how different signals get transformed (e.g. white noise versus standard images).

For a transform to be invertible, we examine the measure of how tightly its basis functions tile the Fourier plane with its Littlewood-Paley function:

$$A(\omega) = |\mathcal{F}\phi_J(\omega)|^2 + \sum_{\lambda} |\mathcal{F}\psi_\lambda(\omega)|^2 \tag{1.1.29}$$

If the tiling is α-tight, then $\forall\omega \in \mathbb{R}^2$:

$$1 - \alpha \leq A(\omega) \leq 1 \tag{1.1.30}$$

and the wavelet operator, $\mathcal{W}$ is an α frame. If $A(\omega)$ is ever close to 0, then there is not a good coverage of the frequency plane at that location. If it ever exceeds 1, then there is overlap between bases. Both of these conditions make invertibility difficult[2]. Figure 1.6 show the invertibility of a few families of wavelets used by Mallat et. al.. Invertibility is possible, but not guaranteed for all configurations. The Fourier transform of the inverse filters are defined by:

$$\mathcal{F}\phi_J^{-1}(\omega) = A(\omega)^{-1}\mathcal{F}\phi_J(\omega) \tag{1.1.31}$$

$$\mathcal{F}\psi_\lambda^{-1}(\omega) = A(\omega)^{-1}\mathcal{F}\psi_\lambda(\omega) \tag{1.1.32}$$

[2]In practise, if $A(\omega)$ is only slightly greater 1 for only a few small areas of ω, approximate inversion can be achieved

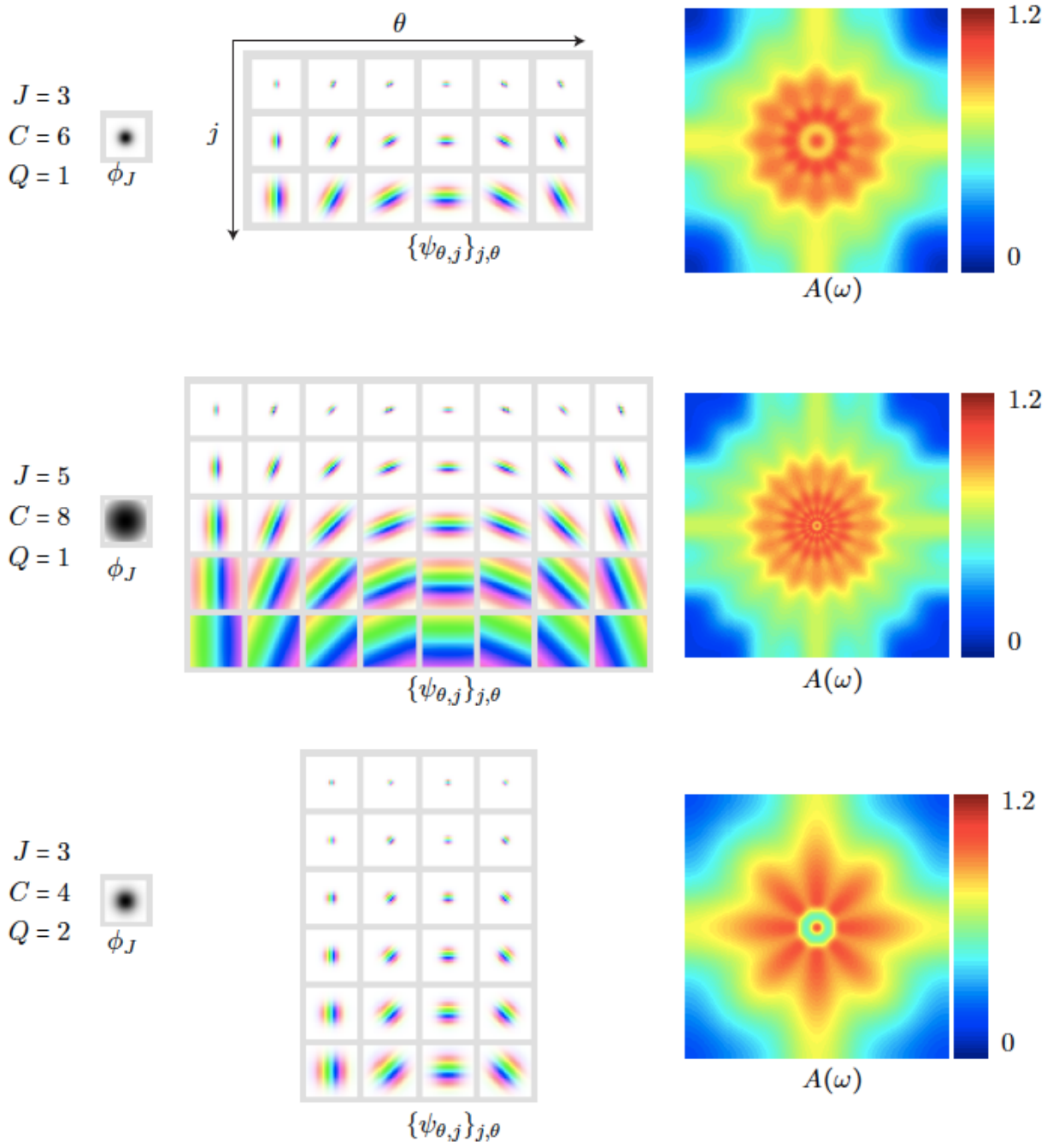


Figure 1.6: Three Morlet Wavelet Families and their tiling of the frequency plane. For each set of parameters, the point spread functions of the wavelet bases are shown, next to their Littlewood-Paley sum $A(\omega)$. None of the configurations cover the corners of the frequency plane, and they often exceed 1. Increasing J, L (Sifre uses C in these diagrams) or Q gives better frequency localization but at the cost of spatial localization and added complexity. Image taken from [21].

1.1.7 The DT$\mathbb{C}$WT

The DT$\mathbb{C}$WT was first proposed by Kingsbury in [8], [9] as a way to combat many of the shortcomings of the DWT, in particular, its poor directional selectivity, and its poor shift invariance. A thorough analsysis of the properties and benefits of the DT$\mathbb{C}$WT is done in [4], [10]. Building on these properties, it been used successfully for denoising and inverse problems [23]–[26], texture classification [27], [28], image registration [29], [30] and SIFT-style keypoint generation matching [31]–[35] amongst many other applications.

Compared to Gabor (or Morlet) image analysis, the authors of [4] sum up the dangers as:

> A typical Gabor image analysis is either expensive to compute, is noninvertible, or both.

This nicely summarises the difference between this method and the Fourier based method outlined in subsection 1.1.6. The DT$\mathbb{C}$WT is a filter bank (FB) based wavelet transform. It is faster to implement than the Morlet analysis, as well as being more readily invertible.

1.1.7.1 Deisgn Criteria for the DT$\mathbb{C}$WT

It was stated in subsection 1.1.5 that if the mother (and daughter) wavelets were complex, with their real and imaginary parts forming a Hilbert pair, then the wavelet transform of a signal with these $\{\psi_{j,n}\}_{j,n}$ would give a representation that had nice shift properties[3], was insensitive to zero crossings of the wavelet, and had good directional selectivity.

As in subsection 1.1.5, we want to have a complex mother wavelet ψ_c that satisfies (1.1.23), but now achieved with filter banks. A slight deviation from standard filter bank notation, where h_0, h_1 are the analysis and g_0, g_1 are the synthesis filters. We define:

- h_0, h_1 the low and high-pass analysis filters for ψ_r (henceforth called ψ_h)
- g_0, g_1 the low and high-pass analysis fitlers for ψ_i (henceforth called ψ_g)
- $\tilde{h}_0, \tilde{h}_1$ the low and high-pass synthesis filters for $\tilde{\psi}_h$.
- $\tilde{g}_0, \tilde{g}_1$ the low and hig pass synthesis filters for $\tilde{\psi}_g$.

[3] in particular, that a shift in input gives the same shift in magnitude of the wavelet coefficients, and a linear phase shift

Figure 1.7: Analysis FB for the DT$\mathbb{C}$WT@. Top 'tree' forms the real component of the complex wavelet ψ_r, and the bottom tree forms the imaginary (Hilbert pair) component ψ_i. Image taken from [4].

The dilation and wavelet equations for a 1D filter bank implementation are:

$$\phi_h(t) = \sqrt{2}\sum_n h_0(n)\phi_h(2t-n) \tag{1.1.33}$$

$$\psi_h(t) = \sqrt{2}\sum_n h_1(n)\phi_h(2t-n) \tag{1.1.34}$$

$$\phi_g(t) = \sqrt{2}\sum_n g_0(n)\phi_g(2t-n) \tag{1.1.35}$$

$$\psi_g(t) = \sqrt{2}\sum_n g_1(n)\phi_g(2t-n) \tag{1.1.36}$$

This implementation is shown in Figure 1.7.

Designing a filter bank implementation that results in Hilbert symmetric wavelets does not appear to be an easy task. However, it was shown by Kingsbury in [10] (and later proved by Selesnick in [36]) that the necessary conditions are conceptually very simple. One low-pass filter must be a *half-sample shift* of the other. I.e.,

$$g_0(n) \approx h_0(n-0.5) \rightarrow \psi_g(t) \approx \mathcal{H}\{\psi_h(t)\} \tag{1.1.37}$$

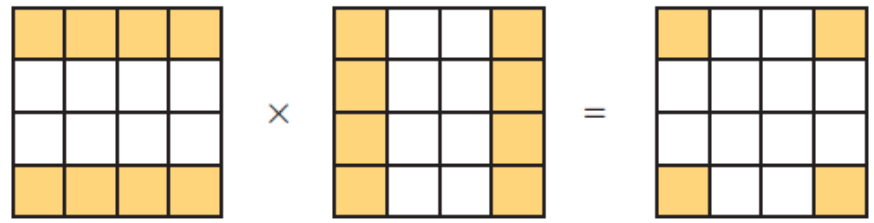
As the DT$\mathbb{C}$WT is designed as an invertible filter bank implementation, this is only one of the constraints. Naturally, there are also:

- Perfect reconstruction
- Finite support
- Linear phase
- Many vanishing moments at $z = -1$ for good stopband properties

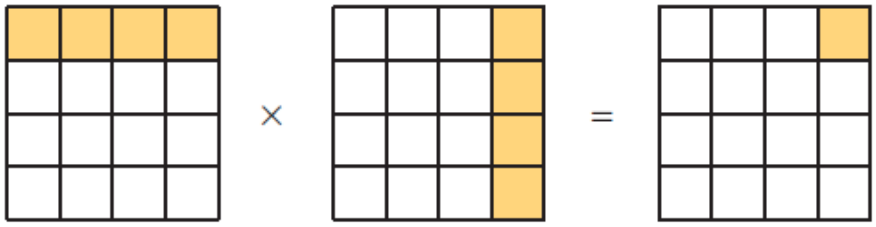
to consider when building the h's and g's. The derivation of the filters that meet these conditions is covered in detail in [13], [37], and in general in [4]. The result is the option of three families of filters: biorthogonal filters ($h_0[n] = h_0[N-1-n]$ and $g_0[n] = g_0[N-n]$), q-shift filters ($g_0[n] = h_0[N-1-n]$), and common-factor filters.

1.1.7.2 The Resulting Wavelets and their Properties

While analytic wavelets in 1D are useful for their shift invariance, the real beauty of the DT$\mathbb{C}$WT is in its ability to make a separable 2D wavelet transform with oriented wavelets.



(a)



(b)

Figure 1.8: (a) The high-high DWT wavelet having a passband in all 4 corners of the frequency plane vs (b) the high-high DTℂWT wavelet frequency support only existing in one quadrant. Taken from [4]

1.8a shows the spectrum of the wavelet when the separable product uses purely real wavelets, as is the case with the DWT. 1.8b however, shows the separable product of two complex, analytic wavelets resulting in a localized and oriented 2D wavelet.

I.e., for the $+45°$ wavelet[4] (which is high in both ω_1 and ω_2), the separable product is:

$$\begin{aligned}\psi(\omega_1,\omega_2) &= \psi_c(\omega_1)\overline{\psi_c(\omega_2)} && (1.1.38)\\ &= (\psi_h(\omega_1)+j\psi_g(\omega_1)\overline{(\psi_h(\omega_2)+\psi_g(\omega_2))}\\ &= \psi_h(\omega_1)\psi_h(\omega_2)+\psi_g(\omega_1)\psi_g(\omega_2)\\ &+ j\left(\psi_g(\omega_1)\psi_h(\omega_2)-\psi_h(\omega_1)\psi_g(\omega_2)\right) && (1.1.39)\end{aligned}$$

Similar equations can be obtained for the other five wavelets and the scaling function, by replacing ψ with ϕ for both directions, and not taking the complex conjugate in (1.1.40) to get the right hand side of the frequency plane.

Figure 1.9 shows the resulting wavelets both in the spatial domain and their idealized support in the frequency domain.

Figure 1.10 shows how the DTℂWT compares with the DWT with a shifting input.

[4] note that 1.8b shows the $135°$ wavelet

Figure 1.9: Wavelets from the 2d DT$\mathbb{C}$WT@. **Top:** The six oriented filters in the space domain (only the real wavelets are shown). **Bottom:** Idealized support of the Fourier spectrum of each wavelet in the 2D frequency plane. Spectra of the the real wavelets are shown — the spectra of the complex wavelets $(\psi_h + j\psi_g)$ only has support in the top half of the plane. Image taken from [4].

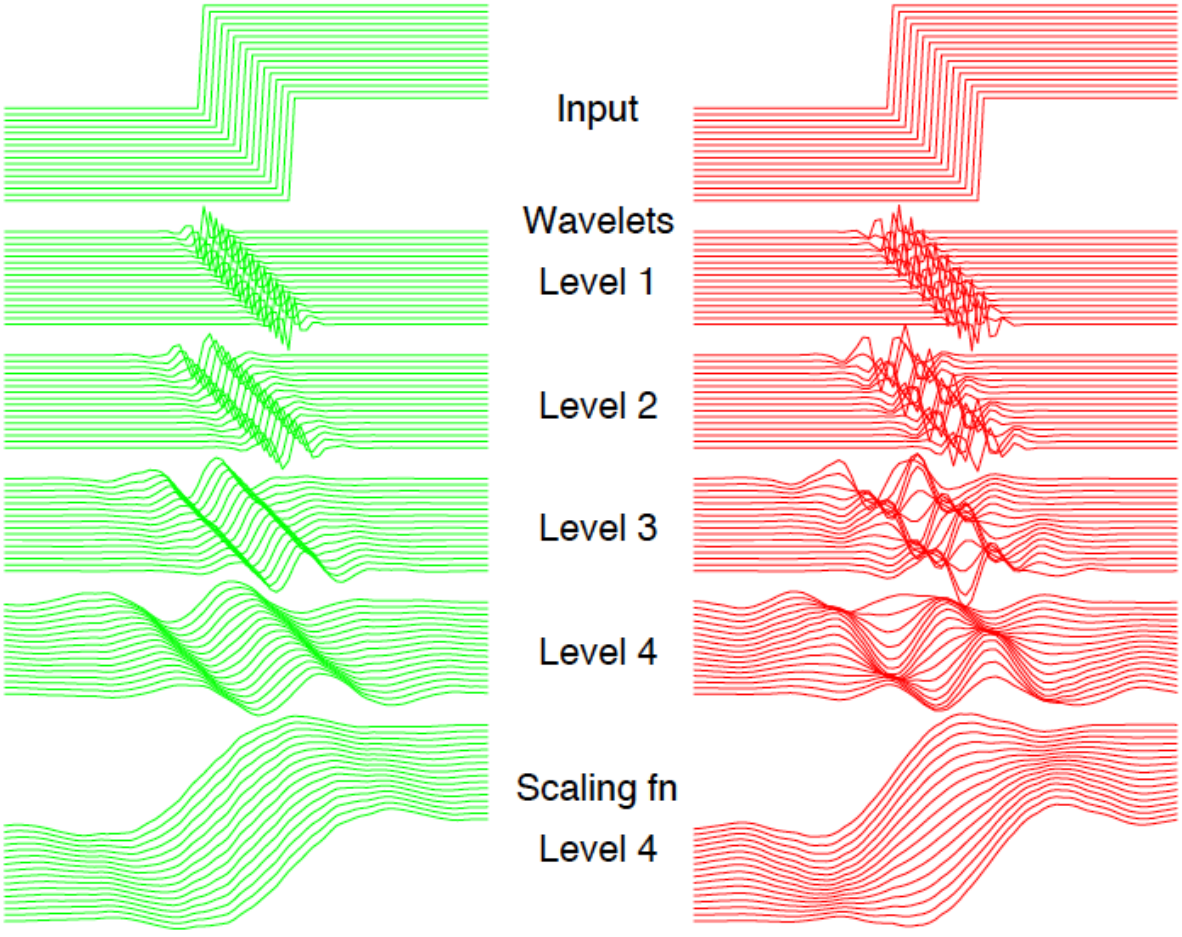


Figure 1.10: **The shift invariance ofthe** DT$\mathbb{C}$WT **vs. the real DWT.** The DT$\mathbb{C}$WT (left) linearizes shifts in the phase change of the complex wavelet. The DWT (right) cannot do this and suffers from shift variance. Image taken from [8].

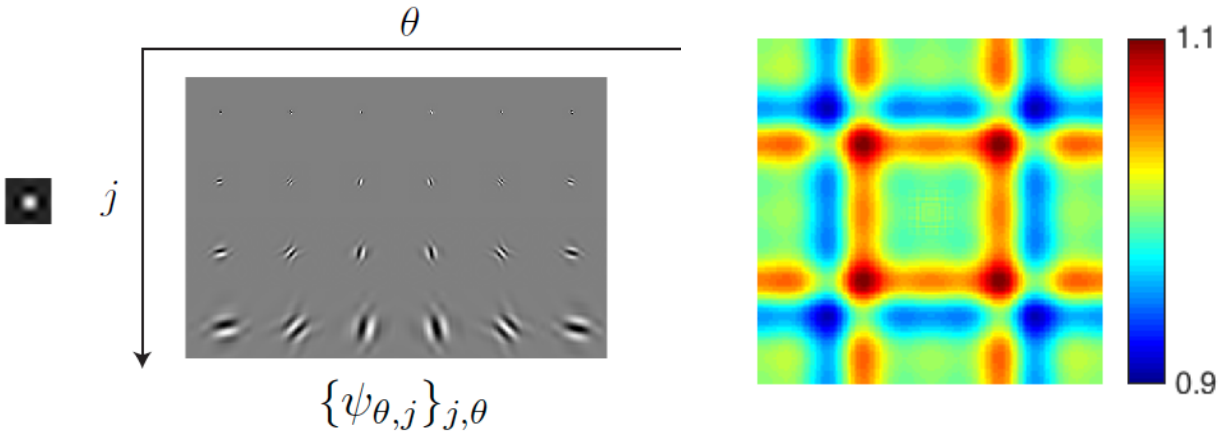


Figure 1.11: Four scales of the DT$\mathbb{C}$WT (left) and its associated frequency coverage, or $A(\omega)$ (right). Note the reduced scale compared to Figure 1.6.

1.1.7.3 Invertibility and Energy Conservation

We analysed the Littlewood-Paley function for the Morlet-Fourier implementation, and saw what areas of the spectrum were better covered than others. How about for the DT$\mathbb{C}$WT?

It is important to note that in the case of the DT$\mathbb{C}$WT, the wavelet transform is also approximately unitary, i.e.,

$$\|x\|^2 \approx \|\mathcal{W}x\|^2 \tag{1.1.40}$$

and the implementation is perfectly invertible as the Littlewood-Paley function is unity (or very near unity) $\forall\omega$. See Figure 1.11. This is not a surprise, as it is a design constraint in choosing the filters, but nonetheless is important to note.

A beneficial property of energy conservation is that the noise in the input will equal the noise in the wavelet coefficients. When we introduce Scatternets, we can show that we can keep the unitary property in the scattering coefficients. This is an important property, particularly in light of the recent investigations in [38]. This paper saw that it is easy to find cases in CNNs where a small amount of input perturbation results in a completely different class label (see Figure 1.12). Having a unitary transform limits the amount the features can change, which will make the entire network more stable to distortion and noise.

1.1.8 Summary of Methods

One final comparison to make between the DT$\mathbb{C}$WT and the Morlet wavelets is their frequency coverage. The Morlet wavelets can be made to be tighter than the DT$\mathbb{C}$WT, which gives better angular resolution — see Figure 1.13. However it is not always better to keep getting finer

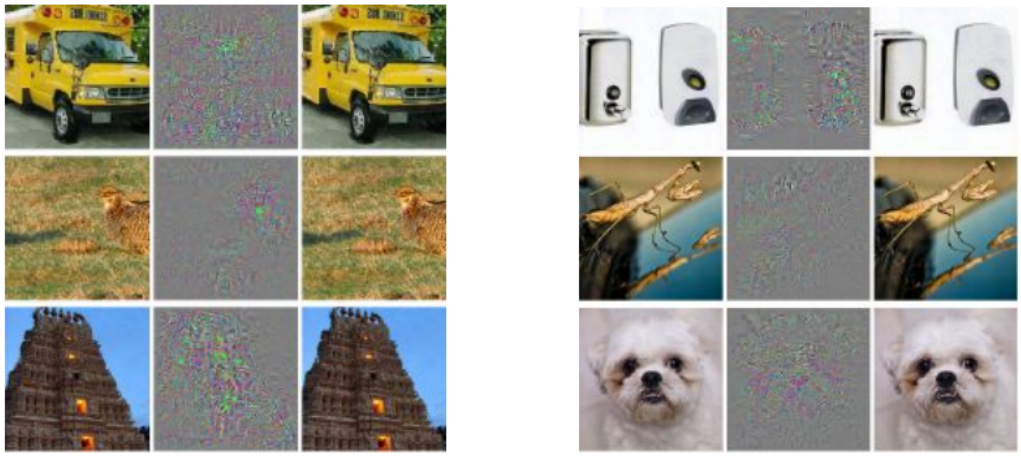
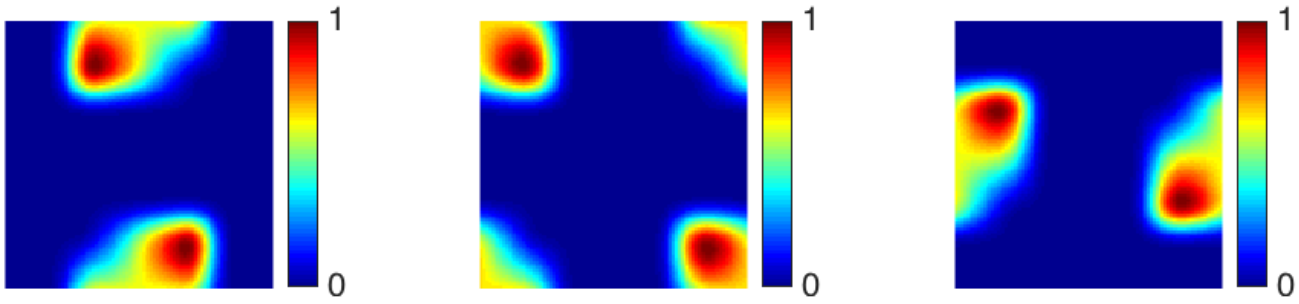


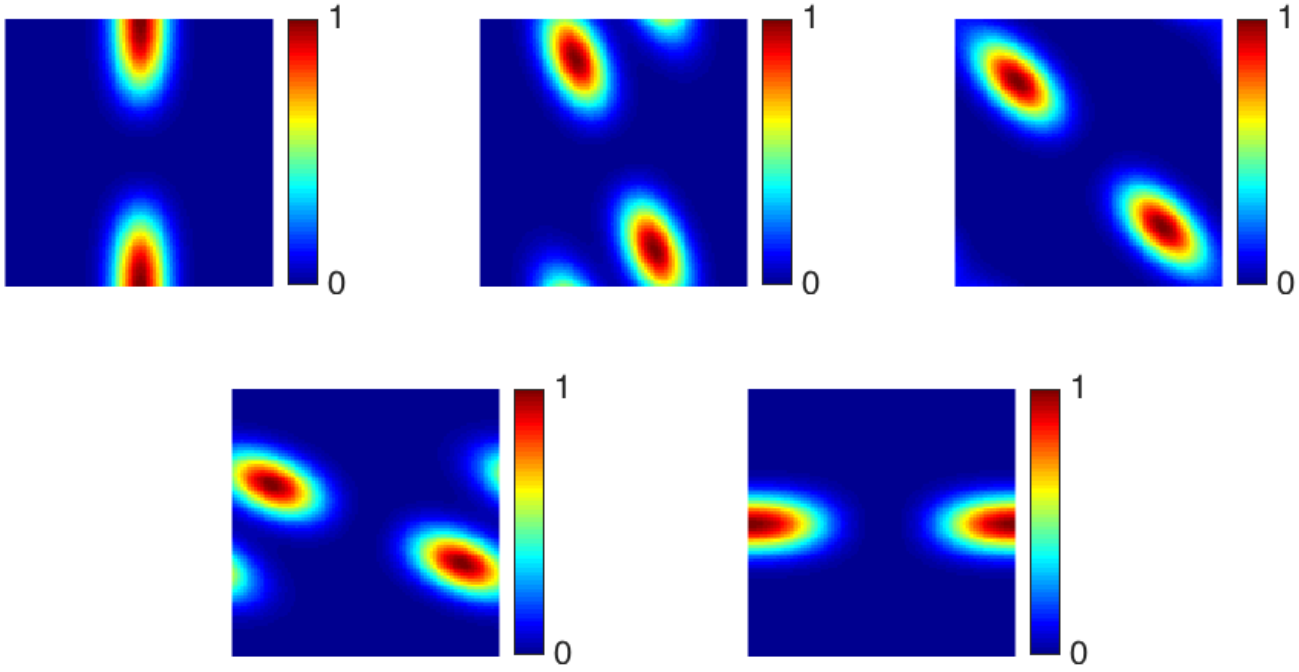
Figure 1.12: **Two adversarial examples generated for AlexNet.** The left column shows a correctly predicted sample, the right column an incorrectly predicted example, and the centre column the difference between the two images, magnified 10 times. Image taken from [38].

and finer resolutions, indeed the Fourier transform gives the ultimate in angular resolution, but as mentioned, this makes it less stable to shifts and deformations.

?? compares the advantages and disadvantages of the wavelet methods discussed in this chapter.



(a) DTℂWT wavelets (left to right) — 15°, 45° and 75°



(c) Morlet wavelets (left to right) — 0°, 22.5°, 45°, 67.5°, and 90°

Figure 1.13: Normalized Energy spectra of the DTℂWT wavelets versus the preferred 8 orientation Morlet wavelets by Mallat for the second quadrant. Orientations listed refer to the edge orientation in the spatial domain that gives the highest response. All wavelets have been normalized to be between zero and one. The Morlet wavelets have finer angular resolution, which can give better discrimination, at the cost of decreasing stability to deformations, and requiring larger spatial support.

References

[1] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, 2nd ed., ser. Prentice Hall Signal Processing Series. Prentice Hall PTR, 2007.

[2] J. Antoine, R. Murenzi, P. Vandergheynst, and S. Ali, *Two-Dimensional Wavelets and Their Relatives*. 2004.

[3] J. Kovacevic and A. Chebira, *An Introduction to Frames*. Hanover, MA, USA: Now Publishers Inc., 2008.

[4] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform", *Signal Processing Magazine, IEEE*, vol. 22, no. 6, pp. 123–151, 2005.

[5] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[6] R. R. Coifman and D. L. Donoho, "Translation-Invariant De-Noising", en, in *Wavelets and Statistics*, ser. Lecture Notes in Statistics 103, A. Antoniadis and G. Oppenheim, Eds., Springer New York, 1995, pp. 125–150.

[7] N. Kingsbury and J. Magarey, "Wavelet transforms in image processing", A. Prochazka, J. Uhlir, and P. Sovka, Eds., 1997.

[8] N. Kingsbury, "The Dual-Tree Complex Wavelet Transform: A New Technique For Shift Invariance And Directional Filters", in *1998 8th International Conference on Digital Signal Processing (DSP)*, Utah, Aug. 1998, pp. 319–322.

[9] ——, "The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement", in *Signal Processing Conference (EUSIPCO 1998), 9th European*, Sep. 1998, pp. 1–4.

[10] N. Kingsbury, "Image processing with complex wavelets", *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, vol. 357, no. 1760, pp. 2543–2560, Sep. 1999.

[11] ——, "Shift invariant properties of the dual-tree complex wavelet transform", in *Icassp '99: 1999 Ieee International Conference on Acoustics, Speech, and Signal Processing, Proceedings Vols I-Vi*, 1999, pp. 1221–1224.

[12] ——, "A dual-tree complex wavelet transform with improved orthogonality and symmetry properties", 2000.

[13] ——, "Complex wavelets for shift invariant analysis and filtering of signals", *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.

[14] J. Bruna and S. Mallat, "Classification with scattering operators", in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2011, pp. 1561–1566.

[15] ——, "Invariant Scattering Convolution Networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.

[16] J. Bruna, "Scattering Representations for Recognition", Theses, Ecole Polytechnique X, Feb. 2013.

[17] E. Oyallon, S. Mallat, and L. Sifre, "Generic Deep Networks with Wavelet Scattering", *arXiv:1312.5940 [cs]*, Dec. 2013. arXiv: 1312.5940 [cs].

[18] E. Oyallon and S. Mallat, "Deep Roto-Translation Scattering for Object Classification", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2865–2873.

[19] L. Sifre and S. Mallat, "Rotation, Scaling and Deformation Invariant Scattering for Texture Discrimination", in *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2013, pp. 1233–1240.

[20] L. Sifre and S. Mallat, "Rigid-Motion Scattering for Texture Classification", *arXiv:1403.1687 [cs]*, Mar. 2014. arXiv: 1403.1687 [cs].

[21] L. Sifre, "Rigid-Motion Scattering for Image Classification", PhD Thesis, Ecole Polytechnique, Oct. 2014.

[22] L. Sifre and J. Anden, *ScatNet*, École normale supérieure, Nov. 2013.

[23] P. de Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets", in *2001 International Conference on Image Processing, 2001. Proceedings*, vol. 2, Oct. 2001, 273–276 vol.2.

[24] Y. Zhang and N. Kingsbury, "A Bayesian wavelet-based multidimensional deconvolution with sub-band emphasis", in *Engineering in Medicine and Biology Society*, 2008, pp. 3024–3027.

[25] G. Zhang and N. Kingsbury, "Variational Bayesian image restoration with group-sparse modeling of wavelet coefficients", *Digital Signal Processing*, Special Issue in Honour of William J. (Bill) Fitzgerald, vol. 47, pp. 157–168, Dec. 2015.

[26] M. Miller and N. Kingsbury, "Image denoising using derotated complex wavelet coefficients", eng, *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, vol. 17, no. 9, pp. 1500–1511, Sep. 2008.

[27] S. Hatipoglu, S. K. Mitra, and N. Kingsbury, "Texture classification using dual-tree complex wavelet transform", in *Seventh International Conference on Image Processing and Its Applications*, 1999, pp. 344–347.

[28] P. de Rivaz and N. Kingsbury, "Complex wavelet features for fast texture image retrieval", in *1999 International Conference on Image Processing, 1999. ICIP 99. Proceedings*, vol. 1, 1999, 109–113 vol.1.

[29] P. Loo and N. G. Kingsbury, "Motion-estimation-based registration of geometrically distorted images for watermark recovery", P. W. Wong and E. J. Delp III, Eds., Aug. 2001, pp. 606–617.

[30] H. Chen and N. Kingsbury, "Efficient Registration of Nonrigid 3-D Bodies", *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 262–272, Jan. 2012.

[31] J. Fauqueur, N. Kingsbury, and R. Anderson, "Multiscale keypoint detection using the dual-tree complex wavelet transform", in *Image Processing, 2006 IEEE International Conference On*, IEEE, 2006, pp. 1625–1628.

[32] R. Anderson, N. Kingsbury, and J. Fauqueur, "Determining Multiscale Image Feature Angles from Complex Wavelet Phases", en, in *Image Analysis and Recognition*, ser. Lecture Notes in Computer Science 3656, M. Kamel and A. Campilho, Eds., Springer Berlin Heidelberg, Sep. 2005, pp. 490–498.

[33] ——, "Rotation-invariant object recognition using edge profile clusters", in *Signal Processing Conference, 2006 14th European*, IEEE, 2006, pp. 1–5.

[34] P. Bendale, W. Triggs, and N. Kingsbury, "Multiscale keypoint analysis based on complex wavelets", in *BMVC 2010-British Machine Vision Conference*, BMVA Press, 2010, pp. 49–1.

[35] E. S. Ng and N. G. Kingsbury, "Robust pairwise matching of interest points with complex wavelets", *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3429–3442, 2012.

[36] I. Selesnick, "Hilbert transform pairs of wavelet bases", *IEEE Signal Processing Letters*, vol. 8, no. 6, pp. 170–173, Jun. 2001.

[37] N. Kingsbury, "Design of Q-shift complex wavelets for image processing using frequency domain energy minimization", in *2003 International Conference on Image Processing, 2003. ICIP 2003. Proceedings*, vol. 1, Sep. 2003, I-1013-16 vol.1.

[38] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks", *arXiv:1312.6199 [cs]*, Dec. 2013. arXiv: 1312.6199 [cs].

Appendix A

Shift Invariance of the DT$\mathbb{C}$WT

Firstly, let us look at what would happen if we retained only one of the subbands. Note that we have to keep the same band from each tree. For any pair of coefficients on the tree, this would look like Figure A.2. E.g. if we kept x_{001a} and x_{001b} then $M = 8$ and $A(z) = H_{0a}(z)H_{00a}(z^2)H_{001a}(z^4)$ is the transfer fucntion from x to x_{001a}. The transfer functions for $B(z)$, $C(z)$ and $D(z)$ are obtained similarly. It is well known that:

$$U(z) \downarrow M \rightarrow U(z^M) \tag{A.0.1}$$

$$U(z) \uparrow M \rightarrow \frac{1}{M} \sum_{k=0}^{M-1} U(W^k z^{1/M}) \tag{A.0.2}$$

Where $W = e^{j2\pi/M}$. So downsampling followed by upsampling becomes:

$$U(z) \downarrow M \uparrow M \rightarrow \frac{1}{M} \sum_{k=0}^{M-1} U(W^k z) \tag{A.0.3}$$

This means that

$$Y(z) = Y_a(z) + Y_b(z) = \frac{1}{M} \sum_{k=0}^{M-1} X(W^k z)[A(W^k z)C(z) + B(W^k z)D(z)] \tag{A.0.4}$$

The aliasing terms for which are everywhere where $k \neq 0$ (as $X(W^k z)$ is $X(z)$ shifted by $\frac{2k\pi}{M}$). I.e. to avoid aliasing in this reduced tree, we want to make sure that $A(W^k z)C(z) + B(W^k z)D(z) = 0$ for all $k \neq 0$.

The figure below (Fig 5 from [13]) shows what $A(W^k z)$ and $C(z)$ look like for both the lowpass case (left) and the highpass case (right). Note that these responses will be similar for $B(W^k z)$ and $D(z)$. For large values of k, there is almost no overlap (i.e. $A(W^k z)C(z) \approx B(W^k z)D(z) \approx 0$, but for small values of k (particularly $k = \pm 1$), the transition bands have significant overlap with the central response. It is here that we need to use the flexibility of

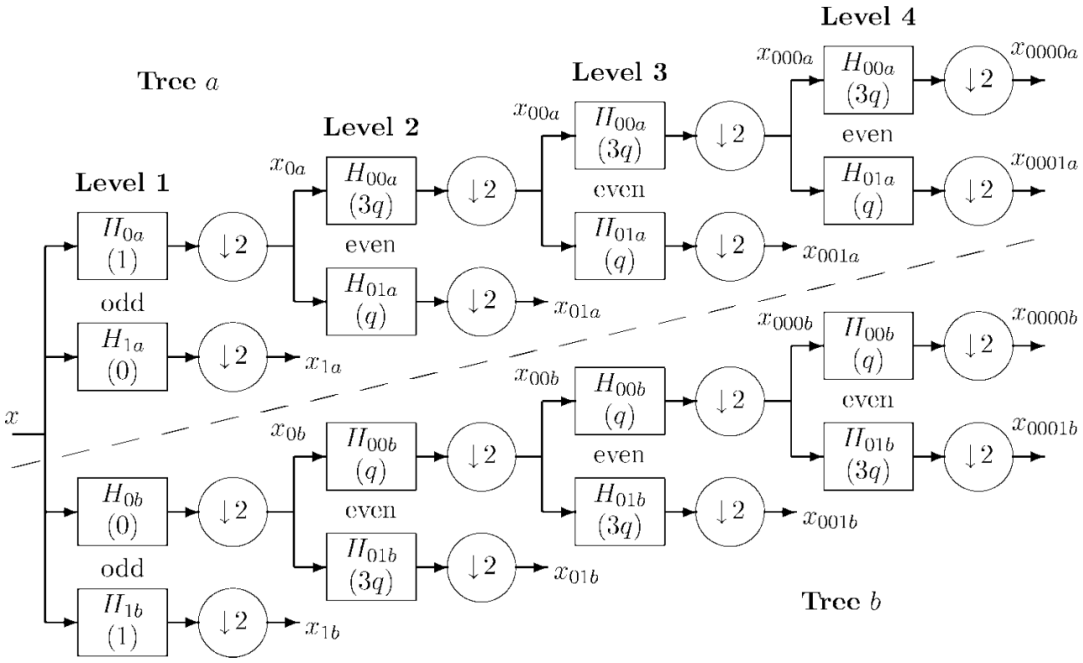


FIG. 3. The Q-shift version of the DT CWT, giving real and imaginary parts of complex coefficients from tree a and tree b, respectively. Figures in brackets indicate the delay for each filter, where $q = \frac{1}{4}$ sample period.

Figure A.1: **Full 1-D** DTℂWT**.**

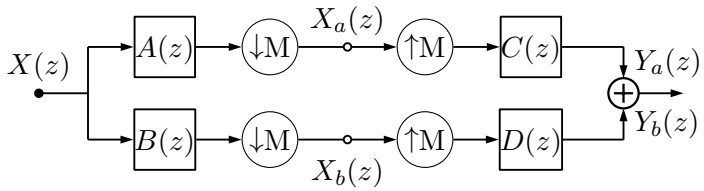
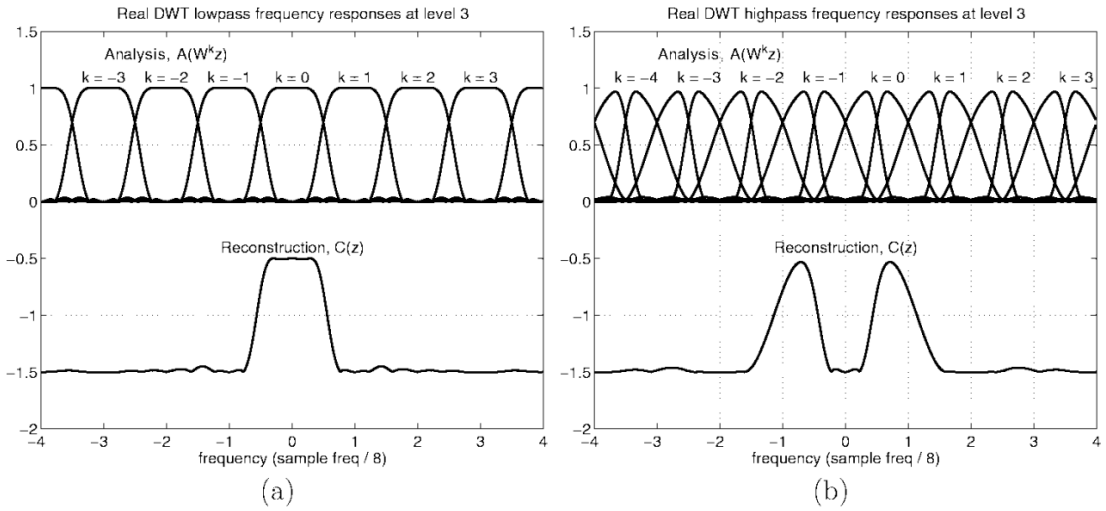


Figure A.2: **Block Diagram of 1-D** DTℂWT**.** Note the top and bottom paths are through the wavelet or scaling functions from just level m ($M = 2^m$). Figure based on Figure 4 in [13].



having 2 trees to ensure that $A(W^k z)C(z)$ and $B(W^k z)D(z)$ cancel out. To do this, let us consider the lowpass filters first. If we let:

$$B(z) = z^{\pm M/2} A(z) \tag{A.0.5}$$

$$D(z) = z^{\mp M/2} C(z) \tag{A.0.6}$$

Then

$$A(W^k z)C(z) + B(W^k z)D(z) = A(W^k z)C(z) + (W^k z)^{\pm M/2} A(W^k z) z^{\mp M/2} C(z) \tag{A.0.7}$$

$$= A(W^k z)C(z) + e^{\frac{jk2\pi}{M} \times (\pm \frac{M}{2})} z^{\pm M/2} z^{\mp M/2} A(W^k z)C(z) \tag{A.0.8}$$

$$= A(W^k z)C(z) + (-1)^k A(W^k z)C(z) \tag{A.0.9}$$

which cancel when k is odd.

Now consider the bandpass case. For shifts of $k = 1$ the right half of the left peak overlaps with the left half of the right peak. For a shift of $k = 2$, the left half of the left peak overlaps with the right half of the right peak. Similarly for $k = -1$ and $k = -2$. For $|k| > 2$, there is no overlap. The fact that we have overlaps at both even and odd shifts of k means that we can't use the same clever trick from before. However, what we do note is that the overlap is always caused by opposite peaks (i.e. the left with the right peak, and never the left with itself, or the right with itself). The solution then is to have B and D have upper and lower passbands of opposite polarity, whil A and C should have passbands of the same polarity.

Consider two prototype complex filters $P(z)$ and $Q(z)$ each with single passbands going from $f_s/2M \to f_s/M$ (or $\frac{\pi}{M} \to \frac{2*\pi}{M}$) - they must be complex to have support in only one half of the frequency plane. Now say $P^*(z) = \sum_r p_r^* z^{-r}$ is the z-transform of the conjugate of p_r, which has support only in the negative half of the frequency plane. Then we can get the

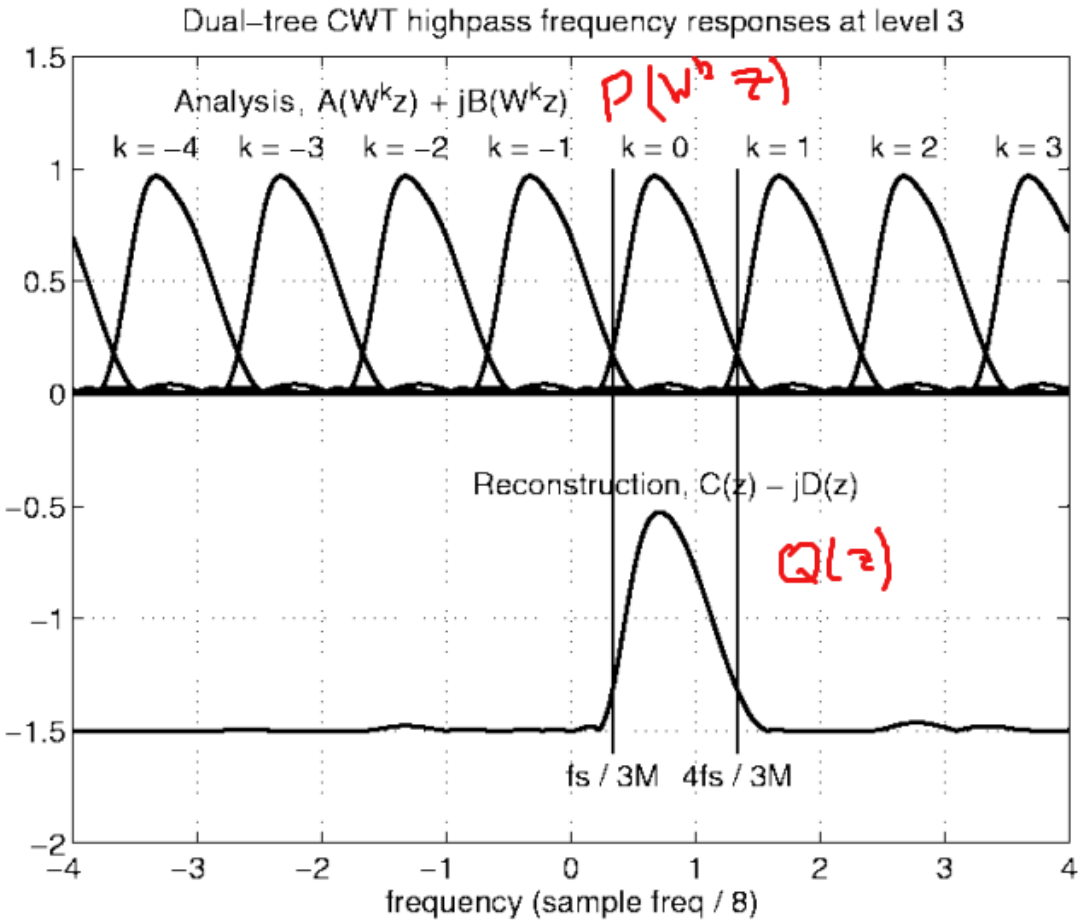
required filters by:

$$A(z) = 2\Re[P(z)] = P(z) + P^*(z) \tag{A.0.10}$$
$$B(z) = 2\Im[P(z)] = -j[P(z) - P^*(z)] \tag{A.0.11}$$
$$C(z) = 2\Re[Q(z)] = Q(z) + Q^*(z) \tag{A.0.12}$$
$$D(z) = -2\Im[Q(z)] = j[Q(z) - Q^*(z)] \tag{A.0.13}$$

Then:

$$\begin{aligned} A(W^k z)C(z) + B(W^k z)D(z) = &\left[P(W^k z) + P^*(W^k z)\right][Q(z) + Q^*(z)] + \\ &(-j * j)\left[P(W^k z) - P^*(W^k z)\right][Q(z) - Q^*(z)] \qquad \text{(A.0.14)} \\ = &P(W^k z)Q(z)[1+1] + P^*(W^k z)Q(z)[1-1] + \\ &P(W^k z)Q^*(z)[1-1] + P^*(W^k z)Q^*(z)[1+1] \qquad \text{(A.0.15)} \\ = &2P(W^k z)Q(z) + 2P^*(W^k z)Q^*(z) \qquad \text{(A.0.16)} \end{aligned}$$

So now we only need to ensure that $P(W^k z)$ overlaps as little as possible with $Q(z)$. This is somewhat more manageable, the diagram below shows the problem.


Dual–tree CWT highpass frequency responses at level 3
Analysis, A(W^k z) + jB(W^k z)
P(W^k z)
k = −4
k = −3
k = −2
k = −1
k = 0
k = 1
k = 2
k = 3
Reconstruction, C(z) − jD(z)
Q(z)
fs / 3M
4fs / 3M
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−4
−3
−2
−1
0
1
2
3
4
frequency (sample freq / 8)

(d)