The Asymptotics of Filter and Wavelet Coefficients

Matthew Stone under the guidance of Mladen Victor Wickerhauser

March 28, 2019

Abstract

We present results of Strang and Shen that locate with high precision the zeros of high degree polynomials used in the construction of an important class of orthogonal wavelets known as the Daubechies wavelets. We then give applications to problems in wavelet analysis and present a similar analysis of zeros of the reversed Bessel polynomials used in the construction of filters.

1 A Note on Notation

The fields of filter design and wavelet analysis developed relatively independently and at different periods in time. While they often deal with related questions, the historical development of these fields means that the notation was and is far from standardized. This unfortunate fact means that in order for this paper to be consistent, we have had to make choices with regard to our notation. We have opted to use the mathematician's notation whenever possible (and when it is not possible we will be explicit about the exception).

Additionally, some captions of figures in this paper are superscripted with a dagger †. These figures are still frames from software written for this paper available at https://www.math.wustl.edu/˜victor/stone/. The programs animate the behavior of roots of specific polynomial functions as the degree of the function is increased. More information is available on the webpage and in the freely available code for those interested.

2 Wavelet Basics

We begin by introducing the basics of wavelet analysis, adapted from [4].

We say that the complex-valued function $\psi(x)$ of the real variable x is a *mother wavelet* if and only if the set of *child wavelets*

$$\{\psi_{jk}(x) := 2^{\frac{j}{2}}\psi(2^j x - k) : j, k \in \mathbb{Z}\} \tag{2.0.1}$$

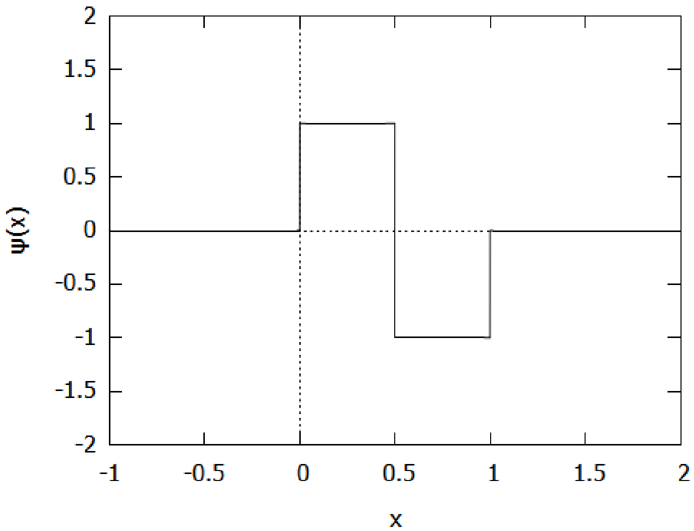

Figure 1: The Haar Wavelet

is an orthonormal basis for $\mathcal{L}^2(\mathbb{R})$ where orthogonality is determined by the $\mathcal{L}^2$ inner product and normality is determined by the $\mathcal{L}^2$ norm

$$< \phi, \psi >= \int_{-\infty}^{\infty} \overline{\phi(x)}\psi(x)dx, \qquad ||\psi|| = \int_{-\infty}^{\infty} |\psi(x)|^2 dx$$

Note that the real requirement is that of orthogonality since the definition of ψ_{jk} includes a normalization factor so that ψ_{jk} has unit norm if ψ has unit norm:

$$||\psi_{jk}|| = \int_{-\infty}^{\infty} |2^{\frac{j}{2}}\psi(2^j(x-k))|^2 dx = (2^{\frac{j}{2}})^2(\frac{1}{2^j}) = 1$$

Note also that a basis determined from a mother wavelet has nested support for fixed k. In particular, fixing $k = n$, we have that $\forall j$ supp$(\psi_{jn}) \subset$ supp$(\psi_{(j+1)n})$.

One simple example of a mother wavelet is the Haar wavelet. The Haar wavelet, displayed in Figure 1, is defined as follows:

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

The Haar wavelet clearly has unit norm, and therefore so does each $\psi_{jk}(x)$. Additionally, each $\psi_{jk}(x)$ is orthogonal to each $\psi_{j'k'}(x)$ for $jk \neq j'k'$, since the child wavelet with smaller support, say $\psi_{jk}(x)$, has its support within an interval where $\psi_{j'k'}(x)$ is constant or is disjoint from the support of $\psi_{j'k'}(x)$ entirely. Finally, Haar showed in his original paper about this system of functions that it indeed forms a basis for $\mathcal{L}^2(\mathbb{R})$ [11].

2.1 Multiresolution Analysis and the Associated Conditions on Wavelet Coefficients

In general, the construction of a mother wavelet is done not directly as in the above example but by using a framework known as multiresolution analysis to define another function ϕ called the scaling function. A *multiresolution analysis* of $\mathcal{L}^2(\mathbb{R})$ is an ascending filtration $\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \ldots$ of subspaces V_j of $\mathcal{L}^2(\mathbb{R})$ such that

- $V_n \to \mathcal{L}^2(\mathbb{R})$ as $n \to \infty$
- $V_n \to \{\emptyset\}$ as $n \to -\infty$
- $f(x) \in V_n \iff f(2^{-n}x) \in V_0$
- $\exists\ \phi(x) \in \mathcal{L}^2(\mathbb{R})$ such that $\{\phi(x-k) : k \in \mathbb{Z}\}$ is a Hilbert basis for $V_0 = \overline{\text{span}}\{\phi(x-k) : k \in \mathbb{Z}\}$

A multiresolution analysis us gives a way to look at $\mathcal{L}^2(\mathbb{R})$ with higher and higher resolution as we increase j, increasing the size of the space V_j we are considering. The wavelets then lie in the 'spaces between resolutions', that is, each $\psi_{jk} \in W_j$ where $V_{j+1} = V_j \oplus W_j$.

The main advantage of this approach is that finding the scaling function is much easier than finding a wavelet directly. One way, for example, to find such a scaling function is to solve a fixed-point problem. In particular, since $\phi \in V_0$ and $V_0 \subset V_1$, $\phi \in V_1$ and so $\{\phi(2x-k)\}$ is a basis for V_1. Therefore,

$$\phi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \phi(2x-k) \tag{2.1.1}$$

This is the *two-scale equation*, and we call the function h that applies the expansion in h_k a *low pass filter*. Since h is a contraction mapping (technically a non-expansion, but we will forget about this point), Banach's Fixed Point Theorem guarantees such a ϕ and gives us a way to find it [13].

We then define our wavelet as

$$\psi(x) = \sqrt{2}\sum_{k} (-1)^k \overline{h_{1-k}} \phi(2x-k) \tag{2.1.2}$$

In a similar fashion, letting $g_k := (-1)^k \overline{h_{1-k}}$ allows us define to the function g that applies the expansion in g_k a *high pass filter*.

We can recast our earlier example of the Haar wavelet in this new language. Choosing $\phi = \mathbf{1}_{[0,1]}$, we see that our filter coefficients $h_0 = h_1 = \frac{1}{\sqrt{2}}$. After setting $g_k = (-1)^k \overline{h_{1-k}}$ as above, we have recovered $\psi(x) = \mathbf{1}_{[0,\frac{1}{2})} - \mathbf{1}_{[\frac{1}{2},1)}$, the Haar wavelet.

Additionally, the added machinery gives us a set of necessary conditions on a scaling function. Suppose that our scaling function ϕ is both integrable and square-integrable with non-zero integral over $\mathbb{R}$. Then by the two scale equation

$$\int_{\mathbb{R}} \phi(x)dx = \int_{\mathbb{R}} \sum_k h_k \sqrt{2}\phi(2x-k)dx = \sum_k h_k \int_{\mathbb{R}} \sqrt{2}\phi(2x-k)dx = \frac{1}{\sqrt{2}} \sum_k h_k \int_{\mathbb{R}} \phi(x)dx$$

and therefore

$$\sum_k h_k = \sqrt{2} \tag{2.1.3}$$

Additionally, since $\{\sqrt{2}\phi(2x-k) : k \in \mathbb{Z}\}$ is an orthonormal basis for $\mathcal{L}^2(\mathbb{R})$, we have that

$$1 =< \phi(t), \phi(t) >= \sum_i \sum_j \overline{h(i)} h(j) \delta(j-i) = \sum_{k \in \mathbb{Z}} |h_k|^2 \tag{2.1.4}$$

where $\delta(n)$ is the Kronecker delta. Finally, the filter must be of even length, that is, for some $n \in \mathbb{Z}$

$$\inf_N \{N : h_k = 0 \ \forall k \text{ outside a contiguous interval of length N}) = 2n \tag{2.1.5}$$

If there were instead a filter of length $N = 2n+1$, the orthogonality condition would stipulate that

$$0 = \sum_k \overline{h_k} h_{2n+k} = \overline{h_0} h_{2n}$$

implying that h_0 or h_{2n} is zero, contradicting that the filter is of length N.

3 Construction of the Daubechies Wavelets

With this framework in place, we can consider more interesting wavelets than Haar's. One class of wavelets of great practical and historical importance are the Daubechies wavelets, designed by Ingrid Daubechies and detailed in [1].

The construction of wavelets and filters is first and foremost a design problem: it is necessary to define what properties are desirable. Once this is defined, we can proceed to derive those wavelets and filters that have these properties. For Ingrid Daubechies, a "good" orthogonal wavelet had to be (1) compactly-supported on the real line and (2) have many vanishing moments.

We say a wavelet ψ is said to have p vanishing moments if and only if $\int_{-\infty}^{\infty} t^k \psi(t)dt = 0 \ \forall k$ such that $0 \leq k < p$. The number of vanishing moments, then, in some sense estimates how quickly the function decays. A wavelet having p vanishing moments can be shown to imply that a linear combination of the scaling function is able to represent polynomials of degree less than p. Roughly speaking, we can think of this condition saying that more vanishing moments means that a function can be represented by a sparser set of wavelet coefficients, which is computationally advantageous.

The design criteria give a set of constraints that allow for a class of compactly-supported, orthogonal wavelets that have a maximal number of vanishing moments to be constructed. These are

1. To ensure compact support for a wavelet ψ, we choose a compactly-supported scaling function as this choice implies that the wavelet itself will be compactly-supported by 2.1.2.
2. For compactly-supported scaling functions, we have that $m_0(x) = \frac{1}{\sqrt{2}} \sum_n h_n e^{-inx}$ is a trigonometric polynomial.
3. The orthogonality constraint implies that $|m_0(x)|^2 + |m_0(x+\pi)|^2 = 1$ by splitting up $\sum_n |h_n|$ into even and odd parts and using the normality condition in 2.1.4.
4. The vanishing moments constraint give us that $m_0(x) = \frac{1+e^{ix}}{2}\mathbf{L}(x)$, where $\mathbf{L}(x)$ is a trigonometric polynomial.

Combining these conditions along with the clever substitution setting

$$|m_0(x)|^2 = (\cos^2(\frac{x}{2}))^N \mathbf{L}(x) := (\cos^2(\frac{x}{2}))^N \mathbf{B}_p(\sin^2(\frac{x}{2})) \tag{3.0.6}$$

gives us that

$$(1-y)^N \mathbf{B}_p(y) + y^N \mathbf{B}_p(1-y) = 1 \quad \forall y \in (0,1]$$

where $\mathbf{B}_p(y)$ is an arbitrary polynomial that will allow us to solve for $|m_0(x)|^2$ and in turn $m_0(x)$ and finally the sequence of low-pass coefficients given by h. To solve for $\mathbf{B}_p(y)$, we can use Bézout's Identity, an old and well-known theorem from number theory. The application of this theorem gives us that

$$\mathbf{B}_p(y) = \sum_{k=0}^{p-1} \binom{p+k-1}{k} y^k \tag{3.0.7}$$

3.0.7 defines the Daubechies polynomial of p terms. To construct the Daubechies wavelets, we need to back out the filter coefficients h_i from the expression for $|m_0(x)|^2$ in 3.0.6. This is done in accordance with the following procedure, adapted from [1] by David Meyer [3], for a wavelet with filter length $2N$:

1. Extract the $N-1$ roots of the polynomial $\mathbf{B}_N(y)$ in 3.0.7. Call these $\{Y_i\}_{i=1}^{N-1}$.
2. Find the solutions $\{Z_j\}_{j=1}^{2(N-1)}$ that satisfy the rule $Z_j + Z_j^{-1} = 2 - 4Y_i$. Note that we get twice as many solutions Z_j as we had zeros Y_i.

3. Choose those $N-1$ $\{Z_j\}$ that lie within the unit disk, and call these $\{r_i\}_{i=1}^{N-1}$.

4. Define the polynomial $\tilde{h}(z) = (z+1)^N(\prod_{i=1}^{N-1}(z-r_i))$ and normalize it such that $\sum_i |h_i| = 2$ by dividing by the constant $2^{N-1/2}\prod_{i=1}^{N-1}(1-r_i)$. This normalized $\tilde{h}(z)$ is $h(z)$.

The advantage of this form will become clear later, but for now it is important to note that the zeros of the Daubechies polynomials figure centrally in the construction of the Daubechies wavelets.

4 The Zeros of the Daubechies Polynomials

We now, as did Strang and Shen in [6], describe the asymptotic behavior of the zeros of the Daubechies polynomials as the degree p increases.

From the coefficients of the Daubechies polynomials alone, we already have information about the location of their zeros. A classical theorem due to Eneström and Kakeya presented in [15] states that any polynomial of degree n with all coefficients a_i real and positive has its zeros Y within the annulus

$$\min_i \frac{a_i}{a_{i+1}} \leq |Y| \leq \max_i \frac{a_i}{a_{i+1}} \quad 0 \leq i \leq n-1$$

Since all coefficients of the Daubechies Polynomials are real and positive, the theorem applies, and $\frac{a_i}{a_{i+1}} = \frac{i+1}{p+i}$ simply by expressing the binomial coefficients in their factorial form. These ratios are monotonically increasing in i, and so the zeros Y must lie in the annulus

$$\frac{1}{p} \leq |Y| \leq \frac{1}{2} \tag{4.0.8}$$

The refinement in [15] gives $|Y| < \frac{1}{2}$ for $p > 2$. Note that the inner radius tends toward 0 as $p \to \infty$, which means that our zeros could be anywhere in the disk $|Y| < \frac{1}{2}$ for filters of long length. This result, however, can be improved.

The key observation of Strang and Shen is that the Daubechies polynomial $\mathbf{B}_p(y)$ coincides exactly with the first p terms in the Maclaurin series for the function $(1-y)^{-p}$. This means that Taylor's formula for the remainder $\mathbf{R}_p(y) = (1-y)^{-p} - \mathbf{B}_p(y)$ can be viewed not as an error estimate for truncations of the series for $(1-y)^{-p}$ but as an asymptotic result about the Daubechies polynomial as p, half the filter length of the Daubechies wavelets, increases. In particular, Taylor's remainder formula gives that

$$\mathbf{R}_p(y) = (2p-1)\binom{2(p-1)}{p-1} y^p \int_0^1 (1-t)^{p-1}(1-yt)^{-2p} dt \tag{4.0.9}$$

but since this expression is monotonic in y for y in the unit disk and all zeros satisfy 4.0.8, our remainder must always be less than $\mathbf{R}_p(\frac{1}{2}) = 2^{p-1}$ for any zero Y. For any zero Y of $\mathbf{B}_p(y)$, then, the remainder term becomes $(1-Y)^{-p}$ and we have that

$$|4Y(1-Y)|^{-p} = |4^{-p}Y^{-p}\mathbf{R}_p(Y)^{-p}| < |4^{-p}\frac{1}{2}^{-p}\mathbf{R}_p(\frac{1}{2})^{-p}| = \frac{1}{2} \qquad (4.0.10)$$

In other words, the curve $|4Y(1-Y)| > 2^{\frac{1}{p}}$ and so the zeros Y of the Daubechies polynomial are bounded outside the limiting curve $|4y(1-y)| = 1$. We summarize the discussion leading up to 4.0.10 in the following theorem.

Theorem 4.1. *The zeros of $\boldsymbol{B}_p(y)$ all satisfy $|4Y(1-Y)| > 2^{1/p}$.*

Strang and Shen go on to prove that the zeros do indeed converge to the limiting curve $|4y(1-y)| = 1$. Figures 2 through 4 plot these zeros for Daubechies polynomials different degree, demonstrating how these zeros converge to this limiting curve. Due to slow convergence about the point $y = \frac{1}{2}$, the proof is broken up to consider separately the region near $y = \frac{1}{2}$ and the region away from $y = \frac{1}{2}$.

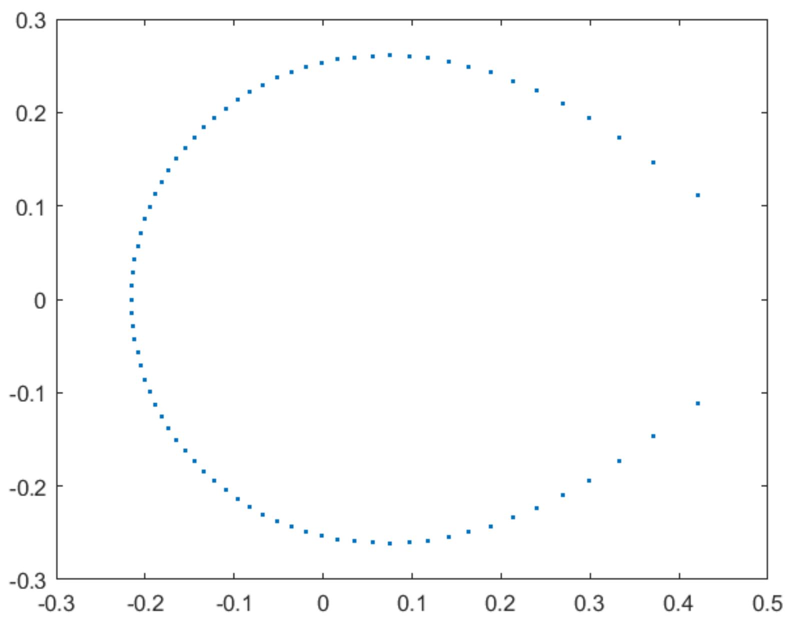

Figure 2: The zeros of $\mathbf{B}_{80}(y)$ †

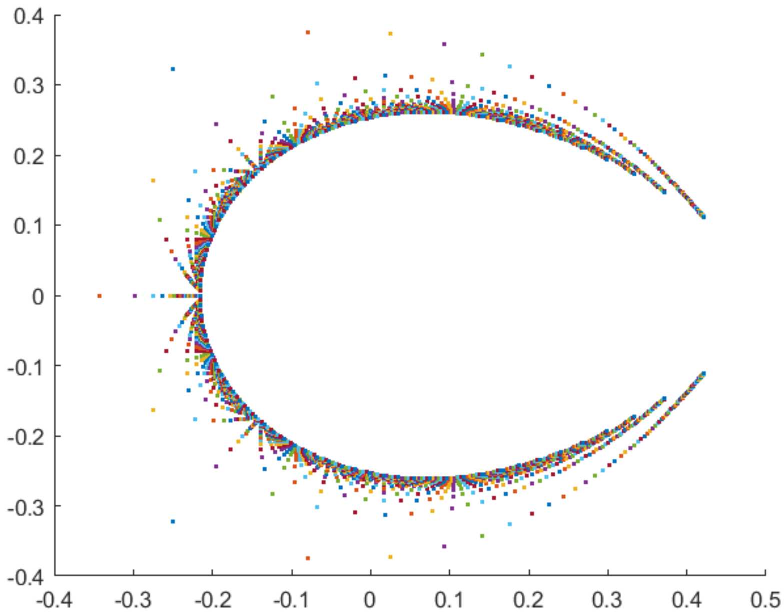

Figure 3: For fixed p, each zero of $\mathbf{B}_p(y)$ is the same color. The points are colored according to the sequence {blue, red, gold, purple, green, cyan, maroon} and p ranges from 3 to 80 †

Theorem 4.2. *The zeros outside the circle $|y - \frac{1}{2}| = \delta$ are not farther than $c(\delta)p^{-2}$ from the curve*

$$|4y(1-y)| = (4\pi p)^{\frac{1}{2p}} |1-2y|^{\frac{1}{p}}$$

for some factor c dependent on δ.

Theorem 4.3. *If W is a zero of the complementary error function erfc(w), there is a zero Y of $\boldsymbol{B}_p(y)$ such that*

$$Y = \frac{1}{2} + \frac{W}{2\sqrt{p}} + \mathcal{O}(p^{-\frac{3}{2}})$$

Both of these results, as did Theorem 4.1, follow from heavy analysis of the remainder term in 4.0.2. Those readers interested in how exactly this core insight of representing the Daubechies polynomial as the truncation of the Maclauren series of $(1-y)^{-p}$ can be used again to yield these two theorems should refer to [6].

Since we now have information about the behavior of the Daubechies polynomials as the number of filter coefficients $2p$ goes to infinity. Since the construction of the Daubechies wavelets is determined by the zeros of the Daubechies polynomials, we might ask what we know about the wavelets of long length.

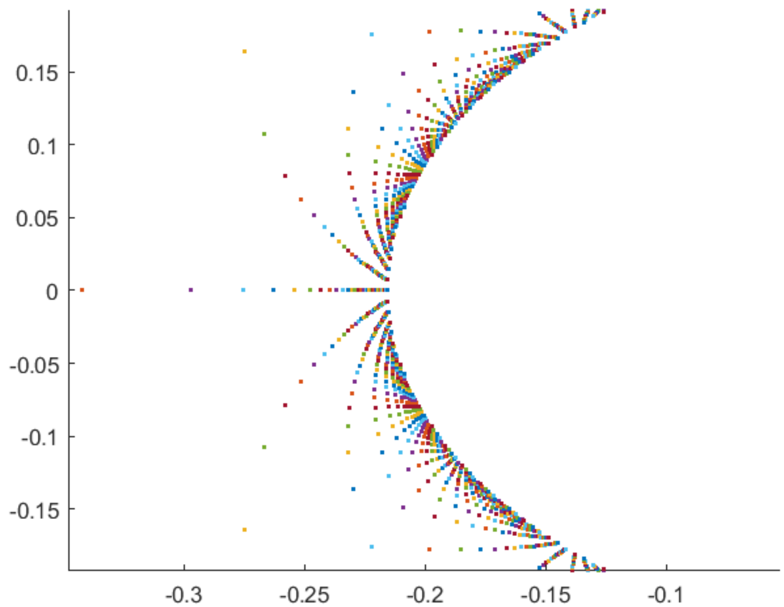

Figure 4: A neighborhood around the value -0.2 with the figure generated as in Figure 3 †

5 An Application of Daubechies Asymptotics: Nearest Neighbor Lifting

This asymptotic result, in fact, has applications to problems in wavelet analysis. In this section, we present one such application. We have already seen how to construct a wavelets in multiple ways: by solving the fixed-point problem in the two-scale equation and by finding necessary and sufficient conditions on specific types of wavelets using multi-resolution analysis as done by Daubechies. Daubechies, together with Wim Sweldens, pioneered yet another way to design wavelets known as the *lifting scheme* [2]. We present the lifting scheme and then show how our asympototic results on the Daubechies polynomials has applications to this method.

5.1 The Lifting Scheme

The approach considers a class of filters with a finite number of non-zero coefficients that satisfy a property known as *perfect reconstruction*. To define this property, we first introduce some useful notation. First, note that for these finitely-supported filters, we may decompose the filter f into its even and odd parts. In particular,

$$f(z) = f_e(z^2) + z^{-1} f_o(z^2)$$

where $f_e(z) = \sum_k f_{2k} z^{-k}$, $f_o(z) = \sum_k f_{2k+1} z^{-k}$. We then assemble the *polyphase matrix*

$$P(z) = \begin{bmatrix} h_e(z) & g_e(z) \\ h_o(z) & g_o(z) \end{bmatrix}$$

where h and g, as before, represent the low and high pass filters, respectively. Now, we define filters h and g as satisfying the *perfect reconstruction* property if and only if

$$P(z)\tilde{P}(z^{-1})^T = \mathbf{I}$$

where $\tilde{P}$ is the dual of P. For orthogonal wavelets, we have that $h = \tilde{h}$ and $g = \tilde{g}$.

Now, if h and g are perfect reconstruction filters,

$$1 = \det(\mathbf{I}) = \det(P(z)\tilde{P}(z^{-1})^T) = \det(P(z))\det(\tilde{P}(z^{-1})^T)$$

and so the determinant of P is a monomial. As such, we can rescale g to ensure P has determinant 1. Filters h and g that give a P with determinant 1 are said to be *complementary* filters that make constitute the *complementary pair* (h, g).

Daubechies and Sweldens then showed the following:

Theorem 5.1. *Let (h, g) be a complementary pair. Then any h^{new} complementary to g is of the form*

$$g^{new}(z) = g(z) + h(z)t(z^2)$$

where $t(z)$ is a Laurent polynomial. Conversely, any filter of this form is complementary to h.

In other words, $P^{new}(z) = P(z) \begin{bmatrix} 1 & 0 \\ t(z) & 1 \end{bmatrix}$ for $t(z)$ as above.

Theorem 5.2. *Let (h, g) be a complementary pair. Then any g^{new} complementary to h is of the form*

$$h^{new}(z) = h(z) + g(z)s(z^2)$$

where $s(z)$ is a Laurent polynomial. Conversely, any filter of this form is complementary to g.

In other words, $P^{new}(z) = P(z) \begin{bmatrix} 1 & s(z) \\ 0 & 1 \end{bmatrix}$ for $s(z)$ as above.

Finally, since Laurent polynomials form a Euclidean domain, the Euclidean algorithm applies to find the gcd of $\begin{bmatrix} h_e(z) \\ h_o(z) \end{bmatrix}$. Since the determinant of P is 1, we have that h_e and h_o are relatively prime, and so this algorithm gives a series of shear matrices whose product is P. Daubechies and Sweldens use this to show their main result:

Theorem 5.3. *Given a complementary pair (h, g), there always exists Laurent polynomials $s_i(z)$ and $t_i(z)$ for $1 \leq i \leq m$ where $m = \frac{n}{2} + 1$ and a non-zero constant K so that*

$$P(z) = \prod_{i=1}^{m} \begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & 1/K \end{bmatrix}$$

We call each shear matrix with shear factor s_i a lifting step and each shear matrix with shear factor t_i a dual lifting step.

Theorem 5.3 gives us the theoretical framework for the wavelet generation procedure known as the *lifting scheme*. We start with $P(z) = \mathbf{I}$, called the *Lazy wavelet* in this context, and then apply a sequence of m lifting and dual lifting steps followed by a scaling. This then produces a complementary filter pair (h, g) completely characterizing a new wavelet.

5.2 Nearest Neighbor Lifting Asymptotics

We have seen that every complementary h and g satisfying the perfect reconstruction property has a lifting factorization. It was shown later shown by Wickerhauser and Zhu that every lifting factorization can be a *nearest neighbor* factorization [14].

A lifting factorization of a complementary pair (h, g) is defined to be *nearest neighbor* whenever

$$s_i(z) = \alpha_i + \beta_i z^{-1}, \quad t_i(z) = \gamma_i z + \delta_i$$

where $s_i(z)$ and $t_i(z)$ are as in Theorem 5.3 and $\alpha_i, \beta_i, \gamma_i, \delta_i \in \mathbb{C}$. Filters with a nearest neighbor factorization only need to access coefficients directly preceding or directly proceeding a given coefficient for each lifting step. This allows for, among other things, the creation of wavelets to be done on a computer system with few distant memory accesses, which, for long filters, lowers the number of cache misses and therefore increases the speed with which the computation can be completed.

While every complementary filter pair admits a nearest neighbor factorization of its polyphase matrix, it often comes at a cost. Reducing the degree of the lifting steps may increase the actual number of lifting steps from the $\frac{n}{2} + 1$ required by Theorem 5.3. Therefore, those of particular interest, at least for the computational advantages discussed earlier, are those nearest neighbor factorizations that do not increase the number of lifting steps required. We call such factorizations *direct nearest neighbor*.

David Meyer investigated whether the Daubechies wavelets had such a direct nearest neighbor factorization in [3], and key to his results was knowledge of the asymptotic behavior of the zeros of the Daubechies polynomials. We now present his results.

Meyer first showed that, for the Daubechies wavelets, if the sequence of divisions on the filter coefficients of h associated with the lifting scheme reduces the degree of the resulting quotient by exactly one after each division, then a direct nearest neighbor factorization of the polyphase matrix exists. Recall that the degree of a Laurent polynomial is not the maximum sum of exponents on each monomial but the smallest integer k such that all non-zero coefficients h_n satisfy $m \leq n \leq m+k$ for some fixed integer m. This means that if we can show the existence of a sequence of divisions that decreases the degree by exactly one after each division, we have shown that the Daubechies wavelets have a direct nearest neighbor factorization.

To do this, we can use Vièta's formulas in tandem with the results of Strang and Shen. Vièta's formulas [12] relate a polynomial's coefficients to products of its roots using the Fundamental Theorem of Algebra. Vièta's formulas are obtained by solving for the coefficients a_k in the expression

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = a_n \prod_{i=1}^{n}(x - r_i)$$

where r_i are the roots of f. Explicitly, Vièta's formulas give that

$$\sigma_k = (-1)^k \frac{a_{n-k}}{a_n} \tag{5.2.1}$$

where σ_k is the sum of all products of exactly k roots r_i and is known as the k^{th} elementary symmetric sum.

Since we have asymptotic information about the roots of the Daubechies polynomials, Vièta's formulas show us that we also have asymptotic information about their coefficients, which we can use to analyze the divisions in the lifting scheme. Note that the construction of the Daubechies low-pass filter presented in section 3 gives only non-negative powers of z and thus Vièta's formulas apply directly. We first need a lemma that reformulates the Daubechies asymptotics convergence results.

Define the set of asymptotic Daubechies polynomial roots in the upper half plane as $A_j^{\Delta} = \{z : |z+1| \leq \sqrt{2}, \arg(Z_{\frac{N}{\Delta_j}}) \leq \arg(z) \leq \arg(Z_{\frac{N}{\Delta_{j+1}}})\}$ where $\Delta_j = 2 + (j-1)\Delta$ and Δ is a small positive real number.

Lemma 5.4. *Let $\{r_i\}_{i=1}^{N-1}$ and $\{Z_i\}_{i=1}^{N-1}$ be the Daubechies polynomial roots and associated asymptotic estimates as in (2) and (3) in Section 3. Then, for any A_j^{Δ},*

$$1 - o(1) < \frac{\#\{r_i \in A_j^{\Delta}\}}{\#\{Z_i \in A_j^{\Delta}\}} < 1 + o(1)$$

Proof. The proof of the lemma follows almost directly from Theorems 4.1 and 4.2 in the Strang and Shen section, the fact that the Z_i are asymptotically distributed evenly along the liming curve, and the definition of A_j^{Δ}. A complete proof can be found in [3]. $\square$

We now can find bounds for the relevant sums of products of roots that appear in Vièta's formulas. We start with σ_1.

Theorem 5.5. *Let $\{r_i\}_{i=1}^{N-1}$ be the roots of the N^{th} Daubechies polynomial as in (2) in Section 3. Then for sufficiently large N,*

$$\sum_{i=1}^{N-1} r_i < 0.36343N$$

Proof. Since $\mathbf{B}_p(y) \in \mathbb{R}[X]$, their $N-1$ roots must come in complex conjugate pairs. Additionally, by the construction of the polynomials, $\mathbf{B}_p(y)$ has at most one real positive root if N is even. For simplicity, we assume N is odd. Ordering the roots increasing by argument and setting $M = \frac{N-1}{2}$, we have that

$$\sum_{i=1}^{N-1} r_i = \sum_{i=1}^{M} (r_i + \bar{r_i}) = 2\sum_{i=1}^{M} \mathrm{Re}(r_i)$$

Instead of summing over the M roots of $\mathbf{B}_p(y)$, we sum over the regions A_j^{Δ} in which they are found. Using Lemma 5.4, we have that

$$\sum_{i=1}^{M} \mathrm{Re}(r_i) \leq \sum_{j=1}^{\infty} \mathrm{Re}(Z_{\frac{N}{\Delta_j}}) \leq \sum_{j=1}^{L} \mathrm{Re}(Z_{\frac{N}{\Delta_j}}) + \frac{N}{\Delta_{L+1}} \mathrm{Re}(Z_{\frac{N}{\Delta_{L+1}}})$$

The bound is tighter as L increases and Δ decreases. Setting $L = 10^6$ and $\Delta = 0.01$, we have that

$$\sum_{i=1}^{N-1} r_i \leq 2[\sum_{j=1}^{L} \mathrm{Re}(Z_{\frac{N}{\Delta_j}}) + \frac{N}{\Delta_{L+1}} \mathrm{Re}(Z_{\frac{N}{\Delta_{L+1}}}] < 0.36343N$$

$\square$

In a similar fashion, first breaking the roots into complex conjugate pairs and then summing over the regions in which the roots can be found and applying the asymptotic results of Strang and Shen, Meyer proves bounds on the elementary symmetric sums σ_2 and σ_3.

Theorem 5.6. *Let $\{r_i\}_{i=1}^{N-1}$ be the roots of the N^{th} Daubechies polynomial as in (2) in Section 3. Then for sufficiently large N,*

$$\sum_{1 \leq i < j \leq N-1} r_i r_j < 0.072753N^2$$
$$\sum_{1 \leq i < j \leq N-1} r_i r_j > 0.063902N^2$$
$$\sum_{1 \leq i < j < k \leq N-1} r_i r_j r_k > 0.04223N^3$$

These together show that the first division step is as required for a direct nearest neighbor factorization, all from the use of the Strang and Shen's asymptotic results and the use of Vièta's formulas.

Theorem 5.7. *Given a degree N Daubechies filter $h = \{h_0, h_1, \ldots, h_{2N-1}\}$, the first division does indeed decrease the degree by exactly one for sufficiently large N.*

Proof. The first division decreases the degree by at least one, and so to show it decreases the degree by exactly one, it suffices to show that the extremal terms of the remainder are non-zero. Formally, we must have that

$$h_{2N-2} - \frac{h_0 h_{2N-1}}{h_1} \neq 0 \tag{5.2.2}$$

$$h_2 - \frac{h_0 h_3}{h_1} \neq 0 \tag{5.2.3}$$

5.2.2 follows directly from inspection of the signs of each term. From the orthogonality condition $h_0 h_{2N-2} + h_1 h_{2N-1} = 0$. Since h_0 and h_1 are always positive, it must be that h_{2N-1} and h_{2N-2} are of opposite signs, proving 5.2.2.

The analysis of 5.2.3, however, is not as simple. Knowing that h_1 is positive, we see that 5.2.3 is equivalent to $h_2 h_1 - h_0 h_3 \neq 0$. We now use Vièta's formula (see 5.2.1) letting $\{R_i\}$ denote the set of roots $\{r_i\}$ along with the N roots at $z = -1$.

$$h_1 h_2 = [(-1)(h_0) \sum_{i=1}^{2N-1} R_i][h_0 \sum_{1 \leq i < j \leq 2p-1} R_i R_j]$$

and

$$h_0 h_3 = (h_0)[(-1)h_0 \sum_{1 \leq i < j < k \leq 2N-1} R_i R_j R_k]$$

Therefore, it suffices to show that

$$\sum_{1 \leq i < j < k \leq 2N-1} -R_i R_j R_k < [\sum_{i=1}^{2N-1} -R_i][h_0 \sum_{1 \leq i < j \leq 2N-1} R_i R_j]$$

This follows from the inequality in Theorem 5.5 and the system of inequalities in Theorem 5.6.

□

The result is true for sufficiently large N and were verified by Meyer for N up to 220.

6 Bessel Polynomials and the Bessel Filter

A natural question would be to ask what other wavelets have such convergence properties. Unfortunately, few other wavelets are derived from polynomials, and even fewer are truncations of a Taylor series. It turns out, however, that the field of filter design is filled with such examples. Additionally, the poles and zeros of rational functions play an important role in the physical properties of the system, and so the study of the asymptotic behavior of roots and poles could be of practical importance too. We present one such asymptotic analysis for the Bessel polynomials.

The Bessel polynomials are a sequence of orthogonal polynomials defined by

$$y_n(x) = \sum_{j=0}^{n} \frac{(n+j)!}{(n-j)!j!}(\frac{x}{2})^j \tag{6.0.4}$$

The engineering literature prefers the so-called reverse Bessel polynomials, which are defined by replacing each variable x^j with the variable x^{n-j} in the Bessel polynomial, that is,

$$\theta_n(x) = x^n y_n(\frac{1}{x}) \tag{6.0.5}$$

These definitions are from [9].

These reversed polynomials are used to construct the Bessel filter, whose transfer function is $T_n(s) = \frac{\theta_n(0)}{\theta_n(s)}$ [7]. In the same way that we analyzed the zeros Daubechies polynomials used in the construction of the Daubechies wavelets, we analyze the zeros of the reversed Bessel polynomials used in the construction of the transfer function of the Bessel filter. Note that the zeros of the reversed Bessel polynomials correspond to poles of the transfer function.

Asymptotic bounds for the Bessel polynomials can be found via an approach similar to that of Strang and Shen's for determining the zeros of the Daubechies polynomials. Where the analysis of the Daubechies polynomials relied on the fact that the coefficients of the Daubechies polynomials of degree p were identical to the truncated Maclauren series of $(1-y)^{-p}$, the coefficients of the Bessel polynomials can also in some sense be thought of as truncations of a Maclauren series.

6.1 Basics of Padé Approximation

We give a basic overview of Padè approximation adapted from [5] that will be important for the asymptotic results presented later.

The Maclauren series of an analytic function f is the limit as n tends to infinity of the sequence of polynomials $P_n(x)$, where $P_n(x)$ is a polynomial of degree n that agrees with $f(x)$ and its first n derivatives at $x = 0$. Each $P_n(x)$ then constitutes an approximation to f by a polynomial of degree n. We can generalize this formulation in a few ways, one of which is to consider not only polynomial functions $P_n(x)$ but rational functions $R_{n,m}(x) = \frac{P_n(x)}{Q_m(x)}$, where $Q_m(x)$ is a polynomial of degree m. Without loss of generality, we may consider that the constant term in one of the polynomials, say Q_m, to be one, as we may divide any the top and bottom by a fixed constant. Note that Padé approximation gives a superset of the possible approximations allowed by the Maclauren series approach as setting $m = 0$ reduces to this earlier case. Indeed, for fixed degree the Padé approximation of a function gives more agreement with derivatives than the Maclauren expansion and can in some cases converge where the Maclauren series will not.

We calculate a Padé approximation for the analytic function $f(x) = \sum_{i=0}^{\infty} a_i x^i$by finding the coefficients p_j and q_k of P_n and Q_m respectively. In particular, we would like to find P_n and Q_m such that

$$f(x) = \frac{P_n(x)}{Q_m(x)} + \sum_{t=n+m+1}^{\infty} c_t x^t$$

or, in other words, to find p_j and q_k such that

$$(\sum_{i=0}^{\infty} a_i x^i)(\sum_{j=0}^{\infty} q_j x^j) - (\sum_{k=0}^{\infty} p_k x^k) = \sum_{t=n+m+1}^{\infty} c_t x^t$$

Since c_t is zero for all $0 \leq t \leq n+m$ the expression on the left hand side in each power x^n to 0 gives us a system of $n+m+1$ equations in $n+m+1$ unknowns that allows for us to solve for the coefficients p_j and q_k:

We have the following system of $n+1$ equations

$$\begin{aligned} a_0 - p_0 &= 0 \\ q_1 a_0 + a_1 - p_1 &= 0 \\ &\vdots \\ q_M a_{n-m} + q_{m-1} a_{n-m+1} + \cdots + a_n - p_n &= 0 \end{aligned}$$

and the following system of m equations

$$\begin{aligned} q_m a_{n-m+1} + q_{m-1} a_{n-m+2} + \cdots + q_1 a_n + a_{n+1} &= 0 \\ q_m a_{n-m+2} + q_{m-1} a_{n-m+3} + \cdots + q_1 a_{n+1} + a_{n+2} &= 0 \\ &\vdots \\ q_m a_n + q_{m-1} a_{n+1} + \cdots + q_1 a_{n+m-1} + a_{n+m} &= 0. \end{aligned}$$

Solving this system of linear equations then gives us the coefficients for a Padé approximation to the function f.

For example, the Padé approximate $R_{2,2}$ to e^x can by this method be shown to be $\frac{1+x/2+x/12}{1-x/2+x/12}$. When $n = m$ as above, we call $R_{n,n}$ the diagonal approximant R_n. In general, the diagonal approximant of e^z can be shown [8] to have numerator

$$P_n(z) = \sum_{i=0}^{n} \frac{(2n-i)!n!z^i}{2n!i!(n-i)!} \tag{6.1.1}$$

and denominator

$$Q_n(z) = \sum_{i=0}^{n} \frac{(2n-i)!n!(-z)^i}{2n!i!(n-i)!} \tag{6.1.2}$$

6.2 The Poles of the Bessel Filter Transfer Function

The numerator in this approximation bears close resemblance to the Pade numerator of e^z. In fact, we can directly verify from 6.0.4 and 6.1.1 that the Bessel polynomial is related to the Pade numerator $P_n(\frac{2}{z})$ by the equation

$$y_n(z) = \frac{2n!}{n!}(\frac{z}{2})^n P_n(\frac{2}{z}) \tag{6.2.1}$$

It was realized by de Bruin, Saff, and Varga [9] that this could be leveraged to give asymptotic results about the zeros of Bessel polynomials.

Theorem 6.1. *$\hat{z}$ is a limit point of zeros of $\theta_n(nz)$ if and only if*

$$\hat{z} \in D_1 := \{z \in -i\overline{\mathbb{H}} : |\frac{ze^{\sqrt{1+z^2}}}{1+\sqrt{1+z^2}}| = 1, |z| \le 1\}$$

Notice that the polynomial θ_n in the above theorem has a normalization factor of n in its argument that allows for this sequence to converge. Additionally, while the above theorem is applicable to only $\theta_n(nz)$, inversion of $z \leftarrow \frac{1}{z}$ allows us to get asymptotic results about the zeros of the normalized Bessel polynomials.

The proof relies on a theorem of Saff and Varga [8] that describes the location of the zeros of normalized Padé approximates of e^z, which we state below as another theorem. The theorem relies on steepest descent methods outside the scope of this paper, and so we treat this theorem as a black box.

Theorem 6.2. *Suppose $\{R_{n_j,m_j}\}_{j=1}^{\infty}$ is a sequence of Padé approximants to e^z such that $\lim_{j\to\infty}(n_j) = \infty$ and $\lim_{j\to\infty}(\frac{m_j}{n_j}) = \sigma$. Let*

$S_\sigma := \{z : |arg(z)| > cos^{-1}(\frac{1-\sigma}{1+\sigma})\}$

$g_\sigma(z) = \sqrt{1 + z^2 - 2z\frac{1-\sigma}{1+\sigma}}$

$w_\sigma(z) = \frac{4\sigma^{\frac{\sigma}{\sigma+1}} ze^{g_\sigma(z)}}{(1+\sigma)(1+z+g_\sigma(z))^{\frac{2}{1+\sigma}}(1-z+g_\sigma(z))^{\frac{2\sigma}{1+\sigma}}}$

Then $\hat{z}$ is a limit point of zeros of the normalized Padé approximants $\{R_{n_j,m_j}((n_j+m_j)z)\}_{j=1}^{\infty}$ if and only if

$$\hat{z} \in D_\sigma := \{z \in \overline{S_\sigma} : |w_\sigma(z)| = 1, |z| \le 1\}$$

We now provide an alternate proof to that given in [10] to prove Theorem 6.1 for our special case of $n = m$.

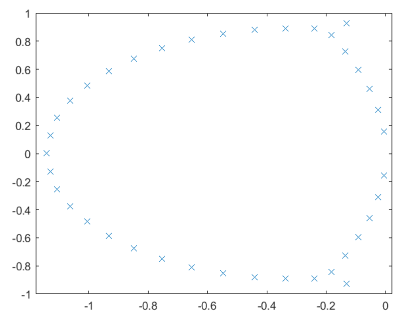

Figure 5: Poles of $T_{80}(y)$ plotted as 'x' as per engineering convention †

Proof. By the definition of $\theta_n(z)$ in 6.0.2 and the equality in 6.2.1, we have that

$$\theta_n(\frac{z}{2}) = \frac{1}{2^n}\frac{(2n)!}{n!}P(z)$$

Using the change of variables $z \leftarrow 2nz$, we have the equality

$$\theta_n(nz) = \frac{1}{2^n}\frac{(2n)!}{n!}P_n(2nz) \tag{6.2.2}$$

recognizing $\theta_n(nz)$ as the normalized reversed Bessel polynomial and $P_n(2nz)$ as the normalized Padé approximant.

Now, since $n = m$, we have that $\lim_{j\to\infty}(n_j) = \infty$ and $\lim_{j\to\infty}(\frac{m_j}{n_j}) = 1$. Applying Theorem 6.2, all zeros of $P_n(2nz)$ lie on the curve D_1. Once we substitute $\sigma = 1$ in Theorem 6.2, we see that D_1 is

$$\{z \in -i\overline{\mathbb{H}} : |\frac{ze^{\sqrt{1+z^2}}}{1+\sqrt{1+z^2}}| = 1, |z| \leq 1\}$$

Finally, since each fixed n is constant, 6.2.2 shows that the zeros of $\theta_n(nz)$ must also lie on D_1, proving the theorem. □

This theorem is substantially more general in de Bruin, Saff, and Varga and is subsequently sharpened in [9] and [10], but for our purposes, such a convergence result is enough. This convergence result shows, in a similar vein to the results of Strang and Shen, that recognizing a particular function as the truncation of another suggests not only a result about approximation error but an asymptotic one too.

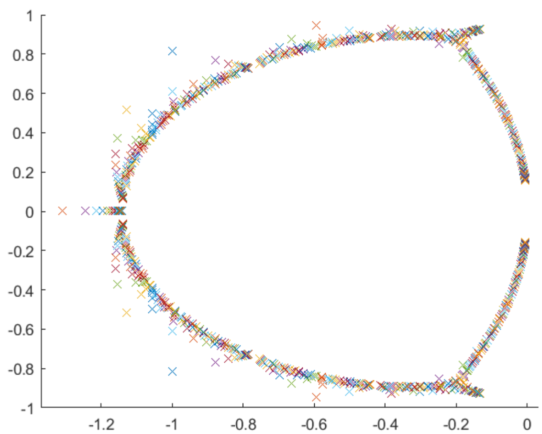

Figure 6: For fixed n, each zero of $T_n(y)$ is the same color. The points are colored according to the sequence {blue, red, gold, purple, green, cyan, maroon} and p ranges from 3 to 40 †

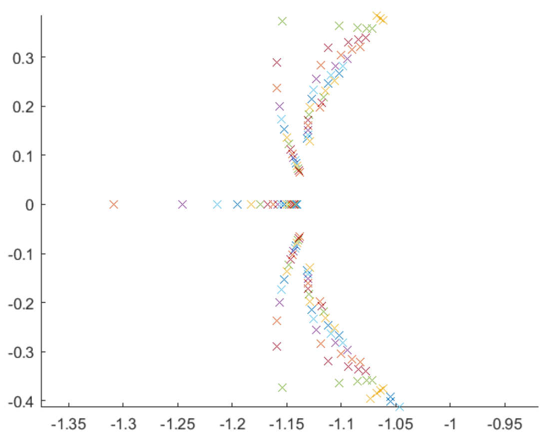

Figure 7: A neighborhood around the value -1.1 with the figure generated as in Figure 6 †

7 Acknowledgements

I would like to thank my advisor, Professor Victor Wickerhauser, for his guidance over the past few months, my family and friends for their encouragement and support, and the Freiwald Summer Scholars program for funding the research around which this paper has been based.

References

[1] Daubechies, I. (1992). Ten lectures on wavelets. Montpelier, Vermont: Capital City Press.

[2] Daubechies, I., & Sweldens, W. (1998). Factoring wavelet transforms into lifting steps. *The Journal of Fourier Analysis and Applications, 4*(3), 247-269.

[3] Meyer, D. (2015). Wavelet factorization and related polynomials. *Arts & Sciences Electronic Theses and Dissertations*. 492.

[4] Wickerhauser, M. V. (2010). *Mathematics for multimedia*. Boston, Massachusetts: Birkhäuser.

[5] Mathews, J. H., & Fink, K. D. (1999). *Numerical methods using MATLAB* (3rd ed.).

[6] Shen, J., & Strang, G. (1996). The Zeros Of The Daubechies Polynomials. *Proceedings of the American Mathematical Society, 124*(12), 3819-3833.

[7] Storch, L. (1954). Synthesis of constant-time-delay ladder networks using bessel polynomials. *Proceedings of the IRE, 42*(11), 1666-1675.

[8] Saff, E. B., & Varga, R. S. (1978). On the zeros and poles of Padé approximants. III. *Numerische Mathematik, 30*, 241-266.

[9] de Bruin, M. G., Saff, E. B., & Varga, R. S. (1981). On the zeros of generalized bessel polynomials. I. *Indagationes Mathematicae (Proceedings) 84*(1), 1-13.

[10] de Bruin, M. G., Saff, E. B., & Varga, R. S. (1981). On the zeros of generalized bessel polynomials. II. *Indagationes Mathematicae (Proceedings) 84*(1), 14-25.

[11] Haar, A. (1910). Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen, 69*, 331-371.

[12] van der Waerden, B. L. (1993). *Algebra, Vol. 1.* New York: Springer-Verlag.

[13] Thomson, B. S., Bruckner, J. B., & Bruckner, A. M. (2001). *Elementary Real Analysis*. Upper Saddle River, New Jersey: Prentice-Hall.

[14] Zhu, W., & Wickerhauser, M. V. (2013). Wavelet Transforms by Nearest Neighbor Lifting. Excursions in Harmonic Analysis, 2, 173-192.

[15] Marden, M. (1966). Geometry of Polynomials. *Mathematical Surveys (3)* 179.