Asymptotics of Wavelets and Filters

by

Jianhong (Jackie) Shen

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1998

Author ....
Department of Mathematics
May 1, 1998

Certified by ....
Gilbert Strang
Professor of Mathematics
Thesis Supervisor

Accepted by ....
Hung Cheng
Chairman, Applied Mathematics Committee

Accepted by ....
Richard B. Melrose
Chairman, Department Committee on Graduate Students

Asymptotics of Wavelets and Filters

by

Jianhong (Jackie) Shen

Submitted to the Department of Mathematics
on May 1, 1998, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In Wavelet Theory, the most significant (both historically and in the present stage) family of wavelets is the Daubechies orthogonal wavelets with compact supports. The most powerful source of curiosity and imagination is the key equation–the Refinement Equation (or Dilation Equation). And finally, it is the keyword "filter" that has served as the major bridge connecting mathematicians, physicists, and engineers, and made Wavelets Theory one of the very few examples in this century that was rooted in several different fields, and was someday unified in the paradise of mathematics (Harmonic Analysis and Approximation Theory), and eventually found itself in the broad market of engineering fields (especially in the information processing technology). This thesis is a mixture of my research results on analyzing, generalizing, and developing Daubechies family of orthogonal wavelets, the Refinement Equation, and the design of digital filters. The main tool is asymptotic analysis.

To study Daubechies' family of wavelets, we first study the associated Daubechies lowpass filters (or polynomials). The distribution of zeros is closely studied and its asymptotic pattern is obtained. The transition bandwidth of the filter is found to be proportional to the inverse of the square root of the number of zeros at the "highest" frequency $\omega = \pi$ (a key number also determining the smoothness of the wavelets). This first step breaks the secret of the nonlinear phase information of the filters. The leading linear term approximation of the phase makes it possible to carry out asymptotic analysis on Daubechies wavelets and scaling functions. The energy significant parts of the wavelets and scaling functions are determined by the stationary phase method.

Our study of the Refinement Differential Equations (RDE) was the first one in the literature. It is motivated by a perturbation of the Refinement Equation and the search for wavelets-like functions. We establish the connection between wavelets and solutions to certain types of functional differential equations, a hot topic near 1970s. We reveal the general structure of solutions to RDE's and discover that RDE's are naturally connected to a class of Refinement Functional Equations (RFE). The Continuous Subdivision Algorithm finds its dominant place in solving RDE's. The vague probability idea of Rvachev (1971) is developed more completely.

In spite of the simple formulation of various (weighted or unweighted, real or complex domains, simply connected or multiply connected domains) Chebyshev (L^∞) polynomial approximation problems, the analytic behavior of solutions remains a mystery except for some simple cases, due to the lack of geometric structures. This is the underlying reason why engineers working on filter designs are frequently puzzled by certain behaviors of *optimal filters*. Based on Fuchs' work, we improve and interpret a widely-used empirical formula established by Kaiser based on his numerical data in 1972. Our new asymptotic formula proves to be more accurate than Kaiser's. To compute the critical constant (related to the Green's function of the underlying domain), we study properties of the Green's function for a multi-interval domain as well as its equilibrium distribution measure. This result also contributes to the numerical analysis of partial differential equations (the Stokes equation in fluid dynamics, for example).

Thesis Supervisor: Gilbert Strang
Title: Professor of Mathematics

Credits

Most of the material in this thesis has appeared or will appear in some journals, or has been submitted. The work on Daubechies filters, scaling functions, and wavelets in Chapter 2 is joint work with Gilbert Strang. The analysis and computation of the zero distribution appeared in [65, 1996], and the asymptotic analysis of the scaling functions and wavelets will appear in [66, 1998]. I would like to thank Cleve Moler (MathWorks Inc.) for providing his numerical observations. The material in Chapter 3 has been submitted [64, 1997] and presented in the 9th International Conference of Approximation Theory. I would like to thank Gilbert Strang and Dingxuan Zhou (Hong Kong City University) for introducing me Rvachev's work in 1970s and providing useful references. The material in Chapter 4 is partly joint work with Gilbert Strang and will appear in [67, 1998]. The newest parts [68, 1998] are now developing from a conformal mapping idea of Nick Trefethen (Oxford Computing Lab). I would like to thank Alan Oppenheim (MIT EECS) and Jim Kaiser (Bell Lab) for discussing REAL problems in the design of digital filters.

I would also like to thank the many people who have made useful suggestions to me in my research. Among them, I would especially mention Tomas Arias, Dmitri Betaneli, Hung Cheng, Ingrid Daubechies, Alan Edelman, Ross Lippert, Truong Ngyuen, Alan Oppenheim, Gilbert Strang, Vasily Strela, Nick Trefethen and Andy Wathen.

Acknowledgement

I would like to express first my abstract but deep appreciation: to the ancient Chinese philosophy and tons of wisdom stories in her history – making me aware of the importance of "being balanced" in every aspect of my life; to the great man **Deng Xiaoping** – Lincoln liberated millions of slaves of USA, while Mr. Deng has set free a billion of *heads* in China (and mine is one epsilon among them); and finally, to the **United States**, for her generosity and kindness to foreigners, and her spirit of being mixed, not to chaos, but to the benefits of all our human beings.

I also feel deeply inside that I owe each person in the department a "thank you", for their friendly helps and daily "hello"s, with which, I found my big family in the USA. Especially, I would like to thank **Linda**, for her candies and cookies in Room 233, where, consciously or not, I loved to drop by, and for her heart-shaped Valentine Day cards to every graduate; and **Shirley, Sueli, and Nini**, with whose "universal" keys, I never worried about being locked out of my office; and **Eda**, for each of her heart-warming emails reminding our proctoring assignments; and **Tivon**– my personal calls made in the headquarter could never be free :-(, yet personal faxes for me always reached me so timely :-) .

I am also grateful to those in the department who have taught me mathematics and academic skills: to **Greenspan and Malkus**, for introducing me to the world of fluid mechanics and solar dynamos; to **Edelman**, for the influence on me of his love of numerical linear algebra (and his one summer support, which was so important to me as a foreign student); to **Stroock**, for teaching me the beauty of theoretical probability and Martingale theory; to **Toomre** – being his TA on numerical analysis was the most unforgettable experience in the past four years. I owe special thanks to **Hung Cheng**, with whom I can speak Chinese and express more personal feelings about the new life here, not mentioning his teaching me on integral equations and asymptotics of ODE's. To complete my moral payment, I would like to thank **Gian-Carlo Rota**, from whom I learned the beautiful classical theory of polynomials, umbral calculus, exterior algebra, Clifford algebra, and invariant theory, as well as the "stories" or "philosophy" behind them; and from whom I learned how to be persistent in research, and how to create something from nothing.

I owe a particular debt of gratitude to all my former 2-342 officemates and friends: **Dave,**

Betaneli, Peter, Mats, Lior, Mathew, Radica and Lisa. To me, they were the most helpful dictionaries when I could not figure out an English word or wanted to know a certain side of the western life; the most enjoyable group of people talking with either about a small funny thing I discovered or a homework problem.

I would like to thank all my Chinese friends in the department: those still here or having graduated. Besides mathematics and speaking English, speaking Chinese was so important a part of my daily life. When words are merely symbols, feelings can be evaporated. But when words are feelings, then to speak is to touch, to express, to enjoy, and to generate solutions to every hard problem in life. I don't know how to weigh the importance of a native language and native friends.

I would also like to thank especially my chinese friends in other departments, who have had great influence on my life. Among them, I would like to mention **Chuan He, Wen Zhang, Qiang Zhu, and Shanhui Fan.** We approximate each other to a very high precision in many sides of our personality and academic styles. The lunch table was a constant source of jokes and knowledge. I cannot imagine a life without them. I really felt it a privilege talking about the protein folding problem when our girlfriends were hunting in Macy or Sears for their suits. And my words feel too powerless to take a photo of my happy time, sitting together with them on the lawn of Boston Botany Garden in the fresh spring time, playing cards from dawn to sunset.

I have benefited greatly from my practicing of teaching for more than two years at the **Tutoring Room Service**, and the **Experimental Study Group** at MIT. Special thanks must also go to the **Wellesley College** for offering me one semester of teaching job. Also allow me to thank the **Graduate Student Council** of MIT for supporting me in the 1998 AMS annual meeting.

Finally, I feel I can never pay back to my advisor **Gilbert Strang** and my closest friend **Tianxi Cai**. They have defined my well-balanced and happy academic life and "plain" life, on this land far away from my family and homeland. let me dedicate to them the following famous poem by Bo Wang (Chinese, 650–676) (hope my translation is an isomorphism, of both words and thoughts):

Where there are seas and lands,
there are those knowing you,
as deep as yourself to you...
there are those with you,
as friendly as your neighbors to you...

In the memory of my grandma.

Contents

Chapter 1

On Asymptotics

1.1 What Is Asymptotics

There are many textbooks and conference procedings on asymptotic analysis and its applications. But none of them has given an overall synthesis on this subject, or has updated the content of asymptotics. In this first chapter, I give my attempt. Though asymptotic analysis is only one tool in my thesis (not the subject), I still feel it valuable to freshen and broaden our viewpoints on this old but never dormant subject, because the way we view things, determines the way we act and justify our actions. Also the discussion of general asymptotics may provide some important background for this thesis.

Almost all asymptotic analysis textbooks consist of three major parts: how to sum infinite series, how to evaluate integrals with large parameters (Laplace or Fourier types), and how to solve differential equations with a small or large parameter (second order differential equations typically). They are three major columns for the hall of classical asymptotic analysis. But to me, asymptotic analysis has already been scattered in several fields: classical analysis, probability, combinatorics, dynamic systems, and so on. It depends on our understanding of the meaning of "asymptotics", and in the following I have chosen bravely the widest (and therefore maybe wildest) one.

Using the least number of words, asymptotics means trends.

What studied by asymptotic analysis is *a system of objects*. It can be a family of integrals or differential equations, or a sequence of polynomials, or a dynamic process (such as matrix iterations in numerical linear algebra and iterations of maps in a dynamic system), or a collection of random variables. In any case, the target objects must be connected by at least *one parameter*, which can be either the real time (as in differentiable dynamic systems), or a discrete "time"(as in various

iterations), or a crucial system parameter (such as the number of vanishing moments for Daubechies family of wavelets with compact supports, the size of a gap between two close intervals, or a dimensionless physical constant in differential equations). The central task of asymptotic analysis is to detect and classify trends of the systems as the parameters vary (especially toward some extremal values), to predict the "speed"(temporal) or the "scale"(spatial) of the trends and to give simple but practically useful approximations to the trends. Therefore asymptotic method is one among a handful of powerful "applied" methods.

1.2 Examples of Asymptotics

An ancient Chinese poet wrote: "you cannot see the real face of Mount LuShan [1], only because you are on it." The same applies when we observe a system of objects. You cannot feel the trend of a system unless you allow the system parameter to vary in a very large range ("watch it from a distance"). The behavior of any individual object is often hard to understand, not transparent to analysis, and even unpredictable — because its behavior is the mixed effect of many factors, and many relations that can be random or deterministic, and linear or nonlinear. Only in the asymptotic case, one can consider only very few dominant factors or relations. This can make things much simpler than usual.

Let us look at several examples scattered in different contexts.

Riemann-Lebesgue Lemma The first simple example is the Riemann-Lebesgue Lemma in analysis. Let $f(x)$ be any L^1 integrable function on $[a, b]$ (either a or b can be ∞). Then

$$\lim_{\lambda \to \infty} \int_a^b e^{i\lambda x} f(x)\, dx = 0.$$

For each individual λ, the integral obviously depends on $f(x)$, a and b, and its exact evaluation can be very hard (except by numerical methods). The Lemma captures such a simple asymptotic behavior universally shared by this (Fourier) type of integral. It provides a simple necessary condition for a function to be the Fourier or Laplace transform of an L^1 function. In the Riemann-Lebesgue Lemma, the asymptotic trend is the cancellation. The parameter λ is the frequency of canceling periods.

Law of Large Numbers and Central Limit Theorem The second familiar example comes from probability. For one ideally random toss of a coin, the result of being head or tail is unpredictable.

[1] LuShan is one of the most beautiful mountains in China.

However, after tossing it for many times, "almost surely", for nearly half of the times we must get heads and for the other half, tails. This half to half behavior is an asymptotic one and the parameter is the number of tosses. Generally, this example is summarized by the celebrated Law of Large Numbers, and the Central Limit Theorem describes even more detailed asymptotic behavior. These two fundamental theorems of probability are both asymptotic results. They describe the universal asymptotic behavior shared by a fairly large class of random events. In this example, asymptotics is the sibling of another familiar word: statistics, and basically, the trend is still cancellation or averaging—an individual random variable X is usually complicated, yet asymptotically, there is a lot of cancellation in the independent sum $(X_1 + X_2 + \cdots + X_N)/N$.

Attractor, Ergodicity and Ergodic Theorem The third example is from dynamic systems. In a differentiable dynamic system (on a compact manifold, say), the evolution of an individual state or phase often allows a very wide degree of freedoms and therefore usually has no simple closed form. Fortunately, asymptotically, or as the counting time goes to infinity, the trajectory must exhibit certain universal behaviors such as being attracted by an attractor (or a "sink"), which can be either an attracting state, or a stable limiting circle (mostly in the plane phase case), or even a strange attractor (in a high dimensional phase space). Each flow can be complicated, yet its asymptotic behavior can be simply identified and classified. Another asymptotic example in dynamic system is the set of concepts like "ergodic", "mixing", and "exact". Each of them describes one typical sort of asymptotic behavior of the semigroup generated from the iteration of a given map. The celebrated Birkhoff's Ergodic Theorem is yet another asymptotic example in dynamic system and it is more or less related to the Central Limit Theorem, whose asymptotic meaning is just discussed above.

Regular and Singular Perturbation The fourth example, which is more classical, is the perturbation method of linear or non-linear differential equations. Even for second order linear ordinary equations, except for some simple valuable cases such as Cauchy equations, equations with constant coefficients, and equations with analytic coefficients, there is no closed form for the solutions. Fortunately, in application, the dimensionless equation obtained from a physical system often contains a large or small parameter. This usually makes the problem much simpler since the leading terms of the solution can be easily obtained by regular or singular perturbations (though the problem of matching is usually non-trivial). For example, the (leading term) solution to the following equation containing a small parameter

$$\epsilon y'' - x^2 y' - y = 0, \quad y(0) = y(1) = 1$$

can be easily found by solving one first order outer (or slowly varying) problem and two second order inner (or boundary layer, or rapidly varying) problems. In this example, the asymptotic means the trend of spatial variation: as ϵ gets smaller and smaller, the region with rapid spatial variation tends to be more and more concentrated near one of the boundary points—the famous phenomenon of boundary layer.

Polynomial Sequences The last example concerns polynomial sequences. Let us consider two classes of polynomial sequences that are closely related to Chapter 2 and Chapter 4 in this thesis. The first class is the partial sum sequence of the power series (at some point) of a meromorphic function. For example,

$$q_n(z) = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^n}{n!}$$

for e^z at $z = 0$. A question asked and answered by Szegö is the zero distribution pattern of $q_n(z)$. It is hard to describe precisely the zeros of $q_n(z)$ for each individual n (except for $n = 1, 2, 3, 4$). However, asymptotically, the zeros of $q_n(z)$ behave very regularly—after a simple elementary transform, the zeros are nearly equidistributed along the unit circle. And this asymptotic pattern is universally shared by this class of polynomial sequences. The second class of polynomial sequences are Chebyshev polynomials for a domain and a given function. That is, given a function $f(z)$ and a domain K in the complex plane, $p_n(z)$ minimizes the error

$$\|f - p_n\|_{L^\infty(K)}$$

among all polynomials of degree n. The behavior of $p_n(z)$ obviously depends on $f(z)$, which can be very complicated and arbitrary: entire, meromorphic, analytic, smooth, or only continuous. Besides, the domain can also have a large degree of freedom. However, as the approximation order gets larger, the sequence always exhibits certain common asymptotic behavior. In Chapter 4, we study the asymptotic behavior of a polynomial approximation problem from digital filter design.

Summary These examples are scattered in different fields and have never been seen via a unified viewpoint. The purpose of of the listing is to extract something common hidden in them and therefore to find a quasi-foundation and methodology for them.

1.3 Mechanisms for Asymptotics

If we are the loyal followers of the cause-effect philosophy, there must exist some common causes leading to asymptotics or trends. These causes, or "forces" as I would like to call in the following, in my opinion, include the following three major ones: cancellation, averaging, and attraction.

Cancellation The cancellation mechanism sees itself in the Riemann-Lebesgue Lemma, the method of Stationary Phase, and even in the Law of Large Numbers. We can understand it in the following quasi-philosophical way. Inside each object of a given system, there is more than one force (typically two, Yin and Yang, for example). Those forces are not balanced in an individual object because of random factors, and thus make the individual objects varying and complicated. Those objects are linearly ordered according to a certain system parameter (frequency, say), which more or less characterizes the cancellation degree of those forces. As the parameter increases, the cancellation gets stronger and certain steady (or stationary) trends can appear. This is the asymptotics arising from cancellation.

Averaging The averaging mechanism to asymptotics appears in the Law of Large Number, the Central Limit Theorem, and the Birkhoff Ergodic Theorem. It is more or less associated to statistics. The asymptotic or trend in this case is obtained from averaging a large sample of objects. The individual irregularities cancel out each other during the averaging process. In this case, the objects of the system must carry certain degree of randomness or diversity. For example, the iterations of an ergodic map must be able to send any non-zero mass almost everywhere. In signal processing, averaging means the elimination (or filtering) of high frequencies. Asymptotic trend is usually steady and stationary, and therefore corresponds to low frequencies. They are preserved and even amplified during the averaging (lowpass filtering) process.

Attraction Attraction is probably the most common way leading to asymptotics or trends. Unlike cancellation, whose mechanism depends on two or more *internal* forces, and averaging, which requires certain degree of *randomness* from the system, attraction is caused by certain deterministic "external forces." The simplest example is the iterations of an initial state under a contracting map. The "external force" is the contracting mechanism, which for a linear system, is usually caused by the spectral radius of a linear operator (less than 1). The "external force" is also the unique fixed point (suppose the metric space is complete) in the sense that the trajectory of any initial state is attracted to it. Sinks, limiting circles, and strange attractors are more examples of "external forces." [2]

[2] Of course, being external and internal is relative. An external force for one object (say, the trajectory of a state)

Let us enumerate more examples. Consider the classical one:

$$\int_0^1 e^{-\lambda x} f(x)\, dx,$$

where λ is a positive parameter. Suppose $f(0)$ is not zero (and assume f is smooth for simplicity). Then the leading term as λ gets very large is simply $f(0)/\lambda$. It is obtained by replacing $f(x)$ by $f(0)$ in the integral. As λ gets larger, the effect of $x = 0$ becomes more and more dominant. It acts like a strong force pulling the whole weight of integration round it.

Also consider the boundary layer phenomenon. In this case, fluid mechanics leads to a more vivid picture of the "external force"–the drag of a plate or some boundary material to the liquid. As viscosity gets smaller (or the Reynolds number gets larger), the propagation of this influence through shear stress of the liquid is more confined near the boundary and we observe thinner boundary layers.

Finally, let us look at the asymptotics of the zeros of the partial sum polynomial sequence of a meromorphic function. The scaled zeros will converge to a limiting curve (see Chapter 2, for example), which acts as an attracting force. In fact, the underlying mechanism for this attracting force is exactly the same as that discussed in the second paragraph.

Summary The understanding of asymptotic mechanisms helps us to adopt appropriate methodology in applications. For attraction, the main task of asymptotic analysis is to identify the "dominant force" and find a suitable approach to amplify its influence. For averaging, it is often inevitable to turn to the methodology of statistics and operator theory. For cancellation, it is crucial to locate the states where cancellation is the least (since they will determine the leading terms).

1.4 Introduction to the Thesis and Its Asymptotic Contents

Chapter 2 studies the Daubechies miniphase orthogonal wavelets with compact support (for spline wavelets, the work has been carried out by other people). The major difficulty of analyzing this family is caused by the complicated non-linear phases of the associated filters. The wavelets also have the same property—with simple magnitudes but very complex phases (in the Fourier domain). From a certain angle, this family of wavelets is very alike the Airy function, whose Fourier transform is also purely phased. Our analysis starts with the asymptotic pattern of the zeros of the filters and ends at the asymptotic structure of the wavelets. Basically, the asymptotics of a polynomial sequence (derived from truncating the power series of a family of meromorphic functions) and the method of stationary phase are used in this chapter.

can be the internal force for a larger system (the whole dynamic system).

In Chapter 3, we study a generalization of the Refinement Equation. A refinement equation has the following form

$$\phi(x) = h[0]\phi(2x) + h[1]\phi(2x-1) + \cdots + h[N]\phi(2x-N)$$

for some real coefficients (or filter coefficients) $h[0], \cdots, h[N]$. The Refinement Equation plays a crucial role in the theory of wavelets with compact support. It also links Wavelet Theory to signal processing and other fields. After perturbing this equation, we obtain a new class of equations called Refinement Differential Equations. Our major achievement in this chapter is the discovery of the link of Wavelet Theory to the theory of functional differential equations and probability theory. This chapter contains the least content of asymptotics, however.

Chapter 4 studies a particular polynomial approximation problem arising from digital filter designs, and also the associated potential theory for a several-interval domain. The asymptotic behavior of the optimal polynomial sequence is often determined by the singular locations of the target function (i.e. poles) and the critical points of the Green's function for the working domain (the "external forces" mentioned in the preceding section). The first part impoves an empirical formula discovered by Kaiser regarding the relation of optimal errors to filter lengths. The second part studies the Green's function and equilibrium distribution of a several-interval domain based on the Schwarz–Christoffel mapping. Both structural and asymptotic results are established.

Chapter 2

Asymptotics of Daubechies Mini-phase Filters and Wavelets

Though it now has been a cliché — "to analyze wavelets, first analyze the filters," it never hurts in practice to follow this simple principle.

The first part of the chapter studies the asymptotic behavior of Daubechies filters (polynomials). The zero distribution pattern of the filters is crucial in their filtering effects as well as in the next stage of analysis (on wavelets). Here the objects are discrete (polynomials and their zeros), yet the result is continuous (the existence of the limiting curve). The second part studies the asymptotics of Daubechies scaling functions and wavelets based upon their Fourier integrals. The stationary phase plays an important role here. In this part, the objects are continuous (integrals and wavelets), yet the result is in certain sense discrete (three different scales with separate asymptotics).

2.1 The Zero Distribution of Daubechies Filters

2.1.1 Introduction

The Product Filter $P(z)$: Positivity and Zeros

Let $H(z) = \sum_{n=0}^{N} h[n]z^{-n}$ be a lowpass filter whose associated Refinement Equation

$$\phi(t) = \sum_{n=0}^{N} h[n]\phi(2t-n)$$

yields orthogonal integer translates $\{\phi(t-n) \mid n \in \mathbb{Z}\}$. Its associated "energy" filter, or the product filter $P(z)$ is defined by $\boxed{P(z) = H(z) \cdot H(z^{-1})}$. In order that the integer translates of the scaling function are orthogonal to each other, it is necessary for $H(z)$ to be a *quadrature mirror filter* (QMF), a connection first made by Mallat. This means

$$P(z) + P(-z) = 1. \tag{2.1}$$

Such a filter is called a *halfband* filter. Its impulse response $h[n]$ is always zero at even times $n = 2k$ except when $n = 0$.

The product filter has the following two remarkable properties:

(1) Positivity: $P(z) \geq 0$, for all $|z| = 1$.

Suppose $z = e^{i\omega}$. Then $H(z^{-1}) = \overline{H(z)}$ and $P(z) = |H(z)|^2 \geq 0$. Combined with the symmetry property $P(z) = P(z^{-1})$, we conclude that $P(z)$ must be a nonnegative polynomial of $x = \cos\omega$:

$$p(x) = \sum_{n=0}^{L} c[n]x^n, \qquad p(x) = P\Big(\frac{z+z^{-1}}{2}\Big).$$

(2) Zeros at $z = -1$.

Since $H(z)$ is a lowpass filter, we always impose the lowpass condition: $H(1) = 1$. Then $P(1) = 1$ and Eq.(2.1) implies: $P(-1) = 0$. Therefore $P(z)$ must have zero(s) at $z = -1$, or the highest (digital) frequency $\omega = \pi$. This, in return, implies $H(-1) = 0$.

There is a profound influence of those zeros at $z = -1$ in Wavelet Theory. As shown in Battle [3, 1989], Mayer [50, 1992], Daubechies [10, 1992], and their most recent improvement in Cai and Shen [7, 1998], those zeros are necessary to achieve good smoothness for the wavelets. If an orthogonal wavelet is C^m, then at least m zeros of $H(z)$ should be guaranteed at $z = -1$ (even more in practice).

Whereas image processing engineers debate the real necessity of smoothness in their applications, mathematicians feel no hesitation to have it – for the purpose of regularity analysis of functions and having "good" basis functions for the Wavelet-Galerkin method (for solving PDE's numerically).

A general design problem of orthogonal wavelets starts with the design of $P(z)$. The number of zeros at $z = -1$ must be an even number, say $2p$. It is therefore convenient to factorize it in the following form:

$$P(z) = \left(\frac{1+z^{-1}}{2}\right)^p \left(\frac{1+z}{2}\right)^p Q(z). \tag{2.2}$$

Obviously $Q(z)$ must also be nonnegative on the unit circle and have a symmetric impulse response. It is the $Q(z)$ part that has induced the diversity of orthogonal wavelets with compact supports.

Ingrid Daubechies chose $Q(z)$ in a typical mathematician's way: $Q(z)$ is extremal in certain sense.

Daubechies' Maxflat Condition

The Maxflat Condition asks for the *lowest* order of $Q(z)$ such that $P(z)$ defined by Eq.(2.2) is both a halfband filter and nonnegative when restricted on the unit circle.

It is convenient to introduce another variable y: $y = (1-x)/2$. Here $x = (z+z^{-1})/2$ is the Joukowski transform ($x = \cos\omega$ when z is restricted on the unit circle). Since $Q(z)$ is symmetric, it must be a polynomial of x, and therefore of y. Denote it by $\mathbf{B}(y)$. Then $p(y) = (1-y)^p\mathbf{B}(y)$, if $p(y)$ denotes the "y–transform" of $P(z)$. The halfband condition Eq.(2.1) now becomes

$$(1-y)^p\mathbf{B}(y) + y^p\mathbf{B}(1-y) = 1. \tag{2.3}$$

Notice that $y \in [0,1]$ as z changes on the unit circle. Therefore $\mod y^p$,

$$(1-y)^p\mathbf{B}(y) \equiv 1,$$

or

$$\mathbf{B}(y) \equiv \frac{1}{(1-y)^p} = 1 + py + \binom{p+1}{2}y^2 + \binom{p+2}{3}y^3 + \cdots.$$

Define $\mathbf{B}_p(y)$ to be the following polynomial of degree $p-1$:

$$1 + py + \binom{p+1}{2}y^2 + \cdots + \binom{2p-2}{p-1}y^{p-1}. \tag{2.4}$$

Then $\mathbf{B}(y) \equiv \mathbf{B}_p(y), \mod y^p$. This means the lowest order of $\mathbf{B}(y)$ can be $p-1$. And it is not difficult to see that $p(y) = (1-y)^p \mathbf{B}_p(y)$ is the unique Hermitian interpolation polynomial of degree $2p-1$ that interpolates 1 at $y = 0$ and 0 at $y = 1$ both to order p. By symmetry, $p(y)$ must satisfy the halfband condition Eq.(2.3). Therefore $\mathbf{B}_p(y)$ is indeed the polynomial with the lowest degree. Daubechies made this choice. Various Daubechies families of wavelets have been designed from it. The difference lies in the procedure of factorizing $P(z)$ into a product of $H(z) \cdot H(z^{-1})$ (the so called *spectral factorization*). In this thesis, we only demonstrate the most natural factorization: the **mini-phase spectral factorization**, which leads to the mini-phase orthogonal wavelets.

To start, suppose we have already known the $p-1$ roots $Y_1, Y_2, \cdots, Y_{p-1}$ of $\mathbf{B}_p(y)$. By the rule of Joukowski transform $z + z^{-1}/2 = 1 - 2y(= x)$, in the z-plane, we have $2p-2$ preimages of those Y_i's–exactly half of which lie inside the unit circle. Denote them by $Z_1, Z_2, \cdots, Z_{p-1}$. Then the Daubechies mini-phase filter is defined by

$$H_p(z) = \Big(\frac{1+z^{-1}}{2}\Big)^p \prod_{n=1}^{p-1} \frac{1 - z^{-1} Z_n}{1 - Z_n}. \tag{2.5}$$

If the product factor is omitted, the Refinement Equation produces spline functions–with accuracy p but not orthogonal to their integer translates.

Our main goal is to analyze the zero distribution pattern of $H_p(z)$.

2.1.2 A Note about the Numerical Computation of Zeros

Before we set out applying many analytic methods, let us first mention briefly the numerical computation of the zeros using Matlab, a popular software for signal processing and wavelet analysis.

Matlab creates the companion matrix whose characteristic polynomial is $\mathbf{B}_p(y)$. Then it finds the eigenvalues of that matrix. Without scaling, this breaks down at $p = 35$, because of the wide range in the coefficients of $\mathbf{B}_p(y)$. The first coefficient is 1, and by Stirling's formula, the coefficient of y^{p-1} is

$$\binom{2p-2}{p-1} \simeq \frac{\sqrt{2\pi(2p-2)}}{2\pi(p-1)} \frac{(2p-2)^{2p-2}}{(p-1)^{2p-2}} = \frac{4^{p-1}}{\sqrt{\pi(p-1)}} \tag{2.6}$$

The leading term 4^{p-1} suggests that the variable $4y$ is preferable to y. With this scaling, the Matlab computation remains accurate to $p = 80$. For larger p, a bifurcation (see Figure 2-1) occurs from roundoff error. The coefficient $\binom{p-1+i}{i} 4^{-i}$ of $(4y)^i$ is numbered $b(p-i)$ by Matlab. Then $b(p) = 1$

and the sequence of coefficients is created recursively;

$$\text{for } i = p-1:-1:1,\ b(i) = b(i+1) * (2p-i-1)/(4*(p-i)); \text{ end}$$

The command "$Y = \text{roots}\,(b)/4$" produces the approximate zeros $Y(1), \ldots, Y(p-1)$.

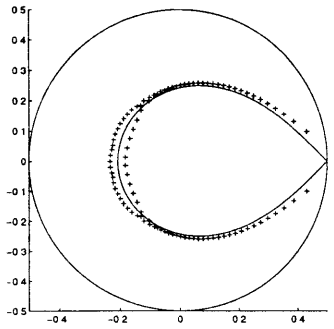

Figure 2-1: A bifurcation occurs from roundoff error, $p = 100$.

Experiments with other root-finding algorithms were less successful, even though working with the companion matrix is *a priori* surprising. A polynomial with repeated roots leads to a defective matrix (not diagonalizable). Algorithms based on Newton's method had difficulty with the accurate evaluation of $\mathbf{B}_p(y)$ and $\mathbf{B}'_p(y)$. Lang's algorithm (Lang and Frenzel [45, 1994]) is comparable to Matlab 'roots', and probably faster.

2.1.3 A Clue from the Three-term Recursive Relation

In this section, we attempt to have a clue of the zero distribution pattern of $\mathbf{B}_p(y)$ from its recursion formula.

From Eq.(2.4),

$$\mathbf{B}_p = 1 + py + \binom{p+1}{2} y^2 + \cdots + \binom{2p-2}{p-1} y^{p-1},$$
$$y\mathbf{B}_p = y + py^2 + \cdots + \binom{2p-3}{p-2} y^{p-1} + \binom{2p-2}{p-1} y^p.$$

Hence

$$(1-y)\mathbf{B}_p = \mathbf{B}_{p-1} + \binom{2p-3}{p-2} y^{p-1}(1-2y).$$

Since

$$\binom{2p-1}{p-1} = \binom{2p-3}{p-2}\frac{(2p-2)(2p-1)}{(p-1)p} = 2(2-p^{-1})\binom{2p-3}{p-2},$$

by setting $c_p = 2(2-p^{-1})$, we have

$$(1-y)\mathbf{B}_{p+1} - [1 + c_p y(1-y)]\mathbf{B}_p + c_p y\mathbf{B}_{p-1} = 0. \tag{2.7}$$

Note that $c_p \to 4$ as $p \to \infty$.

Lemma 1 *Suppose a sequence of meromorphic functions $f_p(y)$, $p = 0, 1, \cdots$ satisfy the following three-term recursive relation:*

$$(1-y)f_{p+1}(y) - [1 + 4y(1-y)]f_p(y) + 4yf_{p-1}(y) = 0.$$

Then **generically** *as $p \to \infty$, zeros of $f_p(y)$ in the holed complex y-plane $\mathbb{C}\backslash\{0, 1, \infty\}$ approach the following lemniscate (see Figure 2-2):*

$$|4y(1-y)| = 1.$$

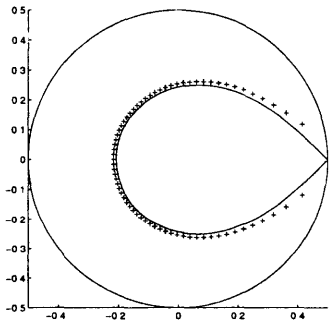

Figure 2-2: Zeros of $\mathbf{B}_p(y)$ approach the lemniscate: $|4y(1-y)| = 1$.

Proof. The auxiliary equation (AE) for the recursion formula is

$$(1-y)\lambda^2 - [1 + 4y(1-y)]\lambda + 4y = 0.$$

Since y changes, λ is a function of y. It has two auxiliary roots:

$$\lambda_1 = \frac{1}{1-y}, \qquad \lambda_2 = 4y.$$

Therefore

$$f_p(y) = A(y)\lambda_1^p(y) - B(y)\lambda_2^p(y)$$

for some suitable coefficients A and B (both depending on y). In fact,

$$A = \frac{f_1 - f_0\lambda_2}{\lambda_1 - \lambda_2}, \qquad B = \frac{f_1 - f_0\lambda_1}{\lambda_1 - \lambda_2}.$$

By "generic", we mean both $A(y)$ and $B(y)$ are not zero functions. Since f_0 and f_1 are meromorphic functions, so are A and B. On the zero set of $f_p(y)$, we have

$$\left[\frac{\lambda_2}{\lambda_1}\right]^p = \frac{A}{B},$$

or

$$|4y(1-y)| = |\lambda_2/\lambda_1| = |A/B|^{1/p}.$$

If $A/B = 0$ or ∞, $\lambda_2/\lambda_1 = 4y(1-y) = 0$ or ∞. This is only possible when $y = 0$, 1, or ∞. On the rest of the y-plane, $|A/B|$ is finitely positive. Hence $|\lambda_2/\lambda_1|$ approaches 1 as $p \to \infty$. □

This lemma gives us a clue of how the zeros of $\mathbf{B}_p(y)$ might behave in the complex plane. However, since c_p is not exactly 4, we cannot apply it directly to $\mathbf{B}_p(y)$. It seems that we have to analyze Eq.(2.7) through perturbing the three-term relation in the lemma. The analysis then gets very involved. We therefore abandon this method and turn to a more analytical and easier method first used by Gabor Szegö.

Let me mention that there is no reason to curse the non-zero $4 - c_p = 2p^{-1}$. It is this small deviation that makes the sequence $\mathbf{B}_p(y)$ a *polynomial* sequence. Generally $f_p(y)$ can be at most a sequence of meromorphic functions.

2.1.4 Two Bounds for the Zeros of $\mathbf{B}_p(y)$

In this section, we prove two bounds for the zeros. Let Y denote an arbitrary zero of $\mathbf{B}_p(y)$, and Z any preimage of $1 - 2Y$ under Joukowski transform.

Theorem 1 *For $p = 2$, the only zero is $Y = -1/2$. For $p > 2$, all the zeros satisfy $|Y| < 1/2$. Especially, in the z-plane,* $\mathrm{Re}(Z) > 0$.

Its proof depends on a result due to Eneström and Kakeya (Marden [49, 1966]).

Lemma 2 (Eneström and Kakeya) *Let $p(y)$ be a polynomial of degree n with all coefficients a_i real and positive. Define $r_i = a_i/a_{i+1}$, $0 \le i \le n-1$. Then all zeros of $p(y)$ must lie in the closed annulus:*

$$\min_i r_i \le |y| \le \max_i r_i.$$

The details about when and how the zeros can indeed lie on the border of the annulus is discussed by Anderson, Saff and Varga [2, 1979].

Proof of Theorem 1. Obviously, all coefficients of $\mathbf{B}_p(y)$ are real and positive. $r_i = (i+1)/(p+i)$ for $0 \le i \le p-2$. Thus $\min r_i = r_0 = 1/p$, and $\max r_i = r_{p-2} = 1/2$. By the lemma and its sharpened form, $|Y| < 1/2$ for $p > 2$. Therefore, $\mathrm{signRe}Z = \mathrm{signRe}(1-2Y) = 1$. □

See Figure 2-3 to visualize this result.

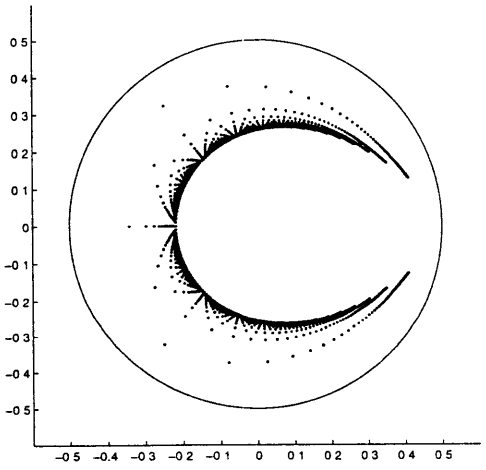

Figure 2-3: All zeros lie inside the circle $|y| = 1/2$, $p = 3:1:60$.

Our next bound is more tricky and the underlying idea is borrowed from Szegö [71, 1924].

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

Here again we see the quadratic polynomial $4Y(1-Y)$ appear (see also the preceding subsection). This time, we give a direct analytic proof.

Proof of Theorem 2 . $\mathbf{B}_p(y)$ is the truncated Taylor series at $y = 0$ for $(1-y)^{-p}$. The pth derivative of this function is $p(p+1)\dots(2p-1)(1-y)^{-2p}$. Then Taylor's integral formula for the remainder $\mathbf{R}_p(y) = (1-y)^{-p} - \mathbf{B}_p(y)$ is

$$\begin{aligned}\mathbf{R}_p(y) &= (2p-1)\binom{2(p-1)}{p-1}\int_0^y (y-s)^{p-1}(1-s)^{-2p}\,ds \\ &= (2p-1)\binom{2(p-1)}{p-1}y^p\int_0^1 (1-t)^{p-1}(1-yt)^{-2p}\,dt\end{aligned}$$

Call this last integral $\mathbf{I}_p(y)$. Since each zero has $|Y| < 1/2$, for any $t \in (0\,,\,1]$, $|1-Yt|^{-1} < (1-t/2)^{-1}$. Hence

$$|\mathbf{I}_p(Y)| < \int_0^1 (1-t)^{p-1}(1-t/2)^{-2p}\,dt = \mathbf{I}_p(1/2).$$

At $y = 1/2$, Eq.(2.3) gives $\mathbf{B}_p(1/2) = 2^{p-1}$. Thus the remainder is

$$\mathbf{R}_p(1/2) = (1-1/2)^{-p} - 2^{p-1} = 2^{p-1}.$$

At each zero of $\mathbf{B}_p(y)$, $\mathbf{R}_p(Y) = (1-Y)^{-p}$. The above equations combine into

$$|4Y(1-Y)|^{-p} = |4^{-p}Y^{-p}\mathbf{R}_p(Y)| < |4^{-p}(1/2)^{-p}\mathbf{R}_p(1/2)| = 1/2.$$

This is the bound $|4Y(1-Y)| > 2^{1/p}$ that puts Y outside the limiting curve, and completes the proof. (Also see Figure 2-4.) □

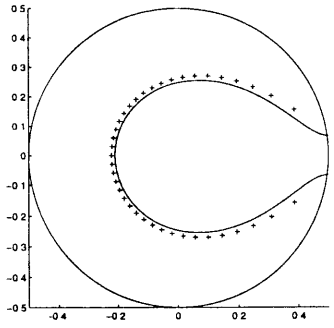

Figure 2-4: All zeros lie outside the curve $|4y(1-y)| = 2^{1/p}$, $p = 40$.

This idea of using the integral remainder belongs to Szegö. In [71, 1924], he studied the asymptotic zero distribution pattern of the partial sum (polynomial) sequence $q_n(z)$, obtained from the

infinite series expansion of e^z:

$$q_n(z) = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^n}{n!}.$$

Recent extension of this work can be found in Varga [75, 1992].

2.1.5 Regular Zeros and Singular Zeros

Further analysis exhibits that the zeros of $\mathbf{B}_p(y)$ should be better grouped into two sets: those away from $y = 1/2$ and those near $y = 1/2$. For convenience, we call them the "regular" zeros and the "singular" zeros.

Regular Zeros

Lemma 3 *Fix a $\delta > 0$. Then uniformly for all $|y| \le 1/2$ and $|y - 1/2| \ge \delta$,*

$$\mathbf{I}_p(y) = \frac{1}{p(1-2y)} + O(p^{-2}).$$

Proof. In the integral $\mathbf{I}_p(y)$, change variables from t to $w = (1-t)/(1-yt)^2$. Then w goes from 1 to 0 and the derivative is $dw/dt = (2y - yt - 1)/(1-yt)^3$. We leave part of the integral in terms of t

$$\mathbf{I}_p(y) = -\int_0^1 w^{p-1}\Big(\frac{1-yt}{2y-1-yt}\Big)dw.$$

As $p \to \infty$ the power w^{p-1} is concentrated near $w = 1$. Around that endpoint the leading term of the expression in parentheses is $(2y-1)^{-1}$. The integration of w^{p-1} gives $1/p$ and completes the proof. □

If Y is a zero of $\mathbf{B}_p(y)$, then $\mathbf{R}_p(Y) = (1-Y)^{-p}$. Therefore

$$\begin{aligned}
[4Y(1-Y)]^{-p} &= 4^{-p}(2p-1)\binom{2p-2}{p-1}\mathbf{I}_p(Y) \\
&= 4^{-p}\binom{2p-2}{p-1}\frac{2}{1-2Y}(1+O(p^{-1})) \\
&= \frac{1}{(1-2Y)\sqrt{4\pi p}}(1+O(p^{-1})).
\end{aligned} \tag{2.8}$$

We have applied Eq.(2.6) in the last step. The pth root displays the equation of the approximate curve C_p and the error term

$$|4Y(1-Y)| = |1-2Y|^{1/p}\,(4\pi p)^{1/2p}\,(1+O(p^{-2})). \tag{2.9}$$

Theorem 3 *Let $\delta > 0$ be any fixed small positive number. Then all zeros outside the circle $|y - 1/2| = \delta$ are not farther than Ap^{-2} from the curve C_p:*

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

The constant A only depends on δ.

Proof. Let y be the point on C_p nearest to Y and $\epsilon = Y - y$. We must show that ϵ is $O(p^{-2})$. Since $|1+\epsilon|^{1/p} = 1 + O(|\epsilon|/p)$, we have

$$\begin{aligned} |1-2Y|^{1/p} &= |1-2y|^{1/p} \cdot \left|1 + \frac{\epsilon}{1-2y}\right|^{1/p} \\ &= |1-2y|^{1/p} \cdot (1 + O(|\epsilon|/p)) \\ |4Y(1-Y)| &= |4y(1-y)| \cdot \left|1 + \frac{1-2y}{y(1-y)} \cdot \epsilon + O(\epsilon^2)\right| \\ &= |4y(1-y)| \cdot |1 + E\epsilon + O(\epsilon^2)| \end{aligned}$$

where $E = (1-2y)/(y(1-y))$. $E = O(1)$ since δ is fixed. Division yields

$$\frac{|4Y(1-Y)|}{|1-2Y|^{1/p}(4\pi p)^{1/2p}} = \frac{|1 + E\epsilon + O(\epsilon^2)|}{1 + O(|\epsilon|/p)} = |1 + E\epsilon + o(|\epsilon|)|.$$

On the other hand, by Eq.(2.9), the right hand side of the last equation is $1 + O(p^{-2})$. Therefore the left hand side implies that ϵ must be of order $O(p^{-2})$. □

Corollary 1 *All zeros outside the circle $|y - 1/2| = \delta$ are not farther than Bp^{-1} from the curve D_p drawn in Figure 2-5:*

$$|4y(1-y)| = 1 + \epsilon_p, \qquad \textit{where } \epsilon_p = \frac{\ln(4\pi p)}{2p}.$$

Here B is a constant only dependent on δ.

A further argument directly based on Eq.(2.8) provides a more detailed information about these regular zeros, which is given in our next theorem:

Theorem 4 *Let $u = 4y(1-y)$, and $r_p = 1 + \epsilon_p$ as defined in the corollary above. Then for any*

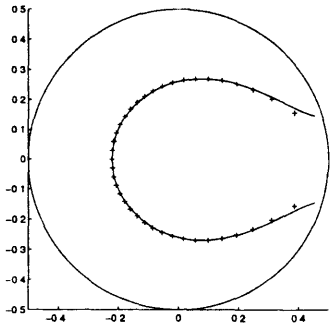

Figure 2-5: D_p is a first order approximation curve for regular zeros. $p = 40$.

fixed (as compared to p) small positive number α,

$$U_k = r_p \exp(2\pi i \frac{k}{p}), \qquad p\alpha \le k \le p(1-\alpha), \quad k \in \mathbb{N},$$
$$Y_k = \frac{1+\sqrt{1-U_k}}{2}, \qquad \textit{(take the negative real part branch of } \sqrt{}\textit{)}$$

gives a first order approximation (i.e. with error of order $O(p^{-1})$*) to the regular zeros lying outside a circle* $|y - 1/2| = \delta(\alpha)$ *for some* $\delta(\alpha) > 0$ *and* $\delta(\alpha) \to 0$ *as* $\alpha \to 0$.

Please note that the theorem says that on the u-plane, the regular zeros are asymptotically equidistributed.

Singular Zeros

The value of $y = 1/2$ is in every respect a singular point for this problem. It corresponds to points $z = i$ and $z = -i$ on the unit circle. We now prove that the zeros Y approach $1/2$ at speed $p^{-1/2}$, as Moler discovered by Matlab experiment. Surprisingly, the coefficient of $p^{-1/2}$ comes from a zero W of the complementary error function

$$\mathrm{erfc}(w) = 1 - \mathrm{erf}(w) = \frac{2}{\sqrt{\pi}} \int_w^\infty e^{-s^2}\, ds.$$

The corollary will improve slightly a known result for the location of these zeros.

Theorem 5 *If* W *is a zero of* $\mathrm{erfc}(w)$, *there is a zero* Y *of* $\mathbf{B}_p(y)$ *and a zero* Z *of* $Q_p(z)$ *such that*

$$Y = \frac{1}{2} + \frac{W}{2\sqrt{p}} + O(p^{-3/2}),$$
$$Z = i - \frac{W}{\sqrt{p}} - \frac{iW^2}{2p} + O(p^{-3/2}).$$

Proof. We introduce a new expression for $p(y) = (1-y)^p \mathbf{B}_p(y)$ (note $p(y) = P(z)$). As a function of y, this is a polynomial of degree $2p-1$ whose derivative has $p-1$ zeros both at $y=0$ and $y=1$. Therefore the derivative is a multiple of $y^{p-1}(1-y)^{p-1}$, and we have an *incomplete beta function*

$$p(y) = (1-y)^p \mathbf{B}_p(y) = 1 - c_p^{-1} 2^{2p-1} \int_0^y t^{p-1}(1-t)^{p-1} dt. \tag{2.10}$$

The number c_p is determined by setting $y=1$:

$$c_p = 2^{2p-1} \int_0^1 t^{p-1}(1-t)^{p-1}\, dt = 2^{2p-1} \frac{\Gamma(p)^2}{\Gamma(2p)} = 2^{2p-1} \left((2p-1) \binom{2p-2}{p-1} \right)^{-1}$$

By Stirling's formula, we have

$$c_p = \sqrt{\frac{\pi}{p}} (1 + O(p^{-1})).$$

By symmetry, the value of the integral above should be $2^{1-2p} c_p/2$. Therefore $P(1/2) = 1/2$. In order to see the detail of the zeros of $\mathbf{B}_p(y)$ near $y = 1/2$, we introduce a new variable by $y - 1/2 = w/2\sqrt{p}$. Then

$$\begin{aligned}
p(y) = p(1/2 + w/2\sqrt{p}) &= p(1/2) - c_p^{-1} 2^{2p-1} \int_0^{w/2\sqrt{p}} (1/2+t)^{p-1}(1/2-t)^{p-1}\, dt \\
&= 1/2 - 2\, c_p^{-1} \int_0^{w/2\sqrt{p}} (1-4t^2)^{p-1}\, dt \\
&= 1/2 - \frac{2\sqrt{p}}{\sqrt{\pi}} \int_0^{w/2\sqrt{p}} e^{-4pt^2}\, dt\, (1 + O(p^{-1})) \\
&= 1/2 - \frac{1}{\sqrt{\pi}} \int_0^{w} e^{-s^2} ds\, (1 + O(p^{-1})) \\
&= 1/2 \mathrm{erfc}(w) + O(p^{-1})
\end{aligned}$$

Let W be a zero of $\mathrm{erfc}(w)$. All zeros are simple, because the derivative e^{-w^2} is never zero. The fundamental theorem of complex analysis says that as $p \to \infty$, $p(1/2 + w/2\sqrt{p})$ is zero at some point $w = W + O(p^{-1})$. In terms of y, $Y = 1/2 + W/2\sqrt{p} + O(p^{-3/2})$. This completes the proof since $\mathbf{B}_p(y)$ shares every zero with $p(y)$ except $y = 1$. □

As an interesting application, we can infer certain behaviors of the zeros of the complementary error function.

Corollary 2 *Every zero of* $\mathrm{erfc}(w)$ *has* $|\arg\ W| < 3\pi/4$.

Proof. The corresponding Y lies outside the limiting curve $|4y(1-y)| = 1$, which intersects itself

at $y = 1/2$ with slopes ± 1. In the limit, $W = (Y - 1/2)/\sqrt{p} + O(p^{-1})$ must have $|\arg\ W| \leq 3\pi/4$. If the equality held, W^2 would be purely imaginary. Then the previous theorem would give

$$|4Y(1-Y)| = |1 - W^2 p^{-1} + O(p^{-2})| = 1 + O(p^{-2})$$

This contradicts the inequality $|4Y(1-Y)| > 2^{1/p}$ in Theorem 2, proving the corollary. □

Fettis, Caslin, and Cramer [26, 1973] computed the zeros of erfc(w) to very high accuracy. They also proved an asymptotic form of the statement $|\arg\ W| \leq 3\pi/4$. It is interesting to see the complete statement (which their numerical table confirms) proved by such an indirect argument involving the zeros of $\mathbf{B}_p(y)$.

These zeros approach $1/2$ at order $p^{-1/2}$, close to the line $Y - 1/2 = W/2\sqrt{p}$. By the corollary, the slope of this line is not ± 1. Therefore the distance from Y_p to the limiting curve C is of strict order $p^{-1/2}$ near $y = 1/2$. In this region, the error order in Eq.(2.9) rises to p^{-1}. This applies in particular to the rightmost zero, which comes from the first W tabulated in [26, 1973], $Y \approx 1/2 + (-1.3548\ldots + i1.9914\ldots)/2\sqrt{p}$.

2.1.6 Transition Bandwidth

It is no surprise to see the connection to the error function. Probability theory has already made the error function a universal cumulative distribution function through the Central Limit Theorem. In analysis, the error-function-like behavior is universally shared by certain integrals with a large parameter. In this subsection, we apply this idea to find the *transition bandwith* of the Daubechies product filter $P(e^{i\omega})$, a quantity very important in signal processing.

A change of variables $t = (1 - \cos\theta)/2$ in Eq.(2.10) produces the integral of $\sin^{2p-1}\theta$. The limits of integration are related by $y = (1 - \cos\theta)/2$. Thus Eq.(2.10) leads to the Meyer's form [50, 1992] of the halfband filter $P(z)$ in Eq.(2.2):

$$P(e^{i\omega}) = 1 - c_p^{-1} \int_0^{\omega} \sin^{2p-1}\theta\, d\theta. \tag{2.11}$$

The zero at $y = 1$ becomes the celebrated "zero at π" for the frequency response $P(e^{i\omega})$. This zero at $\omega = \pi$ is of order $2p$, from the power of $\sin\theta$ in the above integral and the form of $P(z)$ in Eq.(2.2). Factorization gives pth order zeros for the Daubechies polynomials in $P(z) = H(z)H(z^{-1})$. That zero at $\omega = \pi$ and $z = -1$ is responsible for the p vanishing moments in the wavelets.

The trigonometric polynomial $P(e^{i\omega})$ drops monotonically from one to zero on $0 \leq \omega \leq \pi$. The first $2p - 1$ derivatives are zero at $\omega = 0$, and $\omega = \pi$, from the vanishing of $\sin^{2p-1}\theta$. Furthermore,

this integral of $(1-\cos\theta)^{p-1}\sin\theta$ involves only odd powers of $\cos\theta$, and the only even power is the constant term. $P(e^{\imath\omega})$ is odd around its value $1/2$ at $\omega = \pi/2$, and it is called "halfband".

An important question for such a filter is the slope at $\omega = \pi/2$. This slope determines the width of the frequency band, in which P drops from 1 to 0. An ideal filter has a jump; its graph is a brick wall (however, this ideal is not a polynomial). An optimally designed polynomial of order N has slope nearly $O(N^{-1})$. There will be ripples in the graph of $P(e^{\imath\omega})$—a monotonic polynomial cannot provide such a sharp cutoff. The Daubechies filters are necessarily less sharp: $O(N)$ becomes $O(\sqrt{N})$.

Theorem 6 *The slope of $P(e^{\imath\omega})$ is approximately $\sqrt{p/\pi}$ at $\omega = \pi/2$. The transition from nearly 1 to nearly 0 is over an interval (i.e. transition band) of with $2\sqrt{2/p}$.*

Proof. The integral in Eq.(2.11) has derivative $\sin^{2p-1}(\pi/2) = 1$ at $\omega = \pi/2$. The slope of $P(e^{\imath\omega})$ is exactly the constant $-c_p^{-1}$. From the proof of the previous theorem, this is $-\sqrt{p/\pi} + O(p^{-3/2})$. To measure the drop in $P(e^{\imath\omega})$ around $\omega = \pi/2$, we integrate from $\pi/2 - \sigma/\sqrt{p}$ to $\pi/2 + \sigma/\sqrt{p}$. Shifting by $\pi/2$ to center the integral, and scaling by $\theta = \tau/\sqrt{p}$, the drop is

$$c_p^{-1}\int_{-\sigma/\sqrt{p}}^{\sigma/\sqrt{p}} \sin^{2p-1}\theta d\theta \simeq \frac{1}{c_p\sqrt{p}}\int_{-\sigma}^{\sigma}(1-\frac{\tau^2}{2p})^{2p-1}d\tau$$
$$\simeq \frac{1}{\sqrt{\pi}}\int_{-\sigma}^{\sigma} e^{-\tau^2}\,d\tau.$$

Thus 95% of the drop comes for $\sigma = \sqrt{2}$ (within two standard derivatives of the mean, for the normal distribution). This transition interval has width $\Delta\omega = 2\sqrt{2/p}$, as the theorem predicts. That rule was found experimentally by Kaiser and Reed at the beginning of the triumph of digital filters. □

2.2 Asymptotics of Daubechies Mini-phase Wavelets

2.2.1 Introduction

Orthogonal wavelets with compact support were announced by Ingrid Daubechies in 1988. For each $p = 1, 2, \cdots$, she created a wavelet supported on $[0,\ 2p-1]$ with p vanishing moments. Our goal is to understand the asymptotic behavior of the scaling functions and the wavelets as $p \to \infty$. The construction begins with the "maxflat minphase lowpass filter" of length $2p$. From its coefficients $h_p[n]$ we form the transfer function or the filter polynomial $H_p(z)$, and there are four main steps to analyze as $p \to \infty$:

(1) The $2p-1$ zeros of $H_p(z) = \sum_n h_p[n]z^{-n}$,

(2) The phase of $H_p(z)$ on the unit circle $z = e^{i\omega}$(we keep using $H_p(\omega)$ for $H_p(e^{i\omega})$),

(3) The scaling function $\phi_p(t)$ with Fourier transform $\prod_{k=1}^{\infty} H_p(\omega/2^k)$,

(4) The wavelet $w_p(t) = \sum_k (-1)^n h_p[2p-1-n]\phi_p(2t-n)$.

In the previous section, we have analyzed the zero distribution of $H_p(z)$. The phase analysis of $H_p(z)$ was carried out in Kateb and Lemarié [41, 1995]. This section brings step (3) and (4) near to completion, based on the Kateb-Lemarié's analysis of step (2). The phase is of crucial importance because orthogonal filters cannot be symmetric (beyond the Haar case $p = 1$). We show that the filter coefficients and the scaling functions have similar asymptotic behavior (but not identical! See Section 6).

The zeros of $H_{70}(z)$ are shown in Figure 2-6. There are 70 zeros at $z = -1$, or $\omega = \pi$, which makes the function "maxflat". The other 69 zeros are inside the unit circle, which makes it "miniphase". The graph of $|H_{70}(\omega)|$ shows that the filter is "lowpass"; the magnitude is near zero for high frequencies. This graph approaches the ideal one-zero function as $p \to \infty$. Then the magnitude of the infinite product $\widehat{\phi}_p(\omega) = \prod_{k=1}^{\infty} H_p(\omega/2^k)$ approaches the characteristic function of $[-\pi, \pi]$.

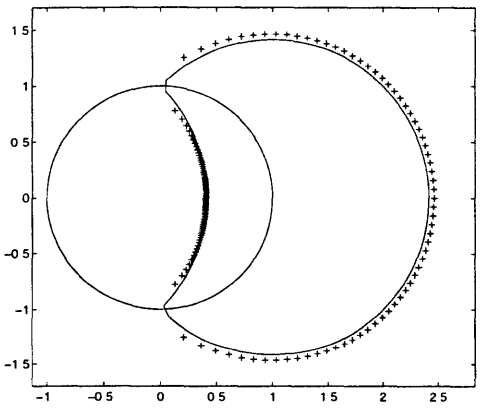

Figure 2-6: H_{70} has 70 zeros at $z = -1$ (not shown in the graph) and 69 zeros inside the unit circle. Those outside the unit circle in the graph are their reciprocals.

The z-transform $(z+z^{-1}/2 = 1-2y)$ of the limit curve $|4y(1-y)| = 1$ in y-plane is two intersected circles $|z \pm 1| = \sqrt{2}$ in z-plane. By Theorem 1, all preimages of Y's are in the right half plane: half inside the unit circle and half outside. We take all the zeros inside to construct the Daubechies mini-phase filter $H_p(z)$ according to Eq.(2.5). Those zeros approach the circular arc $|z+1| = \sqrt{2}$ from inside by Theorem 2 (see also Figure 2-6).

From the two asymptotic formulas for the zeros (along the circular arc and near the end points $\pm i$), Kateb and Lemarié found the leading term in the phase $\arg(H_p(\omega))$. They multiplied the $2p-1$

linear factors and added phases. The result is naturally expressed in terms of the *group delay* grd:

$$\text{grd}(H_p(\omega)) = -\frac{d}{d\omega}(\text{phase of } H_p(\omega)) = p\,g(\omega) + O(p^{1/2}), \tag{2.12}$$

with

$$g(\omega) = \frac{1}{2} + \frac{1}{2\pi}\frac{\cos\omega}{\sin\omega}\ln\frac{1-\sin\omega}{1+\sin\omega}. \tag{2.13}$$

(The $1/2$ term was not in Kateb and Lemarié's paper [41, 1995] and appears here because we shifted the highpass filter to make it causal.) This even function $g(\omega)$ is analytic and convex on $(-\pi, \pi)$. Its Taylor expansion around $\omega = 0$ is $(1/2 - 1/\pi) + \omega^2/6\pi + O(\omega^4)$. Its derivative is infinite at $\omega = \pm\pi$.

Our step (3) in the analysis must work with the infinite product $\widehat{\phi}_p(\omega) = \prod_{k=1}^{\infty} H_p(\omega/2^k)$. The phases add, and the derivative for the group delay contributes a factor $1/2^k$. This makes the infinite sum converge:

$$\text{grd}(\widehat{\phi}_p(\omega)) = p\,G(\omega) + O(p^{1/2}) \tag{2.14}$$

with

$$G(\omega) = \sum_{k=1}^{\infty} \frac{1}{2^k} g(\frac{\omega}{2^k}). \tag{2.15}$$

The function $G(\omega)$ and its derivative are shown in Figure 2-7. The series for $G(\omega)$ gives $G(0) = g(0) = 1/2 - 1/\pi$ and $G''(0) = g''(0)/7 = 1/21\pi$. The numbers $\tau_0 = G(0) \simeq .1817$ and $\tau_1 = G(\pi) \simeq .3515$ will be called the *transition time* in the eventual asymptotic formula for $\phi_p(\tau)$, with $\tau = t/p$.

We note an important difference in the time scale t/p, compared to the asymptotics of B–splines (see Unser, Aldroubi and Eden [46, 1992]). The splines are symmetric. They approach Gaussians with scaling $t/\sqrt{p}$. The spline wavelets approach cosine–modulated Gaussians on that scale too. The Central Limit Theorem is at work. Our problem requires a further step, and the technical tool will be the method of *stationary phase*. This enters when we invert the Fourier transform:

$$\phi_p(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{\phi}_p(\omega) e^{\imath t\omega}\, d\omega \simeq \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ipG^{(-1)}(\omega)} e^{\imath t\omega}\, d\omega = \Phi_p(t). \tag{2.16}$$

The scaled phase is approximately $G^{(-1)}(\omega) = \int_0^{\omega} G(\theta)\, d\theta$. Our main task is to justify this approximation $\phi_p(t) \simeq \Phi_p(t)$ based on magnitude and phase. Then we analyze the asymptotics of

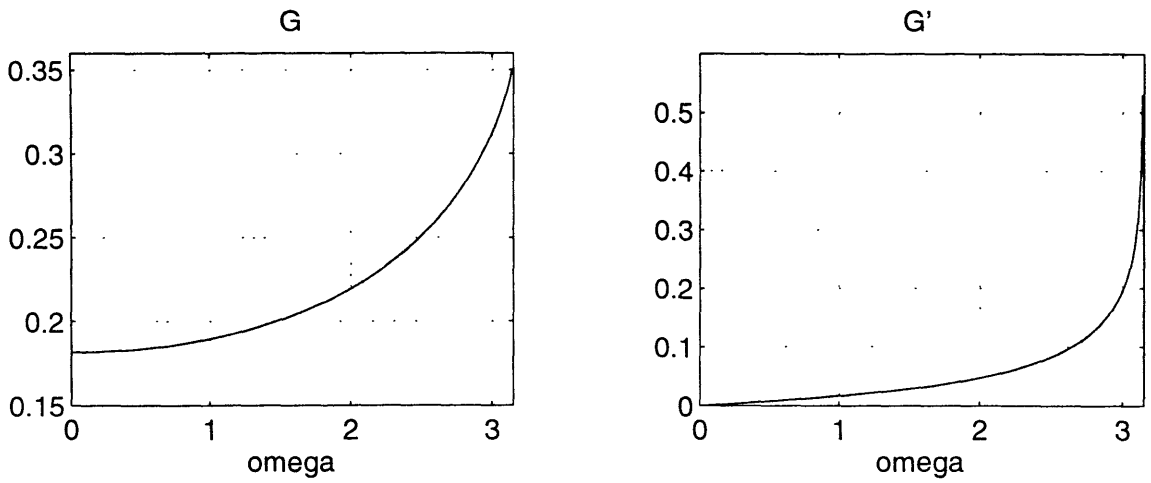

Figure 2-7: $G(\omega)$ and $G'(\omega)$

$\Phi_p(t)$ as $p \to \infty$. The results are summarized in our abstract.

Throughout this section, "$A \simeq B$" means that A and B share the same leading term (when expanded in terms of a certain asymptotic parameter). The symbol "$a \ll 1$" means that a is small enough (usually this can be characterized by some asymptotic parameter). We refer to [4, 51, 56, 81] for a full theory of asymptotic analysis on integrals.

2.2.2 Accuracy of Approximations

We define the following approximations to $\phi_p(t)$:

$$\phi_p^c(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \widehat{\phi}_p(\omega) e^{it\omega}\, d\omega \qquad \text{(frequency limited to } |\omega| \le \pi) \tag{2.17}$$

$$\psi_p(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\arg(\widehat{\phi}_p)} e^{it\omega}\, d\omega \qquad \text{(magnitude taken as 1)} \tag{2.18}$$

$$\Phi_p(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ipG^{(-1)}(\omega)} e^{it\omega}\, d\omega \qquad \text{(leading term of phase)} \tag{2.19}$$

The integral $G^{(-1)}(\omega) = \int_0^\omega G(\theta)d\theta$ approximates the phase. Our main goal in this section is to justify these approximations. And in the next two sections, we will give the asymptotic analysis of $\Phi_p(t)$.

Spectrum of the scaling function $\phi_p(t)$

We take the Fourier Transform and its inverse to be

$$\widehat{\phi}(\omega) = \int_{-\infty}^{\infty} \phi(t) e^{-it\omega} dt \qquad \text{and} \qquad \phi(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{\phi}(\omega) e^{it\omega}\, d\omega$$

Therefore, for any two square integrable functions $f(t)$ and $g(t)$,

$$\langle f(t),\, g(t) \rangle = \frac{1}{2\pi} \langle \widehat{f}(\omega),\, \widehat{g}(\omega) \rangle. \tag{2.20}$$

For each p, let $E_p(\omega) = |\widehat{\phi}_p(\omega)|^2$ and $P_p(\omega) = |H_p(e^{\imath\omega})|^2$. In order to use $P_p(\omega)$ to estimate the L_q-norm of $\widehat{\phi}_p(\omega)$, we need a detailed description of its behavior outside a finite spectral interval. It is known that asymptotically $\phi_p \in C^{-\mu p}$ with $\mu \simeq .2$, so that $\widehat{\phi}_p$ decays like $\omega^{-\mu p}$ at infinity (see Daubechies [10, Chapter 7]). But this is not sufficient to justify (in the sense of L_q-norm) our attempt to drop the spectrum of $\phi_p(t)$ outside $[-\pi,\, \pi]$ without significant loss of energy. Meyer provides the right bound, even though it is rougher in terms of the estimation of regularity compared to the results in Daubechies [10, Chapter 7].

Lemma 4 (See Meyer [50, p. 103]) *There exists a positive number α, such that $E_p(\omega) \le 2^{-2\alpha p j}$ for any p and any $\omega \in [\frac{\pi}{2} 2^j,\, \pi 2^j]$, $j = 0, 1, \cdots$. Therefore $E_p(\omega) \le (\pi/\omega)^{\alpha p}$ for any positive ω.*

Corollary 3 *For any positive q and $0 < \epsilon \ll 1$,*

$$||\widehat{\phi}_p||_{L_q(\mathbb{R})} = ||\widehat{\phi}_p||_{L_q[-\pi-\epsilon,\, \pi+\epsilon]} + \textit{exponentially small term} \qquad \textit{as} \quad p \to \infty \tag{2.21}$$

Lemma 5 $0 < \epsilon \ll 1$. *There exists $\delta_\epsilon \in (0,\, 1)$ such that*

$$E_p(\omega) \begin{cases} \le P_p(\frac{\omega}{2}) = O(\delta_\epsilon^p) & \omega \in [\pi + \epsilon,\, 2\pi] \\ = P_p(\frac{\omega}{2})\,(1 + O(\delta_\epsilon^p)) & \omega \in [0,\, \pi + \epsilon] \end{cases} \tag{2.22}$$

Proof.

1] From Eq.(2.11),

$$P_p(\omega) = 1 - R_p(\omega), \qquad \text{with} \qquad R_p(\omega) = c_p^{-1} \int_0^{\omega} \sin^{2p-1}\theta\, d\theta$$

where c_p is the constant that makes P_p vanish at π. Stirling's formula shows that c_p has the leading term $\sqrt{\pi/p}$. Note $P_p(\omega) = R_p(\pi - \omega)$ is the mirror image of $R_p(\omega)$ with respect to

$\omega = \pi/2$. Since $\frac{2}{\pi}\theta \le \sin\theta \le \theta$ on $[0, \pi/2]$, we have

$$R_p(\omega) = c_p^{-1} \int_0^\omega \sin^{2p-1}\theta \, d\theta \le c_p^{-1} \omega \sin^{2p-1}\omega \le c_p^{-1} \frac{\pi}{2} \sin^{2p}\omega. \tag{2.23}$$

It also gives $R_p(\omega) \le c_p^{-1}\omega^{2p}$.

2] Noticing that $E_p(\omega) = \prod_{k=1}^{\infty} P_p(\omega/2^k)$ and $\prod_{k=1}^{K}(1 - x_k) \ge 1 - \sum_{k=1}^{K} x_k$, for any $x_k \in [0, 1]$, one has, for any $\omega \in [0, \pi + \epsilon]$,

$$\begin{aligned} E_p(\omega) &= P_p(\frac{\omega}{2}) \prod_{k=2}^{\infty} (1 - R_p(\frac{\omega}{2^k})) \\ &\ge P_p(\frac{\omega}{2})(1 - c_p^{-1} \sum_{k=2}^{\infty} \frac{\omega^{2p}}{4^{pk}}) \\ &= P_p(\frac{\omega}{2})(1 - \frac{c_p^{-1}}{1 - 4^{-p}} \frac{\omega^{2p}}{4^{2p}}). \end{aligned} \tag{2.24}$$

For $\omega \in [\pi + \epsilon, 2\pi]$,

$$E_p(\omega) \le P_p(\frac{\omega}{2}) = R_p(\pi - \frac{\omega}{2}) \le c_p^{-1} \frac{\pi}{2} \cos^{2p}\frac{\epsilon}{2}. \tag{2.25}$$

By (2.24) and (2.25), any $\delta_\epsilon \in (\max[(\frac{\pi+\epsilon}{4})^2, \cos^2\frac{\epsilon}{2}], 1)$ makes the lemma true. □

Lemma 6 *For any positive q,*

$$\int_0^{\frac{\pi}{2}} R_p^q(\theta) \, d\theta = O(p^{-1/2}) \qquad \text{as } p \to \infty. \tag{2.26}$$

Proof. For $0 \le \omega \ll 1$,

$$\begin{aligned} R_p(\frac{\pi}{2} - \omega) &= \frac{1}{2} - c_p^{-1} \int_0^\omega \cos^{2p-1}\theta \, d\theta \\ &\simeq \frac{1}{2} - \sqrt{\frac{p - 1/2}{\pi}} \int_0^\omega e^{-(p-1/2)\,\theta^2} \, d\theta \\ &= \frac{1}{2} - \frac{1}{\sqrt{\pi}} \int_0^{\sqrt{p-1/2}\,\omega} e^{-\theta^2} \, d\theta \\ &= \frac{1}{2} \operatorname{erfc}(\sqrt{p - 1/2}\,\omega), \end{aligned}$$

where the complementary error function is defined by

$$\mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2}\, dt.$$

Since $R_q(\omega)$ has a boundary layer near $\omega = \pi$, we have (for any $q > 0$)

$$\begin{aligned}
\int_0^{\frac{\pi}{2}} [R_p(\omega)]^q\, d\omega &= \int_0^{\frac{\pi}{2}} [R_p(\frac{\pi}{2} - \omega)]^q\, d\omega \\
&\simeq \int_0^{\epsilon} [R_p(\frac{\pi}{2} - \omega)]^q\, d\omega \\
&= \frac{1}{2^q} \int_0^{\epsilon} [\,\mathrm{erfc}(\sqrt{p - 1/2}\,\omega)]^q\, d\omega \\
&= \frac{1}{2^q \sqrt{p - 1/2}} \int_0^{\sqrt{p-1/2}\,\epsilon} [\,\mathrm{erfc}(\theta)]^q\, d\theta = O(p^{-1/2})
\end{aligned}$$

□

Now we can analyze the accuracy of our approximations.

Approximating $\phi_p(t)$ with $\phi_p^c(t)$

Theorem 7 *Let $r_p(t) = \phi_p(t) - \phi_p^c(t)$. Then,*

$$||r_p(t)||_{L_\infty(\mathbb{R})} = O(p^{-\frac{1}{2}}) \tag{2.27}$$

$$||r_p(t)||_{L_2(\mathbb{R})} = O(p^{-\frac{1}{4}}) \tag{2.28}$$

Proof. By definition, $\widehat{r}_p(\omega)$ is just the truncated spectrum of $\widehat{\phi}_p(\omega)$ on $\mathbb{R}\backslash[-\pi, \pi]$. By the corollary of Lemma 4, the L_q norm of $\widehat{r}_p$ is determined by its restriction on $[-2\pi, -\pi] \cup [\pi, 2\pi]$ up to a p-exponentially small error. Then,

$$\begin{aligned}
||\widehat{r}_p||_{L_q(\mathbb{R})} &\simeq 2^{1/q} ||\widehat{r}_p||_{L_q[\pi, 2\pi]} = 2^{1/q} ||\sqrt{E_p(\omega)}||_{L_q[\pi, 2\pi]} \\
&\simeq 2^{1/q} ||\sqrt{P_p(\frac{\omega}{2})}||_{L_q[\pi, 2\pi]} \qquad \text{(by Lemma 5)} \\
&= 4^{1/q} ||\sqrt{R_p(\omega)}||_{L_q[0, \pi/2]} \qquad \text{(by mirror relation)} \\
&= O(p^{-1/2q}) \qquad \text{(by Lemma 6)}
\end{aligned} \tag{2.29}$$

Then (2.27) and (2.28) follow immediately from $||r_p(t)||_{L_2(\mathbb{R})} = \frac{1}{\sqrt{2\pi}} ||\widehat{r}_p||_{L_2(\mathbb{R})}$ and $||r_p(t)||_{L_\infty(\mathbb{R})} \leq \frac{1}{2\pi} ||\widehat{r}_p||_{L_1(\mathbb{R})}$. □

Approximating $\phi_p^c(t)$ with $\psi_p(t)$

Theorem 8 *Let $s_p(t) = \phi_p^c(t) - \psi_p(t)$. Then*

$$||s_p(t)||_{L_\infty(\mathbb{R})} = O(p^{-\frac{1}{2}}) \tag{2.30}$$

$$||s_p(t)||_{L_2(\mathbb{R})} = O(p^{-\frac{1}{4}}) \tag{2.31}$$

Proof.

1] For (2.30),

$$\begin{aligned}
|s_p(t)| &= \frac{1}{2\pi}\left|\int_{-\pi}^{\pi}(\widehat{\phi}_p - \widehat{\phi}_p/|\widehat{\phi}_p|)\, e^{\imath t\omega}\, d\omega\right| \leq \frac{1}{2\pi}\int_{-\pi}^{\pi}|1-|\widehat{\phi}_p|\,|\, d\omega \\
&\leq \frac{1}{2\pi}\int_{-\pi}^{\pi}(1-|\widehat{\phi}_p|\,)\,(1+|\widehat{\phi}_p|)\, d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi}(1-E_p(\omega))\, d\omega \\
&\simeq \frac{1}{\pi}\int_0^{\pi}(1-P_p(\frac{\omega}{2}))\, d\omega = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} R_p(\theta)\, d\theta = O(p^{-\frac{1}{2}})
\end{aligned}$$

where the approximation has an exponentially small error (by Lemma 5). This argument is valid for any real t. Therefore (2.30) is true.

2] For (2.31)

$$\begin{aligned}
||s_p||^2_{L_2(\mathbb{R})} &= \frac{1}{2\pi}||\widehat{\phi}_p - \widehat{\phi}_p/|\widehat{\phi}_p|\,||^2_{L_2[-\pi,\,\pi]} = \frac{1}{2\pi}||1-|\widehat{\phi}_p|\,||^2_{L_2[-\pi,\,\pi]} \\
&\leq \frac{1}{2\pi}\int_{-\pi}^{\pi}(1-|\widehat{\phi}_p|)^2(1+|\widehat{\phi}_p|)^2\, d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi}[1-E_p(\omega)]^2\, d\omega \\
&\simeq \frac{1}{2\pi}\int_{-\pi}^{\pi}[1-P_p(\frac{\omega}{2})]^2\, d\omega = \frac{2}{\pi}\int_0^{\frac{\pi}{2}}[R_p(\theta)]^2\, d\theta = O(p^{-1/2})
\end{aligned}$$

□

Approximating $\psi_p(t)$ with $\Phi_p(t)$ (I)

Both $\psi_p(t)$ and $\Phi_p(t)$ are entirely determined by their phases. By (2.14), the phase difference has order $O(p^{1/2})$, which is very large in the usual sense. This prevents any attempt to explain their similarity by estimating the L_p norms of their spectra. A new mechanism has to be introduced to explain their close relation. That is the stationary phase method we will discuss in the next section. Before finishing this section, we interpret the results obtained so far.

By Theorems 7 and 8, one has

$$||\phi_p(t) - \psi_p(t)||_{L_\infty(\mathbb{R})} = O(p^{-1/2}) \tag{2.32}$$

$$||\phi_p(t) - \psi_p(t)||_{L_2(\mathbb{R})} = O(p^{-1/4}) \tag{2.33}$$

(2.32) is not so satisfactory since we will see later that $O(p^{-1/2})$ itself is the characteristic magnitude of the scaling function $\phi_p(t)$ (see section 2.2.4). However, with the help of (2.33), one can show that the set on which $|\phi_p(t) - \psi_p(t)|$ reaches $O(p^{-1/2})$ is small. The exact statement is Theorem 9.

Theorem 9 *For any positive α and C, we define*

$$A_{\alpha,C} = \{t \in \mathbb{R} : \ |\phi_p(t) - \psi_p(t)| \geq C\, p^{-\alpha}\}$$

Then $\mu(A_{\alpha,C}/p) \leq C' p^{2(\alpha-3/4)}$, where μ is Lebesgue measure on the real line.

The proof uses the Chebyshev inequality to estimate the measure of $A_{\alpha,C}$ by the L_2 norm of $\phi_p(t) - \psi_p(t)$.

Corollary 4 *Set $\tau = t/p$, and denote any $f(t) = f(p\tau)$ still by $f(\tau)$ for simplicity. Then for any $\alpha < 3/4$ and $C > 0$, $\lim_{p\to\infty} \mu\{\tau \in \mathbb{R} : \ |\phi_p(\tau) - \psi_p(\tau)| \geq C\, p^{-\alpha}\} = 0$.*

This result tells what one can hope for from the approximation to $\phi_p(t)$ by $\Phi_p(t)$ (or $\phi_p(\tau)$ by $\Phi_p(\tau), \tau = t/p$). The approximations defined by (2.17) and (2.18) have already introduced a non-negligible error of at least order $O(p^{-3/4})$. Therefore for the further approximation by (2.19), it is only meaningful to talk about an accuracy of order $O(p^{-\alpha})$ with $\alpha < 3/4$.

2.2.3 Fourier Integrals with Large Parameters

Before we investigate the asymptotic form of the approximate scaling function $\Phi_p(t)$, it is helpful to review and extend some results in asymptotic analysis.

Fourier integrals with one large parameter have the form

$$I_\lambda = \int_a^b f(\omega) e^{-\imath\lambda F(\omega)}\, d\omega \tag{2.34}$$

with real λ, f, F. The asymptotic analysis usually deals with $\lambda \gg 1$. The interval $[a, b]$ can be finite or infinite. Our problem is the finite case. The basic results can be stated as follows:

Statement 1 (End Point Contribution) *Suppose that F is a C^1 function and has no critical point inside the closed interval (or equivalently, no stationary phase), and f is a continuous function.*

Then the leading asymptotic magnitude is proportional to $1/\lambda$*:*

$$I_\lambda = \frac{1}{i\lambda}\left[f(a)\frac{e^{-i\lambda F(a)}}{F'(a)} - f(b)\frac{e^{-i\lambda F(b)}}{F'(b)}\right] + o(\lambda^{-1}) \tag{2.35}$$

Statement 2 (Stationary Phase Contribution) *Suppose*

1) F *and* f *are* C^1 *and* C^0 *functions on* $[a, b]$ *respectively,*

2) $c \in (a, b)$ *is the only critical point of* F *on* $[a, b]$ *and* $f(c) \neq 0$*,*

3) F *is* C^2 *around this critical point and* $F''(c) \neq 0$*.*

Then the leading asymptotic magnitude is proportional to $1/\sqrt{\lambda}$*:*

$$I_\lambda = f(c)\sqrt{\frac{2\pi}{|\lambda F''(c)|}}e^{-i[\lambda F(c)+sign(F''(c))\frac{\pi}{4}]} + o(\lambda^{-\frac{1}{2}}) \tag{2.36}$$

The proofs of these two statements can be found in many asymptotic analysis textbooks (for instance [1, 5, 7, 11]) with a little modification on the regularity of F.

Next, let's consider the *doubly parameterized Fourier integral* (DPFI):

$$I_\lambda(\tau) = \int_a^b e^{-i\lambda F(\omega,\tau)}\, d\omega, \qquad \text{for real } \tau. \tag{2.37}$$

At any fixed time τ, Statements 1 and 2 can be applied to DPFI. As long as the regularity conditions for ω are satisfied uniformly during a certain period of time, the approximations hold uniformly with respect to τ. Attention must be paid to the so-called *transition period* of τ. It could happen that one period of time belongs to the case of Statement 1 uniformly, and some other period to that of Statement 2, while the rest is a transition period between these two cases.

The transition phenomenon is structurally stable and therefore universal. It occurs near "turning points". To begin, we consider an idealized DPFI with $a = -1, b = 1$, and $F(\omega, \tau) = \frac{\omega^3}{3} - \tau\omega$. We plot the critical points of $F(\omega, \tau)$ as a function of ω with parameter τ on the τ–ω plane. F has two critical points for $\tau \in (0, 1)$. Outside $[0, 1]$, there are no critical points on the interval $[a, b]$. Two classes of transitions with different origins occur here (see Figure 2-8):

1) Near $\tau = 0$, the two critical points coming from the right side collide and cancel each other (and actually go to the imaginary axis).

2) Near $\tau = 1$, the pair of critical points coming from the left go out of the integral domain.

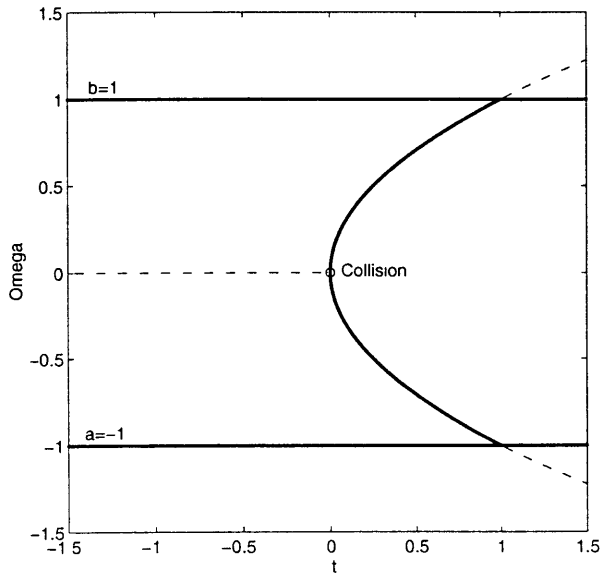

Figure 2-8: Transition Phenomenon

In this example, the leading magnitude of $I_\lambda(\tau)$ is $1/\lambda$ uniformly on any compact set of negative time (by Statement 1) and $1/\sqrt{\lambda}$ uniformly on any compact set of positive time (by Statement 2). The n, how is this jump in magnitude realized around time zero? The answer is given by Theorem 10. We only sketch the proof, since a strict proof may take unsuitably long. Similar work on uniform approximations of integrals can be found in Wong [81, Chapter 7].

Theorem 10 *Let $L(\omega)$ be a function on $[a, b]$ $(a < 0 < b)$ that satisfies*

1) L is a C^1 function;

2) $\omega = 0$ is the unique critical point of L ;

3) L is C^3 around 0, and $L(0) = L'(0) = L''(0) = 0, L'''(0) = \alpha > 0$.

Suppose $F(\omega, \tau) = L(\omega) - \tau\omega$. Then for $|\tau| \ll 1$ (or precisely, $\tau = O(\lambda^{-\frac{2}{3}})$)

$$I_\lambda(\tau) = 2\pi \sqrt[3]{2/\alpha\lambda}\, \mathrm{Ai}(-\sqrt[3]{2\lambda^2/\alpha}\, \tau) + o(\lambda^{-\frac{1}{3}}), \tag{2.38}$$

where Ai *is the Airy function.*

Proof. Since $|\tau| \ll 1$, the leading magnitude of $I_\lambda(\tau)$ is completely determined by the local property of L near $\omega = 0$. Therefore we have

$$\begin{aligned} &= \int_a^b e^{-\imath\lambda(L(\omega)-\tau\omega)}\, d\omega \simeq \int_{-\delta}^{\delta} e^{-\imath\lambda(L(\omega)-\tau\omega)}\, d\omega \\ &\simeq \int_{-\delta}^{\delta} e^{-i\lambda(\frac{\alpha\omega^3}{6}-\tau\omega)}\, d\omega \simeq \int_{-\infty}^{\infty} e^{-i\lambda(\frac{\alpha\omega^3}{6}-\tau\omega)}\, d\omega. \end{aligned} \tag{2.39}$$

Scaling the variables by setting $\tau = -\sqrt[3]{\alpha/2\lambda^2}\, t$ and $\omega = \sqrt[3]{2/\alpha\lambda}\,\theta$ yields

$$\begin{aligned} I_\lambda(\tau) &\simeq \sqrt[3]{2/\alpha\lambda} \int_{-\infty}^{\infty} e^{-i(\frac{\theta^3}{3}+t\theta)}\, d\theta + o(\lambda^{-\frac{1}{3}}) \\ &= 2\pi \sqrt[3]{2/\alpha\lambda}\, \mathrm{Ai}(t) + o(\lambda^{-\frac{1}{3}}) \\ &= 2\pi \sqrt[3]{2/\alpha\lambda}\, \mathrm{Ai}(-\sqrt[3]{2\lambda^2/\alpha}\,\tau) + o(\lambda^{-\frac{1}{3}}). \end{aligned}$$

□

Now let us consider the second type of transition. A generic case is given by the following theorem.

Theorem 11 *Suppose $F(\omega,\tau) = L(\omega) - \tau\omega$, where $L(\omega) \in C^1[-a,a]$ satisfies*

1) $L(\omega) = L(-\omega)$;

2) $\omega = 0$ is the only critical point for L;

3) L is C^2 around $\omega = a$ and $L''(a) \neq 0, \infty$.

Then for $|\tau - A_1| \ll 1$ (precisely, $|\tau - A_1| = O(\lambda^{-\frac{1}{2}})$), we have

$$I_\lambda(\tau) = \sqrt{\frac{8\pi}{\lambda|A_2|}} \left[\cos(\frac{\lambda}{2}T)\,\mathrm{coserf}(\sqrt{\lambda}S) + \mathrm{sign}(A_2)\sin(\frac{\lambda}{2}T)\,\mathrm{sinerf}(\sqrt{\lambda}S)\right] + o(\lambda^{-\frac{1}{2}}) \tag{2.40}$$

$$S = \mathrm{sign}(A_2)\frac{A_1-\tau}{\sqrt{|A_2|}}, \qquad T = 2(\tau a - A_0) + \frac{(A_1-\tau)^2}{A_2}$$

with $A_0 = L(a), A_1 = L'(a), A_2 = L''(a)$, and two Fresnel integrals,

$$\mathrm{coserf}(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} \cos\frac{\theta^2}{2} d\theta, \;\; \mathrm{sinerf}(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} \sin\frac{\theta^2}{2} d\theta. \tag{2.41}$$

Proof. Again we only sketch the proof and refer to Wong [81, Chapter 7].

1] One only needs to prove the case when $A_2 > 0$. For negative A_2, simply replace L and τ by $-L$ and $-\tau$.

2] As $|\tau - A_1| \ll 1$,

$$\begin{aligned}
I_\lambda(\tau) &= \int_{-a}^{a} e^{-\imath\lambda(L(\omega)-\tau\omega)}\, d\omega \simeq \int_{a-\delta}^{a} + \int_{-a}^{-a+\delta} e^{-i\lambda(L(\omega)-\tau\omega)}\, d\omega \\
&= 2\mathrm{Re}\int_{a-\delta}^{a} e^{-i\lambda(L(\omega)-\tau\omega)}\, d\omega \simeq 2\mathrm{Re}\int_{a-\delta}^{a} e^{-\imath\lambda[A_0+A_1(\omega-a)+\frac{A_2}{2}(\omega-a)^2-\tau\omega]}\, d\omega \\
&\simeq 2\mathrm{Re}\, e^{-\imath\lambda(A_0-\tau a)} \int_{-\delta}^{0} e^{-i\lambda[(A_1-\tau)\theta+\frac{A_2}{2}\theta^2]} d\theta \qquad (\theta = \omega - a) \\
&\simeq 2\mathrm{Re}\, e^{-\imath\lambda(A_0-\tau a)} \int_{-\infty}^{0} e^{-i\lambda[(A_1-\tau)\theta+\frac{A_2}{2}\theta^2]} d\theta \qquad (\text{since } |A_1-\tau| \ll 1) \\
&= \frac{2}{\sqrt{\lambda A_2}}\mathrm{Re}\, e^{-i\lambda(A_0-\tau a)}\, e^{\imath\frac{\lambda}{2}\frac{(A_1-\tau)^2}{A_2}} \int_{-\infty}^{\sqrt{\lambda}S} e^{-i\frac{u^2}{2}} du \\
&= \sqrt{\frac{8\pi}{\lambda A_2}}\left[\cos(\frac{\lambda}{2}T)\,\mathrm{coserf}(\sqrt{\lambda}S) + \sin(\frac{\lambda}{2}T)\,\mathrm{sinerf}(\sqrt{\lambda}S)\right].
\end{aligned}$$

□

Unfortunately, in our case, $L(\omega) = G^{(-1)}(\omega)$, hence $A_2 = \infty$. The generic result of Theorem 11 doesn't apply to this special case. In fact for $\Phi_p(\tau)$, one has to deal with the following type of local integral with large parameter λ:

$$I_\lambda(\tau) = \int_0^c e^{-\imath\lambda(\tau\theta - a\theta^2\log\theta + b\theta^2)}\, d\theta \qquad \tau \simeq 0,\ a, b, c > 0 \tag{2.42}$$

Its asymptotic analysis may be very involved. However this transition is of less importance to us for the reason indicated in the last paragraph of section 2.2.2. So we leave this analysis work for the future.

2.2.4 Asymptotic Structure of $\Phi_p(t)$ and $\psi_p(t)$

Asymptotic form of $\Phi_p(t)$

We are ready now to establish the asymptotic form of $\Phi_p(t)$. Recall that we introduced the lower transition time $\tau_0 = G(0) \simeq .1817$ and upper transition time $\tau_1 = G(\pi) \simeq .3515$. The definition of $\Phi_p(t)$ in (2.19) and DPFI in (2.37) implies a scaling $t = \tau p$. For simplicity, we continue to use $\Phi_p(\tau)$ to denote $\Phi_p(p\tau)$.

Result 1 (Stationary Phase) *Uniformly on any compact subset K of (τ_0, τ_1)*

$$\Phi_p(\tau) = \sqrt{\frac{2}{\pi p G'(\omega_\tau)}} \cos[p(G^{(-1)}(\omega_\tau) - G(\omega_\tau)\omega_\tau) + \frac{\pi}{4}] + o(p^{-1/2}). \tag{2.43}$$

ω_τ is the unique $\omega \in (0, \pi)$ such that $G(\omega) = \tau$.

Let $F = G^{(-1)}(\omega) - \tau\omega$ and apply the result of Statement 2 to DPFI.

Result 2 (Airy Transition) *For $|\tau - \tau_0| \ll 1$,*

$$\Phi_p(\tau) = \sqrt[3]{\frac{42\pi}{p}} \operatorname{Ai}(-\sqrt[3]{42\pi p^2}(\tau - \tau_0)) + o(p^{-\frac{1}{3}}) \tag{2.44}$$

Use Theorem 10 with $L(\omega) = G^{(-1)}(\omega)$, and use $\tau - \tau_0$ instead of τ. Note $\alpha = 1/21\pi$ in this case.

Result 3 (End Points) *Uniformly on any compact subset K of $[\tau_0, \tau_1]^c$,*

$$\Phi_p(\tau) = \frac{1}{p\pi(\tau - \tau_1)} \sin[p(G^{(-1)}(\pi) - \tau\pi)] + o(p^{-1}). \tag{2.45}$$

Apply Statement 1 to DPFI in the case $F = G^{(-1)} - \tau\omega, a = -\pi, b = \pi$.

Result 4 (Front Matching) *(2.43) and (2.44) match over interval $p^{-2/3} \ll \tau - \tau_0 \ll 1$.*

Proof. Let $x = \sqrt[3]{42\pi p^2}(\tau - \tau_0)$ and $K(\omega) = G^{(-1)}(\omega) - G(\omega)\omega$.

1] $\tau - \tau_0 \gg p^{-2/3}$ implies $x \gg 1$. Since

$$\operatorname{Ai}(-x) = \frac{1}{\sqrt{\pi}} x^{-1/4} \sin[\frac{2}{3}x^{3/2} + \frac{\pi}{4}] + O(x^{-7/4}) \qquad \text{as } x \gg 1,$$

the leading term of the expression in (2.44) is given by

$$(\frac{42}{\pi})^{1/4} p^{-1/2} (\tau - \tau_0)^{-1/4} \sin[\frac{2}{3}(42\pi)^{1/2} p(\tau - \tau_0)^{3/2} + \frac{\pi}{4}] \tag{2.46}$$

2] $0 \ll \tau - \tau_0 \ll 1$ implies $\omega_\tau \ll 1$ and $G(\omega) \simeq G(0) + \frac{G''(0)}{2}\omega^2$ as $|\omega| \ll 1$. Therefore $G''(0) = 1/21\pi$ gives $\omega_\tau \simeq [42\pi(\tau - \tau_0)]^{1/2}$ and

$$K(\omega_\tau) \simeq \frac{K'''(0)}{6}\omega_\tau^3 = -\frac{G''(0)}{3}\omega_\tau^3 = -\frac{1}{63\pi}\omega_\tau^3 \simeq -\frac{2}{3}(42\pi)^{1/2}(\tau - \tau_0)^{3/2} \tag{2.47}$$

and

$$G'(\omega_\tau) \simeq G''(0)\omega_\tau \simeq (\frac{2}{21\pi})^{1/2}(\tau - \tau_0)^{1/2} \tag{2.48}$$

Substituting these two equations into the expression in (2.43), the leading term of (2.43) is given exactly by (2.46) for $0 \ll \tau - \tau_0 \ll 1$. □

The end point contribution is of magnitude $O(p^{-1})$. This order is more delicate than our approximation precision (see the last paragraph of section 2.2.2). Therefore this modulated sine wave of order p^{-1} is only the behavior of $\Phi_p(t)$, not that of the scaling function. These ripples are introduced by the ideal lowpass filtering (truncation) in (2.17). However, as one can see from Figure 2-9, part of the earliest ripples given by (2.45) can match $\phi_p(\tau)$ quite well.

The stationary phase contribution lasts asymptotically $\tau_1 - \tau_0 = 0.1698$ in the scaled time τ. During this period, the magnitude has an order of $O(p^{-1/2})$ as noted below (2.33). A better way to interpret (2.43) is that the phase inside the cosine is basically amplified (by p) and shifted (by $\pi/4$) from the Legendre transform of $G^{(-1)}$. $G(\omega)$ provides a natural parameterization for the scaled time period (τ_0, τ_1). Therefore, $\tau = G(\omega_\tau)$ together with (2.43) is the ω_τ-parameterized version of $\Phi_p(\tau)$ on (τ_0, τ_1). This is quite useful for computer plotting (the explicit inverse of G is unnecessary). The total wave number k during this period is entirely determined by the area bounded by $G = G(\omega)$, $\omega = 0$, and $G = \tau_1$ in the left subplot of Figure 2-7, which is approximately 0.4152. Therefore $k \simeq 0.4152p/2\pi \simeq 0.0661p$. Every increment of 30 in p adds two complete waves during this period, asymptotically. This is confirmed by numerical results.

The Airy transition reveals the rich structure of the main lobe (or the wavefront). The main lobe lasts about $p^{-2/3}$ in the scaled time τ and has a magnitude $O(p^{-1/3})$. This makes it the real leader of all waves that follow it: it is much wider than other waves ($p^{-2/3}$ to p^{-1}) and also much higher ($p^{-1/3}$ to $p^{-1/2}$). However, from the viewpoint of energy, Airy transition is insignificant since its total energy is of order $O(p^{-1/3})$.

We have plotted our approximation in Figure 2-9.

Approximating $\psi_p(t)$ with $\Phi_p(t)$ (revisited)

Now we show that $\psi_p(\tau)$ and $\Phi_p(\tau)$ will share the same envelope for the stationary phase period and their graphs have separation $O(p^{-1})$ during this period.

For convenience, let $Q(\omega) = \mathrm{grd}(\widehat{\phi}_p)/p$ and $Q^{(-1)} = \int_0^\omega Q(\omega)\,d\omega = -\arg(\widehat{\phi}_p)/p$ for any fixed p. We state the following facts:

1) $\psi_p(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\imath p(Q^{(-1)}(\omega)-\tau\omega)}\,d\omega$;

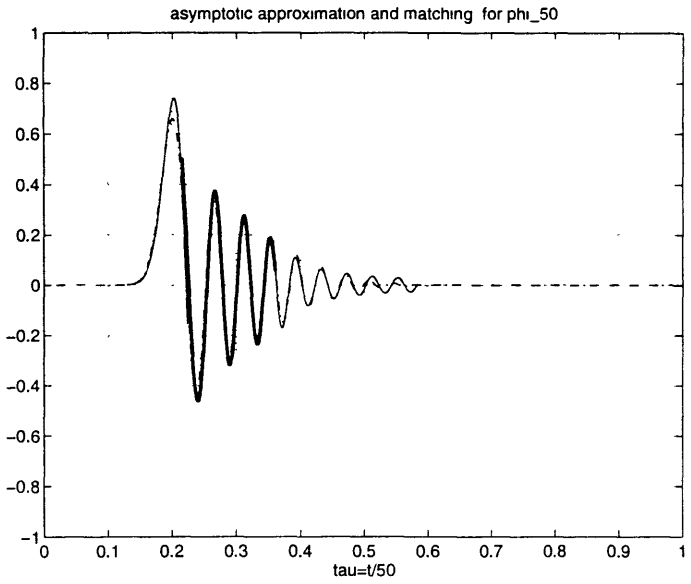

Figure 2-9: Asymptotic approximation and matching for $\phi_{50}(\tau)$ (dashed line). The left and right solid lines are Airy transition and end points contribution. The dark dotted line in the middle is the stationary phase. Each approximation has been uniformly shifted by approximately 0.01 ($= O(p^{-1})$) (see next subsection).

2) $Q(-\omega) = Q(\omega)$, and $Q(\omega)$ is strictly increasing on $[0\,\pi]$;

3) By (2.14), $|Q(\omega) - G(\omega)| \leq cp^{-1/2}$ for any $\omega \in [-\pi, \pi]$. On any compact subset K of $(-\pi, \pi)$ this estimation is differentiable:

$$|Q^{(n)}(\omega) - G^{(n)}(\omega)| \leq c_n\, p^{-1/2} \tag{2.49}$$

Therefore we can apply stationary phase to $\psi_p(\tau)$ uniformly on any compact set K of (τ_0, τ_1) to find that

$$\psi_p(\tau) = \sqrt{\frac{2}{\pi p Q'(\omega'_\tau)}}\cos[p\left(Q^{(-1)}(\omega'_\tau) - Q(\omega'_\tau)\omega'_\tau\right) + \frac{\pi}{4}] + o(p^{-\frac{1}{2}}) \tag{2.50}$$

ω'_τ is the unique $\omega \in (0, \pi)$ such that $Q(\omega) = \tau$.

By (2.49), on the compact set K, one can replace $Q^{(n)}$ and ω'_τ with $G^{(n)}$ and ω_τ (as defined in (2.43)) at the price of an order $O(p^{-1/2})$ phase perturbation, i.e.

$$\psi_p(\tau) = \sqrt{\frac{2}{\pi p G'(\omega_\tau)}}\cos[p\left(G^{(-1)}(\omega_\tau) - G(\omega_\tau)\omega_\tau + O(p^{-1/2})\right) + \frac{\pi}{4}] + o(p^{-\frac{1}{2}}) \tag{2.51}$$

Comparing (2.51) with (2.43), one sees that during the stationary period, $\psi_p(\tau)$ and $\Phi_p(\tau)$ share the same envelope $\sqrt{\frac{2}{\pi p G'(\omega_\tau)}}$. And at each time τ, the phase of the leading term can be made the same (mod 2π) up to an order $O(p^{-1})$ shifting of the scaled time τ. This means that the graphs of $\psi_p(\tau)$ and $\Phi_p(\tau)$ during this period have separation $O(p^{-3/2})$. Since the residue terms in (2.51) and (2.43) actually have magnitude $O(p^{-1})$ (due to the good regularity of both G and Q on K), the graphs of $\psi_p(\tau)$ and $\Phi_p(\tau)$ have distance $O(p^{-1})$.

Similar work can be done for the Airy transition; the phase perturbation still exists.

2.2.5 Asymptotic Structure of Wavelets

Goodness of approximation

The orthogonal highpass filter F_p is the alternating flip of the lowpass filter H_p. In the z–domain, $F_p(z) = -z^{-N}H_p(-z^{-1})$ with $N = 2p-1$. The delay factor z^{-N} makes F_p causal. Its group delay is $\mathrm{grd}(F_p) = N - \mathrm{grd}(H_p)(\omega+\pi)$.

The transform of the wavelet $w_p(t)$ is $\widehat{w}_p(\omega) = F_p(\frac{\omega}{2})\widehat{\phi}_p(\frac{\omega}{2})$, so that $\mathrm{grd}(\widehat{w}_p) = \frac{1}{2}[\mathrm{grd}(F_p)(\frac{\omega}{2}) + \mathrm{grd}(\widehat{\phi}_p)(\frac{\omega}{2})]$. Using the same notations as (2.14) and (2.15), the group delay for wavelets has leading term

$$|\mathrm{grd}(\widehat{w}_p) - [p(\frac{1}{2} + G(\omega)) - \frac{1}{2}]| \le C\sqrt{p}. \tag{2.52}$$

We already know that the spectrum of $\widehat{\phi}_p$ is mainly concentrated on $[-\pi, \pi]$ and the magnitude $|F_p(\omega)|$ converges to the ideal highpass filter. Therefore it is natural to introduce $W_p(t)$ to approximate the wavelet $w_p(t)$:

$$W_p(t) = -\frac{1}{2\pi}\left(\int_{-2\pi}^{-\pi} + \int_{\pi}^{2\pi}\right) e^{-ip(G^{(-1)}(\omega)+\omega/2)+\imath\omega/2}e^{it\omega}\,d\omega \tag{2.53}$$

The minus sign before the integral is because the phase of $\widehat{w}_p$ near $\omega = 0$ is near π by our definition of the highpass filter $F_p(z) = -z^{-N}H_p(-z^{-1})$, though its magnitude is zero at $\omega = 0$. One can repeat the work already done for the scaling functions and obtain the analogues of (2.32) and (2.33).

Asymptotic form of $W_p(t)$

It is always good to use the scaled time $\tau = t/p$. First we have

$$W_p(\tau) = -\frac{1}{2\pi}\left(\int_{-2\pi}^{-\pi} + \int_{\pi}^{2\pi}\right) e^{-ip(G^{(-1)}(\omega)-(\tau-.5)\omega)}e^{\imath\omega/2}\,d\omega \tag{2.54}$$

Let $\tau_0^w = .5 + G(\pi) = .5 + \tau_1 \approx .8515$ and $\tau_1^w = .5 + G(2\pi) \approx 1.0849$ be the lower and upper transition times for the scaled wavelets. Notice that

1) G is smooth inside $(\pi, 2\pi)$ and continuous on $[\pi, 2\pi]$,

2) G is monotonically increasing on $[\pi, 2\pi]$,

Result 5 (Stationary Phase) *Uniformly on any compact subset K of (τ_0^w, τ_1^w)*

$$W_p(\tau) = -\sqrt{\frac{2}{\pi p G'(\omega_\tau)}} \cos[p(G^{(-1)}(\omega_\tau) - G(\omega_\tau)\omega_\tau) + \frac{\pi}{4} + \frac{\omega_\tau}{2}] + o(p^{-\frac{1}{2}}). \tag{2.55}$$

Here ω_τ is the unique $\omega \in (\pi, 2\pi)$ such that $G(\omega) = \tau - .5$.

And similarly, one can write down the endpoint contribution:

Result 6 (End Points) *Uniformly on any compact subset K of $[\tau_0^w, \tau_1^w]^c$,*

$$\begin{aligned} W_p(\tau) = \frac{1}{p\pi} \Bigg(& \frac{1}{\tau_1^w - \tau} \sin[p(G^{(-1)}(2\pi) - (\tau - .5)2\pi) + \pi] \\ & + \frac{1}{\tau - \tau_0^w} \sin[p(G^{(-1)}(\pi) - (\tau - .5)\pi) + \pi/2] \Bigg) + o(p^{-1}). \end{aligned} \tag{2.56}$$

There is no Airy transition for wavelets. The group delay of $\widehat{w}_p$ on $[-2\pi, -\pi] \cup [\pi, 2\pi]$ has no critical point. Another interpretation is that the Airy wavefront has been removed by the highpass filter F_p. Therefore the characteristic scale for wavelets is: magnitude $= O(p^{-1/2})$; lasting time(scaled) $= \tau_1^w - \tau_0^w \approx .2334$. We have plotted our stationary phase approximation in Figure 2-10. Beyond this stationary phase period, the magnitude of $w_p(t)$ is at most $O(p^{-1})$, which is energy insignificant.

2.2.6 Asymptotic Structure of the Filter Coefficients

Three asymptotic regions

The asymptotic analysis method for the scaling function ϕ_p also applies to the filter coefficients $h_p[n]$. Comparing (2.12) to (2.14), one only needs to replace G and $G^{(-1)}$ by g and $g^{(-1)} = \int_0^\omega g(\theta)\, d\theta$. A natural approximation to the impulse response $h_p[n]$ is $h_p^*[n]$

$$h_p^*[n] = \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{-ipg^{(-1)}(\omega)}\, e^{in\omega}\, d\omega. \tag{2.57}$$

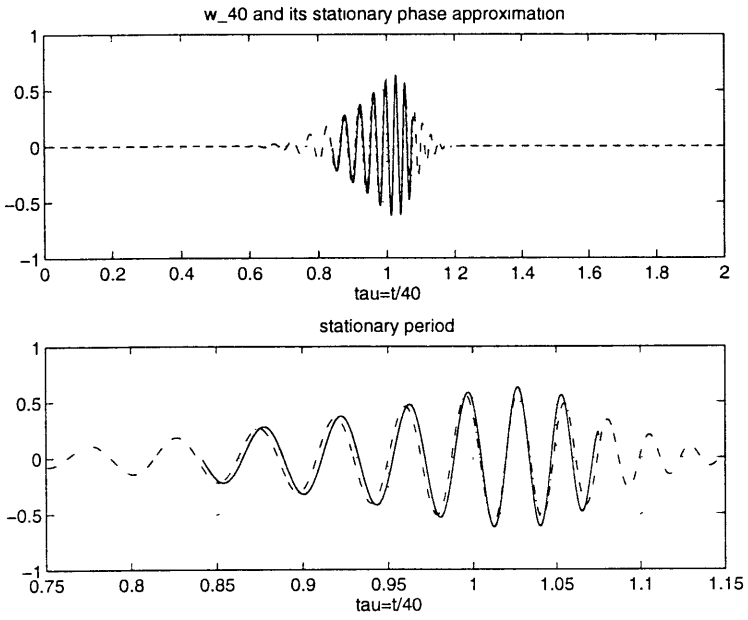

Figure 2-10: Accuracy of the stationary phase approximation for the wavelet $w_p(t)$. The dashed line represents w_{40} and the solid line is given by (2.54) with $p = 40$ and τ shifted by $0.01 = O(p^{-1})$.

We can extend the definition of h_p^* to allow non-integer index t by

$$h_p^*(t) = \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{-\imath p g^{(-1)}(\omega)} e^{\imath t \omega} \, d\omega \tag{2.58}$$

Scaling t by a factor p, $\tau = t/p$ yields

$$h_p^*(\tau) = \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{-\imath p [g^{(-1)}(\omega) - \tau\omega]} \, d\omega. \tag{2.59}$$

Define two filter transition times τ_0^h and τ_1^h by $\tau_0^h = g(0) = G(0) \simeq .1817$ and $\tau_1^h = g(\frac{\pi}{2}) = .5$. Then the asymptotic form of $h_p^*(\tau)$ is described by the following three results:

Result 7 (Stationary Phase) *Uniformly on any compact subset K of (τ_0^h, τ_1^h),*

$$h_p^*(\tau) = \sqrt{\frac{2}{\pi p g'(\omega_\tau)}} \cos[p(g^{(-1)}(\omega_\tau) - g(\omega_\tau)\omega_\tau) + \frac{\pi}{4}] + o(p^{-1/2}). \tag{2.60}$$

Here ω_τ is the unique $\omega \in (0, \frac{\pi}{2})$ such that $g(\omega) = \tau$.

Result 8 (Airy Transition) *For* $|\tau - \tau_0^h| \ll 1$,

$$h_p^*(\tau) = \sqrt[3]{\frac{6\pi}{p}}\,\mathrm{Ai}(-\sqrt[3]{6\pi p^2}(\tau - \tau_0^h)) + o(p^{-\frac{1}{3}}) \tag{2.61}$$

Result 9 (End Points) *Uniformly on any compact subset* K *of* $[\tau_0^h, \tau_1^h]^c$,

$$h_p^*(\tau) = \frac{1}{p\pi(\tau - \tau_1^h)}\,\sin[p(g^{(-1)}(\frac{\pi}{2}) - \tau\frac{\pi}{2})] + o(p^{-1}). \tag{2.62}$$

Of course, the Airy transition and the stationary phase are matched over the interval $p^{-2/3} \ll \tau - \tau_0 \ll 1$.

These results have the following interpretations:

1) **Stationary Phase.** The impulse response $h_p[n]$ in this period is energy significant. Its characteristic magnitude is $p^{-1/2}$. This period lasts approximately .3183 p from $n \simeq \tau_0^h\ p$ to $\tau_1^h\ p = p/2$. Its total energy is of order $O[(p^{-1/2})^2\ (\tau_1^h - \tau_0^h)p] = O[1]$.

2) **Airy Transition.** Though highest, it is energy insignificant. Its leading order is $O[p^{-1/3}]$ but its duration is only $O[p^{-2/3}] \cdot O[p] = O[p^{1/3}]$. Therefore its total energy is of order $O[(p^{-1/3})^2\, p^{1/3}] = O[p^{-1/3}]$.

3) **End Points.** The magnitude is no more than $O[p^{-1}]$. Therefore it is energy insignificant.

In one word, the energy of the impulse response $h_p[n]$ is asymptotically concentrated on the interval $n \in [\tau_0^h p,\ \tau_1^h p]$. Our analysis of these coefficients began with Nico Temme's plot of $h_{100}[n]$ in [72, 96], which is reproduced in Figure 2-11. In this case $p = 100$, and the leading order is $-\log_{10}\frac{1}{\sqrt{100}} = 1$. Readers can observe the stationary period up to $n = p/2 = 50$.

The similarity and difference between $\phi_p(t)$ and $h_p[n]$

A frequent conjecture is that *as* p *goes to* ∞, $h_p[n]$ *should look like* $\phi_p(t)$ *at integer times* $t = n$, $0 \le n \le 2p-1$. This is partly correct and partly wrong. We can summarize three essential points of similarity:

1) Both $h_p[n]$ and $\phi_p(t)$ have the same support interval $[0,\ 2p-1]$.

2) For large p, as n and t increase from 0 to $2p-1$, both $h_p[n]$ and $\phi_p(t)$ undergo the following three stages:

   1] Airy Transition (wavefront) with leading magnitude $O(p^{-1/3})$ and lasting time $O(p^{1/3})$.

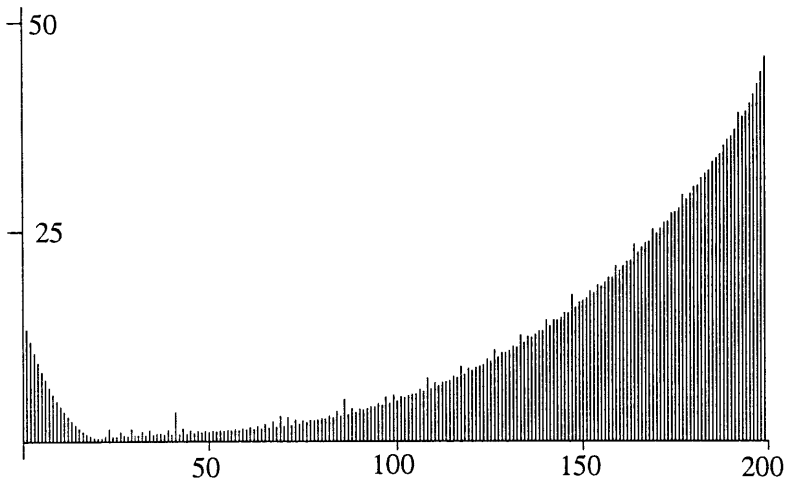

Figure 2-11: Plot of $-\log_{10}|h_p[n]|$ with $p = 100$, $0 \leq n \leq 2p-1 = 199$. The sharp Airy transition ends near $n = \tau_0^h p \simeq 18$, and the stationary phase extends to $n = 50$ with $-\log_{10}|h_p[n]| \simeq 1$. The long tail has small coefficients that are energy insignificant.

2] Stationary Phase (steady oscillation) with leading magnitude $O(p^{-1/2})$ and lasting time $O(p)$.

3] End Points with leading magnitude $O(p^{-1})$.

3) Both wavefronts start near time $p(1/2 - 1/\pi) \simeq .1817p$.

However, this structural similarity *does not* imply that the conjecture is entirely true. A significant difference also exists. *The Stationary Phase period of the filter impulse response $h_p[n]$ is much longer than that of the scaling function.* The scaling function stops near $t \simeq .3515p$, much earlier than the impulse response does (near $n \simeq .5p$). Therefore the impulse response cannot be the sampling of the scaling function.

Numerically the dilation equation is solved by the cascade algorithm, which iterates the lowpass filter (with time rescaling). A natural choice of the initial data is the impulse $\delta[n]$. Then the first iteration gives exactly the filter impulse response $h_p[n]$. Usually within 7 or 8 steps, one can obtain the scaling function with satisfactory accuracy.

From our results, we can see what is happening during this algorithm. After the first step, the values corresponding to the time interval $(.3515p, .5p)$ will be attenuated again and again until the leading magnitude falls from $O(p^{-1/2})$ to $O(p^{-1})$.

Chapter 3

Refinement Differential Equations and Wavelets

In this chapter, we consider the following type of Refinement Differential Equations(RDE)

$$P(D)\phi(x) = 2[H(E)\phi](2x),$$

where $P(\lambda)$ is a real polynomial and $H(z)$ is a real Laurent polynomial; $D = d/dx$ and E is the backward translation operator $Ef(x) = f(x+1)$. If the differential part is $P(\lambda) \equiv 1$, the equation is the famous refinement equation for designing scaling functions in wavelet theory. In this chapter, we reveal the general structure of solutions to RDE's and establish the relation between RDE's and certain types of refinement functional equations (RFE). This makes it possible to solve RDE's using the generalized subdivision scheme. The probability idea of Rvachev and Derfel is explored in a more systematic way. Our results are finally applied to the construction of *smoothed wavelets* and *quasi-multiresolution.*

3.1 Introduction

Let E and D denote the translation operator and derivative operator

$$f(x) \to E(f) = f(x+1), \qquad f(x) \to D(f) = f'(x).$$

Let $P(\lambda)$ be a polynomial in λ and $H(z)$ a Laurent polynomial in z^{-1} with real coefficients(following the literature of digital signal processing, we use z^{-1} instead of z)

$$P(\lambda) = c_N\lambda^N + c_{N-1}\lambda^{N-1} + \cdots + c_0, \qquad N \geq 0, \ c_N \neq 0,$$
$$H(z) = h_m z^{-m} + h_{m-1}z^{-m+1} + \cdots + h_{m-L}z^{-m+L}, \qquad h_m h_{m-L} \neq 0.$$

In this section, we consider equations of the following form

$$P(D)\phi(t) = 2[H(E)\phi](2t). \tag{3.1}$$

We call it a *refinement differential equation*(RDE) of type (P, H), with *order* N and *length* L.

If $H(z) = 0$, the RDE becomes an ordinary differential equation (linear homogeneous equation of order N). If, on the other hand, $P(\lambda) \equiv 1$, the RDE is called a refinement equation(for designing mother scaling functions with compact supports) in wavelet theory. Therefore in this paper, we shall assume $H \neq 0$ and $\deg P = n \geq 1$. The scalar 2 in the equation is not essential and can be replaced by any general integer of $k > 1$. It is kept here to follow the literature in wavelet theory and digital signal processing (See Daubechies [10, 1992] and Strang and Nguyen [69, 1996].)

RDE's of type (3.1) arise in many contexts. The first work should be mentioned is Mahler's remarkable paper [47, 1940] in 1939. Initiated by an integer partition problem in combinatorics, he studied the following functional equation

$$\frac{\phi(x+a) - \phi(x)}{a} = \phi(qx), \qquad a \neq 0, \ 0 < q < 1.$$

When the "difference parameter" a goes to zero, the equation evolves to an RDE: $\phi'(x) = \phi(qx)$ ($q \neq 2$, however.) He constructed a special solution through a very tricky integral transform. De Bruijn's work [14, 1953] on equations of the following type

$$\Phi'(s) = e^{\alpha s+\beta}\Phi(s-1)$$

gives a complete account of equations like $\phi'(x) = a\phi(qx)$ with $q \in (0, 1)$. The connection is realized

by the following change of variables (Kato and McLeod [42, 1971]):

$$x = e^s,\ q = e^c,\ \Phi(s) = \phi(x).$$

Near 1970, many authors(Fox and Mayers [28, 1971], Kato and McLeod [42, 1971], Frederickson [29, 1971]) studied the following special functional-differential equation:

$$\phi'(x) = a\phi(qx) + b\phi(x),$$

which had arisen from the mathematical modeling of an industrial problem involving wave motion in the overhead supply line for high speed train. It was Kato and McLeod who gave the complete investigation on this equation for all types of parameters. As an initial value problem, they showed the equation is well-posed if $q < 1$ and ill-posed if $q > 1$ (see also next section). Particularly, when $q > 1$, the solution to the IVP is not unique. The non-uniqueness made the equation less interesting to people working on ODE's, who care much about the existence and uniqueness of solutions to initial value problems. The importance of $q > 1$ was only realized recently when we abandoned the ODE sunglass and turned to the modern concept of *multiresolution* in wavelet theory.

The early 70's saw tons of papers (see Myshkis's survey paper [52, 1977]) on functional differential equations, most of which were initiated by applications in circuit and control theory. It was probably Rvachev who first deviated from the ODE point of view of the main stream. He paid attention to the functional properties of solutions to certain functional-differential equations. The equation initially studied by him is the following (see Rvachev's [62, 1990] [63, 1971]):

$$\phi'(t) = 2\phi(2t + 1) - 2\phi(2t - 1). \tag{3.2}$$

In this paper, we call it *the Rvachev equation.* The unique solution with unit total integral is denoted by $\mathrm{up}(t)$. It has the following properties: even, non-negative, C^∞, and with compact support $[-1,\ 1]$. The significance of $\mathrm{up}(t)$ in the function-theoretic sense is that it plays an atomic role in certain spaces consisting of C^∞ functions(Rvachev [62, 1990]), parallel to what the "mother" scaling function does in $L_p(R)$ nowadays. The connection between Rvachev's up function and current wavelet theory was explained by the work of Derfel, Dyn and Levin [17, 1995] on Stieltjes subdivision scheme and non-stationary subdivision process.

Their work on this part can be summarized as follows: the classical subdivision process for the Bernoulli (two-point) random distribution leads to the Haar scaling function, and then the continuous subdivision process associated to Haar scaling function generates Rvachev's up function. Part of

our current work is to reveal the general principles hidden in this intriguing example (section 3.5). We show that generally, for each refinement differential equation of form (3.1), one can associated it with a *refinement functional equation*(RFE) of the following form:

$$\phi(x) = \langle T\,,\, \phi(2x - \cdot)\rangle,$$

where T is a suitable distribution (typically, Schwartz distribution.) If the distribution has a nice "density" function, then the above RFE can be solved using the continuous subdivision process (section 3.6). This opens an entirely new window (compared with the classical ODE method for functional-differential equations) for RDE's. It is one of our initial goals to find the bridge between wavelet theory and the theory of functional differential equations.

If, under certain circumstance, the associated distribution T is a probability distribution, then probability method can be applied to the underlying RDE. For the up-function, this has already been noticed by Rvachev and Derfel, and can be further traced back to Jessen and Wintner's work in 1935 on infinite convolutions of symmetric Bernoulli distributions [39, 1935]. Related work can also be found in Erdös [24, 1939] [25, 1940] and Garsia [34, 1962], and Brown and Moran [6, 1973] (as initially referred in Daubechies [11, 1991].) In section 3.6, we develop the probability interpretation of certain basic RDE's (particularly for the Rvachev equation and the kam equation in section 3.2). The connection between the exponential distribution and the kam equation is entirely new, and provides an efficient approach to construct uniform approximation to the kam function. We also explain why the normal distribution cannot generate new functions along this probability line.

This chapter bas been organized as follows. In section 3.2, we introduce some necessary concepts and make some general assumptions on our work. The main result of the section is the Structure Theorem (Theorem 1), which is not difficult to prove but plays a crucial role in determining the presentation structure. In section 3.3, we present the main results on RDE's of type $(P(\lambda), 1)$. A new function $\mathrm{kam}(x)$ and one new family of functions $\Phi_\theta(x)$ are introduced as the atomic solutions to RDE's of type $(P(\lambda), 1)$ when $P(\lambda)$ contains no purely imaginary roots. For $P(\lambda)$'s possessing at least one purely imaginary root, we construct a set of linearly independent, periodic and C^∞ solutions based on the well adapted structure of Dirichlet series. As a by-product, we also show that RDE's of type $(P(\lambda), 1)$ always carry a single-parameter family of C^∞ and *almost periodic* solutions. In section 3.4, we construct the solutions to general RDE's. The special role of Rvachev's up-function comes up naturally. The main result of the section is that a scaling function designed through the $(1, H(z))$ refinement equation is smoothed by the $(P(\lambda), 1)$ RDE to yield a C^∞ solution to the original $(P(\lambda),\ H(\lambda))$ RDE. Section 3.5 and 3.6 are devoted to developing the connections

among RDE, RFE, the probabilistic method and the generalized subdivision process, as already introduced above. In the last section, we demonstrate one important application of our results in wavelet theory, namely, the construction of smoothed wavelets and quasi-multiresolution.

3.2 Regular Equations and the Structure Theorem

To formulate a well-posed problem, let us first understand the major difference between RDE's of form (3.1) and linear homogeneous ODE's. We illustrate it by considering the initial value problem

$$\phi'(t) + \phi(t) = q\phi(qt), \qquad \phi(0) = 1.$$

First we assume $q \in (0,\ 1)$. Integration gives:

$$\phi(t) = 1 - \int_{qt}^{t} \phi(\tau)\, d\tau = R_q\phi(t).$$

It is clear that the affine operator R_q is contracting if restricted to $C[0, \rho]$ for any $\rho < 1$. Hence the local existence and uniqueness follow immediately from the Contraction Mapping Theorem. Furthermore, the local solution can be obtained from the iterative action of R_q on any initial continuous function. Hence for $q < 1$, the equation is integrally no different from ordinary differential equations.

But it is not the case when $q > 1$. The integration gives

$$\phi(t) = 1 + \int_{t}^{qt} \phi(\tau)\, d\tau = R_q\phi(t).$$

R_q is not contracting any more since the domain is expanding. Expressed in the language of signal processing, the differential system is *non-causal*: what occurs at time t is influenced by the "future" events up to time qt since $q > 1$! This causes the ill-posedness.

However, when $q > 1$, the equation is *backward* well-posed in the following sense. Suppose we already know a "future" segment of the solution

$$\phi(t), \qquad t > q^n, \qquad \text{for some integer } n.$$

Then by solving iteratively the following inhomogeneous equations:

$$\phi'(t) + \phi(t) = f_m(t) = q\phi(qt), \qquad q^{m-1} < t \le q^m, \qquad m \le n,$$

the "history" of $\phi(t) : 0 < t \le q^n$ can be reconstructed! In this way, we find an easy way to construct

infinitely many solutions to the given equation on $(0, \infty)$. Let $k(x) \in C^\infty[1, q]$ satisfy the following compatible condition:

$$k^{(m+1)}(1) + k^{(m)}(1) = q^{m+1}k^{(m)}(q),$$

(for example, any function in $C_0^\infty[1, q]$.) Define $\phi(t) = k(t)$ for $t \in [1, q]$. By the above discussion, $\phi(t)$ on $(0, 1)$ can be determined uniquely; and for $t > q$, $\phi(t)$ is given by the following explicit iteration:

$$\phi(qt) = \frac{1}{q}[\phi'(t) + \phi(t)], \qquad q^n \leq t \leq q^{n+1}, n = 0, 1, \cdots .$$

This example makes it clear that the ODE point of view for RDE's of form (3.1) is inappropriate. In fact, as shown in section 3.6, the subdivision scheme (linear operator) "S" is much better fitted to this case than the integral operator R_q.

Therefore RDE's of form (3.1) are better viewed as a functional equation in certain function space. To ensure existence and uniqueness, the equation must satisfy some solvable conditions, just as the refinement equation does in wavelet theory (Daubechies [10, 1992]). We impose the following conditions.

Definition 1 *An RDE of type $(P(\lambda), H(z))$ is said to be regular if*

$$P(\lambda) = (\frac{\lambda}{2})^r p(\lambda), \qquad H(z) = (1 - z^{-1})^r h(z)$$

for some non-negative integer r and

$$p(0) = h(1) \neq 0.$$

r is called the index *of the equation.*

In this thesis, we only consider regular equations. This type already includes most functionally interesting examples such as the Rvachev equation and kam equation. By a non-trivial solution, we mean, in the most relaxed form, a non-zero C^N function $\phi(x)$ such that Eq. (3.1) is satisfied pointwise. Here N is the order of the equation. It is easily seen that such a solution is necessarily to be C^∞. However, unlike the homogeneous ODE with constant coefficients, it is not C^ω. The smoothness condition in no way ensures the uniqueness of the solution. Hence extra constraints are added. In this paper, we are particularly interested in the following two properties: 1) L_1, and 2) periodicity or almost periodicity (see section 3.3 for definition). If the solution is in $L_1(R)$, we secure

the uniqueness by forcing the integral normalization condition

$$\langle 1\,,\,\phi\rangle = \int_R \phi(x)\,dx = C$$

for some convenient non-zero constant C.

There are two alternative methods for studying regular RDE's, namely, the "frequency" domain method and the "time" domain method (Daubechies [10, 1992].) The first technique is particularly powerful for analyzing the regularity behaviors of the solution, as now widely known and practised in wavelet theory. In this paper, however, we prefer the time domain method because on one hand, unlike wavelets and scaling functions in wavelet theory, solutions to regular RDE's are always C^∞ and hence the regularity analysis is redundant; and on the other hand, the solutions to RDE's carry very rich structures and contents in the "time" domain. The time domain method is based on the convolution operator "$*$". Let us first recall some basic properties of the convolution operator.

(a) Suppose $f(t) \in C^\infty$, and $g(t) \in L^1$ and has compact support. Then $f * g \in C^\infty$.

(b) Suppose $f(t)$ is totally continuous and $g(t)$ as given above. Then

$$D(f * g) = (Df) * g, \qquad E(f * g) = f * (Eg).$$

(c) Suppose $f * g$ is well-defined. Then

$$2(f * g)(2t) = [2f(2t)] * [2g(2t)].$$

Theorem 12 (Structure Theorem) *Consider an RDE of type* $(P(\lambda),\ H(z))$. *Suppose*

$$P(\lambda) = P_1(\lambda) \cdot P_2(\lambda), \qquad H(z) = H_1(z) \cdot H_2(z),$$

where $P_i(\lambda)$ *are polynomials in* λ *and* $H_i(z)$ *Laurent polynomials in* z. *Assume* $\phi_i(x)$ *is the solution to the RDE of type* (P_i, H_i), $i = 1, 2$. *If* $\phi = \phi_1 * \phi_2$ *is well-defined, then it is a solution to the RDE of type* $(P(\lambda), H(z))$.

Proof. Let us check it directly by applying the fundamental properties of convolution.

$$\begin{aligned} P(D)\phi &= P_1(D)\cdot P_2(D)[\phi_1 * \phi_2] \\ &= [P_1(D)\phi_1] * [P_2(D)\phi_2] \\ &= 2[H_1(E)\phi_1](2x) * 2[H_2(E)\phi_2](2x) \\ &= 2[H_1(E)\phi_1 * H_2(E)\phi_2](2x) \\ &= 2[H_1\cdot H_2(E)\phi_1 * \phi_2](2x) \\ &= 2[H(E)\phi](2x). \end{aligned}$$

Hence $\phi(x)$ is the solution to the RDE of type $(P(\lambda),\ H(z))$. □

Corollary 5 (Smoothed Scaling Function) *Suppose a regular equation of type $(P(\lambda), H(z))$ has index $r = 0$. If $\phi_p(x)$ is the solution to $(P(\lambda), 1)$, and $\phi_h(x)$ the solution (scaling function) to $(1, H(z))$ (the refinement equation), and $\phi(x) = \phi_p * \phi_s$ is well-defined, then $\phi(x)$ is the solution to $(P(\lambda), H(z))$.*

This corollary will be very useful in the construction of smoothed wavelets and quasi-multiresolution (see section 3.7).

In the following, we first consider regular equations of type $(P(\lambda), 1)$. The discussion on general regular equations is completed in section 3.4.

3.3 Regular RDEs of Type $(P(\lambda), 1)$: the kam equation

Any real polynomial $P(\lambda)$ with $P(0) = 1$ can be factorized into the following

$$\prod_i (1 + a_i\lambda) \prod_j d_j(\lambda),$$

where a_i's are non-zero real constants and $d_j(\lambda)$'s are irreducible quadratic polynomials with $d_j(0) = 1$. By the preceding proposition, to solve RDE of type $(P(\lambda),\ 1)$, it suffices to consider the following two types of equations:

$$a\phi'(x) + \phi(x) = 2\phi(2x), \tag{3.3}$$

$$a\phi''(x) + b\phi'(x) + \phi(x) = 2\phi(2x). \tag{3.4}$$

$a\phi'(x) + \phi(x) = 2\phi(2x),\ a \neq 0$

By a change of variable: $x \to at$, we can assume $a = 1$.

Kato and McLeod [42, 1971] studied the following initial value problem in full details in 1971:

$$y'(t) = ay(qt) + by(t), \qquad t > 0.$$

Since the idea of multiresolution was not yet on the stage of mathematical analysis at the time, their starting point was the theory of initial value problems in ordinary differential equations: existence, uniqueness, and asymptotic behaviors as $t \to \infty$.

From their results, we have

(i) the only solution to ($a = q = 2$ and $b = -1$)

$$y'(t) + y(t) = 2y(2t), \qquad t > 0$$

that decays faster than $O(t^{-1})$ as $t \to +\infty$ is a constant multiple of the following function (which in fact decays exponentially fast, and is named after the two authors in this paper)

$$\text{kam}(t) = e^{-t}\Big\{1 + \sum_{n=1}^{\infty}(-1)^n \frac{2^n \exp[-(2^n-1)t]}{(2-1)(2^2-1)\cdots(2^n-1)}\Big\}, \qquad t > 0. \tag{3.5}$$

(ii) any other solution is not in $L^1(0,\infty)$. The two authors constructed a family of solutions that have the exact order $O(t^{-1})$ in the ∞.

The Dirichlet series solution in Eq.(3.5) was also studied by Frederickson [29, 1971] at the same time. He considered general solutions of the following form parameterized by β:

$$\phi_\beta(t) = \sum_{n=-\infty}^{\infty} C_{n,\beta} \exp(\beta q^n t)$$

for equations with $q > 1$. For $q > 1$, it can be easily verified that $\beta = -1$ is the only parameter that ensures $c_n = 0$ for all $n < 0$ and hence $\phi_\beta(x) = o(x^{-1})$.

We therefore conclude that if there is a solution $\phi(x) \in C^1(R)$ to

$$\phi'(x) + \phi(x) = 2\phi(2x), \qquad \int_R \phi(x)\,dx = 1,$$

$\phi(x)\Big|_{x>0}$ must be a constant multiple of $\text{kam}(x)$. To make referring easier later in the paper, we call this equation *the* kam *equation.*

For convenience, we introduce the combinatorial notation $(n)_q$ for q–analog number (see Goldman

and Rota [36] [37, 1969]) defined by

$$(n)_q = \frac{q^n - 1}{q - 1} = 1 + q + \cdots q^{n-1},$$

and $(n)_q!$ for the q–analog factorial given by

$$(n)_q! = (1)_q(2)_q \cdots (n)_q.$$

Then kam can be rewritten as

$$\text{kam}(x) = \sum_{n \geq 0} \frac{(-2)^n}{(n)_2!} \exp(-2^n x),$$

where $(0)_q!$ is defined to be 1. Series in this form are called *Euler Series* in combinatorics, if x is treated as a parameter.

Theorem 13 (Properties of $\text{kam}(x)$**)** *Define* $\text{kam}(x) = 0$ *for all* $x \leq 0$. *Then*

(a) The alternating infinite series converge to $\text{kam}(x)$ *at a rate of* $O(2^{-\binom{n}{2}})$ *uniformly for all* $x \geq 0$.

(b) $\text{kam}(x) \in C^\infty(R)$.

(c) $\text{supp}[\text{kam}(x)] = [0, \infty)$, *and* $\text{kam}(x) > 0$ *for all* $x > 0$.

(d) $\int_R \text{kam}(x)\, dx = \exp_2(-1) \in (\frac{2}{7}, \frac{1}{3})$, *where* $\exp_q(x)$ *is the q-analog exponential function defined by*

$$\exp_q(x) = \sum_{n=0}^{\infty} \frac{x^n}{(n)_q!}.$$

Proof. Set

$$a_n(x) = \frac{2^n}{(n)_2!} \exp(-2^n x) = \frac{2^n \exp(-2^n x)}{(2-1)(2^2-1)\cdots(2^n-1)}, \qquad n = 1, 2, \cdots.$$

Then

$$\frac{a_{n+1}(x)}{a_n(x)} = \frac{2}{(2^{n+1}-1)\exp(2^n x)} < 1,$$

for all $n \geq 1$ and $x \geq 0$. Hence the alternating infinite series converge to $\text{kam}(x)$ at a rate of

$$a_n(0) = O(2^{-\binom{n}{2}})$$

for all $x \geq 0$.

Fix any positive integer k,

$$\begin{aligned} |a_n^{(k)}(x)| &= \frac{2^{(k+1)n}\exp(-2^n x)}{(2-1)(2^2-1)\cdots(2^n-1)} \\ &\leq \frac{2^{(k+1)n}}{(2-1)(2^2-1)\cdots(2^n-1)} \\ &= O(2^{-\binom{n}{2}+kn}). \end{aligned}$$

which is uniformly exponentially small for all $x > 0$. Therefore for any positive integer k, and $x > 0$,

$$\frac{d^k}{dx^k}\mathrm{kam}(x) = \sum_{n=0}^{\infty}(-1)^{n+k}\frac{2^{(k+1)n}\exp(-2^n x)}{(2-1)(2^2-1)\cdots(2^n-1)},$$

which implies $\mathrm{kam}(x)$ is C^∞ for all $x > 0$. To show that $\mathrm{kam}(x) \in C^\infty(R)$, It suffices to show that

$$\mathrm{kam}^{(k)}(0^+) = 0 \qquad \text{for all} \quad k = 0, 1, \cdots.$$

In fact,

$$\begin{aligned} (-1)^k\frac{d^k}{dx^k}\mathrm{kam}(0^+) &= 1 + \sum_{n=1}^{\infty}(-1)^n\frac{2^{(k+1)n}}{(2-1)(2^2-1)\cdots(2^n-1)} \\ &= 1 + \sum_{n=1}^{\infty}(-1)^n\frac{2^{kn}(2^n-1)+2^{kn}}{(2-1)(2^2-1)\cdots(2^n-1)} \\ &= -2^k(-1)^{k-1}\frac{d^{k-1}}{dx^{k-1}}\mathrm{kam}(0^+) + (-1)^{k-1}\frac{d^{k-1}}{dx^{k-1}}\mathrm{kam}(0^+) \\ &= (1-2^k)(-1)^{k-1}\frac{d^{k-1}}{dx^{k-1}}\mathrm{kam}(0^+). \end{aligned}$$

An inductive argument completes the proof of (b).

(c) is obvious since $(a_n(x))$ is a strictly decreasing positive sequence. The equality part of (d) can be obtained by integrating the infinite series term by term; and the rest is due to

$$\exp_2(-1) = 1 - 1 + \frac{1}{3} - \frac{1}{3\cdot 7} + \cdots.$$

□

From now on, we will keep using $\mathrm{kam}(x)$ to denote its zero-extended version.

Corollary 6 *When $x \gg 1$, the leading term of* $\mathrm{kam}(x)$ *is* $\exp(-x)$.

Corollary 7 *For any non-zero real constant a, the following first order RDE*

$$a\phi'(x) + \phi(x) = 2\phi(2x), \qquad \int_R \phi(x) = \exp_2(-1)$$

has the unique $C^1(R)$ solution

$$\phi_a(x) = \frac{1}{|a|}\mathrm{kam}(\frac{x}{a}).$$

Especially, $\mathrm{supp}\phi_a = \mathrm{sign}(a) \cdot [0,\ \infty)$.

Proposition 1 *Define*

$$K_+(x) = \int_0^x \mathrm{kam}(t)\,dt, \qquad K_-(x) = \int_x^\infty \mathrm{kam}(t)\,dt, \qquad x > 0.$$

Then

$$K_+(x) = \sum_{n=1}^{\infty} \mathrm{kam}(\frac{x}{2^n}), \qquad K_-(x) = \sum_{n=0}^{\infty} \mathrm{kam}(2^n x).$$

Proof. Take the second equation for example. Call the function on the right hand side $L(x)$. Then

$$\begin{aligned} L'(x) &= \sum_{n=0}^{\infty} 2^n \mathrm{kam}'(2^n x) \\ &= \sum_{n=0}^{\infty} 2^n [2\mathrm{kam}(2^{n+1}x) - \mathrm{kam}(2^n x)] \\ &= \sum_{n=1}^{\infty} 2^n \mathrm{kam}(2^n x) - \sum_{n=0}^{\infty} \mathrm{kam}(2^n x) \\ &= -\mathrm{kam}(x). \end{aligned}$$

Obviously, $L(x) \to 0$ as $x \to \infty$. Hence $L(x) = K_-(x)$. □

Notice that $K_-(0^+) \neq K_-(0)$.

$a\phi''(x) + 2b\phi'(x) + \phi(x) = 2\phi(2x)$, $a > b^2$.

By a suitable change of variable $x = \pm\sqrt{a}\,t$, we can assume $a = 1$ and $b \geq 0$. Hence it suffices to consider the following standard equation

$$\phi''(x) + 2\cos\theta\phi'(x) + \phi(x) = 2\phi(2x), \qquad \theta \in (0, \frac{\pi}{2}]. \tag{3.6}$$

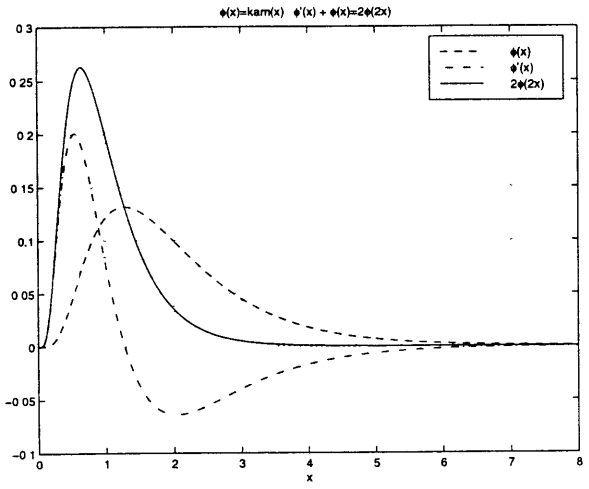

Figure 3-1: $\mathrm{kam}(x), \mathrm{kam}'(x)$ and $\mathrm{kam}(2x)$

To make the solution unique, the following integral normalization condition is applied

$$\int_R \phi(t)\, dt = [\exp_2(-1)]^2. \tag{3.7}$$

The solution is denoted by $\Phi_\theta(x)$.

For a given angle θ, define $\omega = \exp(i\theta)$, and

$$J^\theta(t) = 2\sum_{n=0}^{\infty} \frac{(-1)^m}{(m)_2!} \frac{1}{\bar{\omega} - t2^{-m}\omega} = J_r^\theta(t) + iJ_i^\theta(t),$$

for all $t \geq 0$. Here J_r^θ and J_i^θ are real and imaginary parts of J^θ.

Lemma 7 *Suppose* $0 < \theta \leq \pi/2$. *Then*

$$\sup_{t\geq 0} |J(t)| \leq 2\exp_2(1)\csc\theta.$$

Proof. For any $t \geq 0$,

$$\begin{aligned} |J(t)| &\leq 2\sum_{m\geq 0} \frac{1}{(m)_2!} \sup_{t\geq 0, m\geq 0} \frac{1}{|\bar{\omega} - t2^{-m}\omega|} \\ &\leq 2\exp_2(1) \sup_{t\geq 0} \frac{1}{|\bar{\omega} - t\omega|} \\ &\leq \frac{2\exp_2(1)}{\sin\theta}. \end{aligned}$$

□

Theorem 14 ($\Phi_\theta(x)$) $0 < \theta < \frac{\pi}{2}$. *Eq.(3.6) subjecting to the normalization condition Eq.(3.7) has the following unique solution:*

$$\Phi_\theta(x) = \begin{cases} \sum_{n=0}^{\infty} \frac{(-2)^n}{(n)_2!} \exp(-2^n bx) \left[J_r^\theta(2^n) \cos(2^n ax) + J_i^\theta(2^n) \sin(2^n ax) \right] & x > 0 \\ 0 & x \le 0 \end{cases} \tag{3.8}$$

Here $b = \cos\theta$ *and* $a = \sin\theta$. $\Phi_\theta(x) \in C^\infty(R)$; *and for* $x \ge 0$, *its derivatives can be obtained by differentiating the infinite series in Eq.(3.8) term by term, whose converging rate is* $O(2^{-\binom{n}{2}+kn})$ *for the* k*-th derivative.*

Proof. We only sketch the proof. Uniqueness can be proved by applying the technique of Kato and McLeod [42, 1971] or the frequency-domain method. The rest of the proof is constructed in the following three steps.

First, with the help of Lemma 7, we can show that $\Phi_\theta(x)$ is C^∞ for $x > 0$, and its derivatives can be obtained by taking differentiation on the infinite series term by term, and the derivative infinite series have the given converging rate (see the proof in the preceding theorem.)

Secondly, we show $\Phi_\theta(0^+) = \Phi'_\theta(0^+) = 0$. Take $\Phi'_\theta(0^+) = 0$ for example.

$$\begin{aligned} \Phi'_\theta(0^+) &= -\mathrm{Re} \sum_{n\ge 0} \frac{(-2)^n}{(n)_2!} 2^n \omega J^\theta(2^n) \\ &= -2\mathrm{Re} \sum_{m,n\ge 0} \frac{(-2)^{n+m}}{(n)_2!(m)_2!} \frac{2^n\omega}{2^m\bar{\omega} - 2^n\omega} \\ &= 2 \sum_{m,n\ge 0} \frac{(-2)^{n+m}}{(n)_2!(m)_2!} - 2\mathrm{Re} \sum_{m,n\ge 0} \frac{(-2)^{n+m}}{(n)_2!(m)_2!} \frac{2^m\bar{\omega}}{2^m\bar{\omega} - 2^n\omega} \\ &= -2\mathrm{Re} \sum_{m,n\ge 0} \frac{(-2)^{n+m}}{(n)_2!(m)_2!} \frac{2^n\bar{\omega}}{2^n\bar{\omega} - 2^m\omega} \\ &= 2\mathrm{Re} \sum_{m,n\ge 0} \frac{(-2)^{n+m}}{(n)_2!(m)_2!} \frac{2^n\omega}{2^m\bar{\omega} - 2^n\omega}. \end{aligned}$$

Comparison of the second line with the last line verifies that $\Phi'_\theta(0^+) = 0$.

In the final step, we show that for $x > 0$, $\Phi_\theta(x)$ is the solution to the given RDE. We ask readers to fill in the proof.

The combination of the last two steps and the original RDE implies that $\Phi_\theta^{(k)}(0^+) = 0$ for all non-negative integer k. Hence $\Phi_\theta(x) \in C^\infty(R)$ and satisfies the given RDE. Finally, a direct computation shows that Φ satisfies the prescribed integral normalization condition. This completes the proof. □

Corollary 8 (Damped Oscillation) *If* $0 < \theta < \pi/2$, *or equivalently* $b > 0$, *the leading term of*

the solution to Eq.(3.6) and Eq.(3.7) for $x \gg 1$ is the following damped oscillation:

$$\Phi_\theta(x) \simeq e^{-bx}[A_\theta \cos(ax) + B_\theta \sin(ax)],$$

where, $b = \cos\theta, a = \sin\theta$, and real constants A_θ and B_θ are given by

$$A_\theta + iB_\theta = 2 \sum_{m \geq 0} \frac{(-1)^m}{(m)_2!} \frac{1}{\bar{\omega} - 2^{-m}\omega}.$$

Notice the leading term is a special solution to the second order ordinary differential equation defined by the left hand side of the RDE.

Corollary 9 (Periodic Solution for $\theta = \frac{\pi}{2}$) *Suppose $\theta = \frac{\pi}{2}$. Then $\Phi_{\frac{\pi}{2}}(x)$ defined in Eq.(3.8) is forward periodic with period 2π when $x > 0$; that is,*

$$\Phi_\theta(x) = \Phi_\theta(x + 2\pi), \qquad x > 0.$$

$\Phi_{\frac{\pi}{2}}(x)$ is a special solution to Eq.(3.6).

Remark. It is not difficult to see that both $\Phi_{\frac{\pi}{2}}(x)$ and $\Phi_{\frac{\pi}{2}}(-x)$ are solutions to Eq.(3.6). Hence the solution space is at least 2-dimensional. It can be shown further that in this extremal case, Eq.(3.6) has no solution in $C^1(R) \cap L_1(R)$. We will discuss this case in more details in the coming subsection.

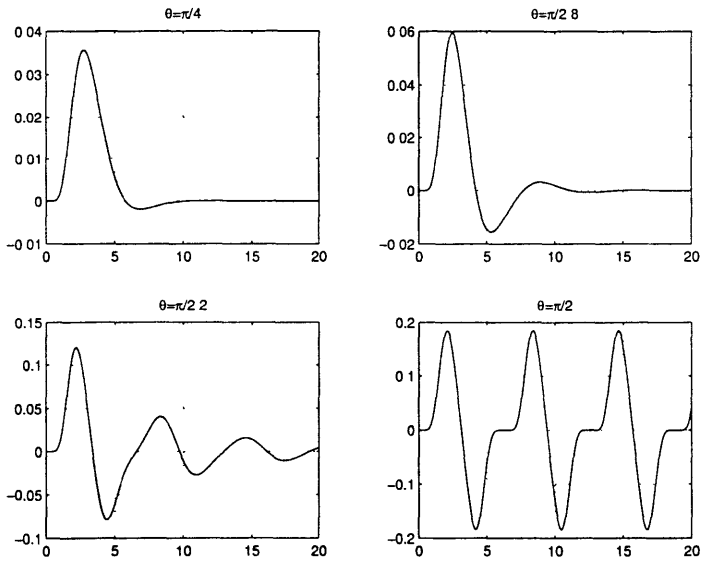

Figure 3-2: $\Phi_\theta(x)$ for different angles: Damped Oscillation ($\theta < \frac{\pi}{2}$) and Periodicity ($\theta = \frac{\pi}{2}$.)

For convenience, we extend the range of the parameter θ in Φ_θ to include $\theta \in (\pi/2,\ \pi)$ by the following

$$\Phi_\theta(x) = \Phi_{\pi-\theta}(-x), \qquad \theta \in (\pi/2,\ \pi).$$

Corollary 10 *Suppose real quadratic polynomial $a\lambda^2 + b\lambda + 1$ has two complex roots $re^{\pm i\theta}$ for some $r > 0$ and $\theta \in (0,\ \pi)\backslash\frac{\pi}{2}$. Then the unique $C^2 \cap L_1$ solution to the following RDE*

$$a\phi''(x) + b\phi(x) + \phi(x) = 2\phi(2x), \qquad \int_R \phi(x)dx = 1$$

is $\Phi_\theta(-rx)$ up to a multiplicative constant.

Finally, by recalling the Structure Theorem, we achieve the main result of this section.

Theorem 15 (Main Theorem. Part I) *Consider an RDE of type $(P(\lambda),\ 1)$, $P(0) = 1$. Suppose $P(\lambda)$ has no root along the imaginary axis. Then Eq.(3.6) has a $C^\infty(R)$ solution which decays exponentially fast near $\pm\infty$. In fact, up to a multiple constant, the solution takes the following form*

$$\phi(x) = \mathrm{kam}(-\lambda_1 x) * \cdots * \mathrm{kam}(-\lambda_k x) * \Phi_{\theta_1}(-r_1 x) * \cdots * \Phi_{\theta_m}(-r_m x).$$

Here, $\lambda_1, \cdots, \lambda_k$ are all real roots of $P(\lambda)$; and $r_j \exp(\pm\theta_j)$, $r_j > 0, \theta_j \in (0,\ \pi)$, $j = 1, \cdots, m$, are all complex roots.

Dirichlet Series Solution and Periodicity

Generally, for a given regular RDE of type $(P(\lambda),\ 1)$, $P(0) = 1$,

$$P(D)\phi(x) = 2\phi(2x),$$

one can always try the following Dirichlet series solution

$$\phi_\beta(x) = \sum_n C_{n,\beta} e^{-2^n \beta x},$$

where the index n runs through all integers and β is a parameter to be discussed below. The necessary condition for such a function $\phi(x)$ to be a solution is the following recursion formula for the coefficients:

$$C_{n,\beta} P(-2^n\beta) = 2C_{n-1,\beta}. \tag{3.9}$$

First, let us assume the parameter β is chosen so that $\lambda = -2^n\beta, n = 0, \pm 1, \cdots$ are all not roots of $P(\lambda)$. Then for any $n > 0$,

$$C_{n,\beta} = \frac{2}{P(-2^n\beta)} C_{n-1,\beta},$$

and $n < 0$,

$$C_{n,\beta} = \frac{1}{2} P(-2^{n+1}\beta) C_{n+1,\beta}.$$

In such a way, we find the unique Dirichlet series "solution" (up to a constant) by choosing $C_{0,\beta} = 1$.

However, we have to check the convergence properties of the resulted series and justify the above differentiation term by term. This is done by the following estimation. Suppose $P(\lambda)$ is of order $N \geq 1$. Then the above recurrence formula implies that

$$\frac{C_{n,\beta}}{C_{n-1,\beta}} = O(2^{-nN}), \qquad n \gg 1,$$
$$\frac{C_{n,\beta}}{C_{n+1,\beta}} \simeq \frac{1}{2}, \qquad n \ll -1,$$

Hence,

$$C_{n,\beta} = O(2^{-\binom{n}{2}N}), n > 0 \qquad \text{and} \qquad C_{n,\beta} = O(2^n), n < 0.$$

This estimation leads to the following.

Definition 2 (Almost Periodicity) *A function $f(x)$ is said to be* almost periodic *if for any $\epsilon > 0$, there exists a periodic function $f_\epsilon(x)$, such that $\|f - f_\epsilon\| < \epsilon$.*

Proposition 2 (Almost Periodic Solution) *For any RDE of type $(P(\lambda), 1)$, and any purely imaginary parameter β, if $P(\lambda)$ has no root in the form of $-2^n\beta$ for certain integer n, then the real and imaginary parts of $\phi_\beta(x)$ with $C_{0,\beta} = 1$ are two linearly independent and $C^\infty(R)$ real solutions to the RDE. Furthermore, they are both almost periodic.*

Now let's consider the case when there exists an integer n such that $P(-2^n\beta) = 0$. Let n_0 be the largest integer satisfying this condition. By choosing $\beta' = 2^{n_0}\beta$, we can assume $n_0 = 0$. From the recursion formula, we have

$$C_{n,\beta} = 0, \ n < 0; \qquad C_{n,\beta} = \frac{2}{P(-2^n\beta)} C_{n-1,\beta}, \ n > 0.$$

Hence by choosing $C_{0,\beta} = 1$, we obtain the unique (complex) Dirichlet series solution:

$$\phi_\beta(x) = \sum_{n\geq 0} C_{n,\beta} e^{-2^n \beta x}.$$

If the real part of β is not zero, the above expression cannot give a global solution since the series diverge for one of the half-axes. It can be fixed, however, if there exists some x_0 in the convergent half-axis such that $\phi_\beta(x)$(or one of its real and imaginary parts) is infinitely vanishing at this point: $\phi_\beta^{(k)}(x_0) = 0$ for $k = 0, 1, \cdots$. This is indeed what has occurred to kam(x) and $\Phi_\theta(x)$ ($x_0 = 0$ for both of them). Otherwise, we cannot obtain a global C^1 solution from the Dirichlet series.

The case when β is purely imaginary leads to the following.

Proposition 3 (Periodic Solution) *Suppose $P(\lambda)(P(0) = 1)$ contains a purely imaginary root $\beta = i\omega$, for which no $-2^n\beta$, $n > 0$ is a root of $P(\lambda)$ any more, then the real and imaginary parts of $\phi_\beta(x)$ with $C_{0,\beta} = 1$ are two linearly independent, periodic and C^∞ real solutions to the RDE of type $(P(\lambda), 1)$.*

Definition 3 (Binary Degree) *Two complex numbers a and b are said to be* binary dependent *if a/b is a an integer power of 2. A set of complex numbers is said to be* binary independent *if no two distinct numbers are binary dependent. The* binary degree *of a finite set is the cardinality of the maximal subset that is binary independent.*

This concept and the above discussion lead to the second part of our Main Theorem.

Theorem 16 (Main Theorem. Part II) *Given a real polynomial $P(\lambda)$, $P(0) = 1$, let Z denote the set of all of its purely imaginary roots with positive imaginary parts. Suppose the binary degree of Z is $d > 0$. Then the RDE of type $(P(\lambda), 1)$ has at least $2d$ linearly independent, periodic and C^∞ solutions.*

Analytic domain

In the above, we have shown that both kam(x) and $\Phi_\theta(x), (\theta \in (0, \frac{\pi}{2}))$ belong to $C^\infty(R)$. Since both only supported in the positive half axis, they cannot by analytic. However, we have the following results.

Proposition 4 (Analytic Extension of kam(x)**)** *There is a unique analytic function $K(z)$ that is defined on the right half plane:* Re$z > 0$ *and continuous to the imaginary axis, such that its restriction on the positive half-axis is* kam(x).

Proof. In fact $K(z)$ is the following function

$$K(z) = \sum_{n=0}^{\infty} \frac{(-2)^n}{(n)_2!} \exp(-2^n z), \qquad \mathrm{Re} z > 0.$$

It is also not difficult to see that $K(z)$ is continuous to the imaginary axis. □

Proposition 5 (Analytic Extension of $\Phi_\theta(x)$) *Suppose $\theta \in (0, \frac{\pi}{2})$. There is a unique analytic function that is analytic inside the angular domain defined by*

$$-\frac{\pi}{2} + \theta < \mathrm{Arg} z < \frac{\pi}{2} - \theta,$$

and continuous to the boundary, such that its restriction on the positive half-axis is $\Phi_\theta(t)$.

Proof. In fact, this function must be given by the following infinite series (see Eq.(3.8))

$$\sum_{n=0}^{\infty} \frac{(-2)^n}{(n)_2!} \exp(-2^n bz) \left[J_r^\theta(2^n) \cos(2^n az) + J_i^\theta(2^n) \sin(2^n az) \right].$$

□

3.4 General Regular RDE's

In this section, we consider general regular RDE's of type $(P(\lambda), H(z))$ and with index r. Since the equation is regular, we can assume

$$P(\lambda) = (\frac{\lambda}{2})^r p(\lambda), \qquad H(z) = (1 - z^{-1})^r h(z),$$

with $p(0) = h(1) \neq 0$. W. l. o. g., assume $p(0) = h(1) = 1$.

Type $(\frac{\lambda}{2}, 1 - z^{-1})$ and Rvachev's up function

An RDE of type $(\frac{\lambda}{2}, 1 - z^{-1})$ has the following form

$$\frac{1}{2}\phi'(x) = 2\phi(2x) - 2\phi(2x - 1), \qquad \int_R \phi(x)\, dx = 1. \tag{3.10}$$

Set

$$y = 2x - 1, \qquad \bar{\phi}(y) = \frac{1}{2}\phi(x) = \frac{1}{2}\phi(\frac{y+1}{2}).$$

Then

$$\bar{\phi}'(y) = 2\bar{\phi}(2x+1) - 2\bar{\phi}(2x-1),$$

which is exactly the Rvachev equation! Therefore,

Proposition 6 *The unique solution to Eq.(3.10) is given by*

$$\mathrm{up}_+(x) = 2\mathrm{up}(2x-1).$$

The Fourier transform of up function first appeared in Jessen and Wintner's paper [39, 1935] in 1935 as an example of infinite convolutions of symmetric Bernoulli distributions. Later in 1971 Rvachev's [63, 1971] studied Eq. (3.2) and obtained up(x) as a solution. Since then, it was re-discovered by many other authors in different contexts (see Kirov and Totkov [43, 1982], de Reina Martinez [15, 1982] for examples,) and its roles in approximation theory and in the representation of smooth functions have been studied extensively. Readers can find more references from Myshkis's survey paper [52, 1977] in 1977 and Rvachev [62, 1990] in 1990.

Integrating both sides of the Rvachev equation

$$\phi'(x) = 2\phi(2x+1) - 2\phi(2x-1),$$

with the assumption that $\phi(x) \in C^1(R) \cap L_1(R)$, we derive the following integral equation:

$$\phi(x) = \int_{2x-1}^{2x+1} \phi(t)\, dt. \tag{3.11}$$

Define the Rvachev operator R as follows

$$Rf = \int_{2x-1}^{2x+1} f(t)\, dt.$$

Restrict Dom$R$ = $L_1(R)$. R has the following three properties: First, the subspace of all $L_1(R)$ functions that are supported in $[-1,\ 1]$ is invariant under R-action; secondly, $\int_{-\infty}^{\infty} f(t)dt$ is conserved by R; and finally, R improves smoothness by order 1.

Rvachev obtained the up function by iterating R on the initial candidate $\phi_0(x) = \frac{1}{2}\mathbf{1}_{[-1,1]}(x)$. That is, define

$$\phi_n = R^n \phi_0,$$

for all $n = 0, 1, \cdots$. He showed that the (spline) sequence $\phi_n(x)$ converges uniformly. The limit is called the up function, which is C^∞, supported in $[-1,\ 1]$, and positive on $(-1,\ 1)$.

We list some other functional properties of $\mathrm{up}(x)$:

(1) $\mathrm{up}(x) + \mathrm{up}(x-1) = 1$ for $x \in [0, 1]$. This follows directly from taking the first derivative. Particularly,

$$\sum_{n=-\infty}^{\infty} \mathrm{up}(x-n) = 1.$$

(2) For any non-negative integer j, $\mathrm{up}^{(j)}(x)$ is a linear combination of the translated and dilated copies $\mathrm{up}(2^j x + k)$ (By induction.)

From the current multiresolution point of view, $\mathrm{up}^{(j)}(x)$ is inside $V_j(\mathrm{up})$, the space spanned (and closed) by all functions $\mathrm{up}(2^j x - k)$ for $k = 0, \pm 1, \cdots$.

Type $(1, H(z))$ and the scaling function

A regular RDE of type $(1, H(z))$ $(H(1) = 1)$

$$\phi(x) = 2 \sum_{k=m-L}^{m} h_k \phi(2x - k)$$

is the famous refinement equation in wavelet theory and computer aided design. Many authors have contributed to the discovery of its importance and the study of its solution behaviors. More references can be found in Daubechies and Lagarias [11, 12, 1991], and Daubechies [10, 1992]. In the following, we summarize some main results on this equation.

Deslauriers and Dubuc [19, 1989] showed that the equation is always solvable in the distributional sense (the solution therefore is called *the scaling distribution* in the following.) The non-trivial distribution solution is supported in $[m - L,\ m]$. Mallat [48, 1989] considered the equation which satisfies the following "orthogonal condition"

$$H(z)H(z^{-1}) + H(-z)H(-z^{-1}) = 1.$$

He showed that a refinement equation with such an "orthogonal" real filter $H(z)$ has a non-trivial $L_2(R)$ solution. Daubechies [9, 1988] studied the equation when the filter $H(z)$ satisfies the following "strong lowpass condition"

$$H(z) = \left(\frac{1 + z^{-1}}{2}\right)^p L(z).$$

She established the following regularity result: if $\|L(z)\|_{S^1} < 2^{p-q}$ for some non-negative number q, then $\phi(x) \in C^q$. Here the norm refers to the supremum norm restricted on the unit circle $S^1 : |z| = 1$. More results on the existence and regularity properties of the solution can be found in Daubechies and Lagarias [11, 12, 1991].

Solutions to General Regular Equations

Combining all the results so far, we achieve the last part of the main theorem.

Theorem 17 (Main Theorem. Part III) *Given a regular equation $(P(\lambda), H(z))$ of order $N = r + n \geq 1$, length $L = r + l$, and index r,*

(a) If $P(\lambda)$ has no purely imaginary roots, then the equation has a unique (up to a multiplicative constant) $C^\infty \cap L_1$ solution of the following form:

$$\phi(x) = \mathrm{up}_+^{*r}(x) * K(x) * \Phi(x) * \phi_h(x),$$

where

$\mathrm{up}_+^{*r}(x)$: *the r-th convolutional power of* $\mathrm{up}_+(x)$,
$K(x)$: *convolutional product of some scaled* $\mathrm{kam}(x)$ *functions,*
$\Phi(x)$: *convolutional product of some scaled* $\Phi_\theta(x)$ *functions,*
$\phi_h(x)$: *certain scaling distribution with a compact support of length l.*

(b) If the set of all purely imaginary roots of $P(\lambda)$ with positive part has a binary degree d, then the equation has at least $2d$ linear independent, periodic and C^∞ real solutions of the following form:

$$\phi(x) = \mathrm{up}_+^{*r}(x) * \Phi_p(x) * \phi_h(x),$$

where $\Phi_p(x)$ is some periodic solution of trigonometric Dirichlet series.

3.5 Distributions and Refinement Functional Equations

The probability explanation of the *up* function can be found in Rvachev [62, 1990]. Derfel [16, 1989] generalized the refinement equation in wavelet theory by allowing arbitrary probability masks(filters). His further work with Dyn and Levin in [17, 1995] studies the convergence problem of the continuous and non-stationary subdivision process with more general Stieltjes masks, which gives further probability explanations to Rvachev's up function. Following this line, we will discuss the probability side

of some typical RDE's in the next section. In this section, we first develop a more general framework based on distribution theory, by which a close connection is found between regular RDE's and a class of refinement equations with abstract "distribution masks", or refinement functional equations (RFE), as called in this paper.

In what follows, We consider only the class $\mathcal{S}$ of Schwartz functions and the space $\mathcal{S}'$ of Schwartz distributions, though some of the theory applies to much larger class of distributions. For those not familiar with distribution theory, Strichartz's "guiding" book [70, 1993] is a friendly and encouraging source to start with.

Recall that a C^∞ function $\phi(x)$ belongs to $\mathcal{S}$ if for any non-negative integer k and N,

$$\sup_{x\in R} |\phi^{(k)}(x)|(1+|x|)^N < \infty.$$

Hence $\mathcal{S}$ is closed under the differentiating operator D. The space of all linear functionals on $\mathcal{S}$ is denoted by $\mathcal{S}'$. Elements in $\mathcal{S}'$ are usually denoted by capital letters $T, F, \cdots$, and called Schwartz distributions (or *tempered distributions*). The value $T(\phi)$ is conventionally denoted by the scalar product $\langle T\,,\,\phi\rangle$.

Given a Schwartz distribution T, let us consider the following refinement functional equation(RFE)

$$\phi(x) = \langle T\,,\,\phi(2x-\cdot)\rangle, \qquad \phi(x)\in\mathcal{S}. \tag{3.12}$$

Here T acts on the variable in the position of $\cdot$ and x plays the role of a parameter. Since $\mathcal{S}$ is invariant under translation ($\phi(t)\to\phi(t-a)$) and dilation ($\phi(t)\to\phi(at)$) for any non-zero constant a, the equation is well-defined.

All solutions to a given RFE is a linear subspace of $\mathcal{S}$, and as one will see, in most interesting cases the solution space is a line. Hence to make the solution unique, we usually add another scalar character for the solution, such as

$$\langle 1\,,\,\phi\rangle = \int_R \phi(t)\,dt = c,$$

for a specified constant $c\neq 0$.

RFE's and RDE's are connected through the following concepts.

Definition 4 (δ–train) *A* δ–train *is the following functional*

$$F = \sum_{n=-\infty}^{\infty} c_n \delta(x-n),$$

with the coefficients c_n *satisfying the* temper growth *condition: there exists an integer* K*, such that*

$$\sup_n \frac{|c_n|}{(1+|n|)^K} < \infty.$$

This condition makes any δ*–train a Schwartz distribution. If there are only finitely many* c_n *that does not vanish, we say the* δ*–train is compactly supported. The maximal non-negative integer* L *such that there exists* m*,* $c_m c_{m-L} \neq 0$*, is called the* length *of the train.*

The most famous δ–train is the Poisson train or the uniform train:

$$P = \sum_{n=-\infty}^{\infty} \delta(x-n),$$

which is closely related to the famous Poisson Summation Formula (see Strichartz [70, 1993]), and is also directly connected to Shannon's Sampling Theorem (See Oppenheim and Schafer [57, 1989]). The interesting class of δ–trains in wavelet theory are those with compact supports.

Definition 5 (δ-simple) *A Schwartz distribution* T *is said to be* δ–simple*, if it is the solution to the following distribution differential equation:*

$$a_N T^{(N)} + a_{N-1} T^{(N-1)} + \cdots + a_1 T' + a_0 T = F, \tag{3.13}$$

for some constants $a_k, k = 0, 1, \cdots, N$*,* $a_N \neq 0$*, and some* δ*–train* F*.* N *is called the* order *of* T*. If* F *is compactly supported and with length* L*,* T *is said to be so.*

It is not difficult to see the following properties of δ-simple distributions:

(a) T is δ-simple if and only if $T(-x)$ is.

(b) If T_1 and T_2 are δ-simple, and $T = T_1 * T_2$ is well-defined, then T is δ-simple.

(c) Suppose T is δ-simple. Then any finite sum of the following form is δ-simple too

$$\sum_k c_k T(x-k).$$

It is easily seen that a δ-simple distribution is a linear combination of integer translated copies of the fundamental solution (to the associated ordinary differential equation.)

Theorem 18 (Second Main Theorem) *Suppose T is a δ-simple Schwartz distribution of order N and length L, and $\phi(x) \in \mathcal{S}$ is the solution to RFE (3.12) with mask T. Then $\phi(x)$ is the solution to an RDE with the same order and length. Conversely, suppose a regular RDE of order N and length L has a non-trivial solution ϕ that belongs to the Schwartz class $\mathcal{S}$. Then $\phi(x)$ must be the solution to an RFE with a δ-simple distribution mask T of the same order and length.*

Proof. Suppose T satisfies

$$a_N T^{(N)} + \cdots + a_1 T' + a_0 T = c_m \delta(t-m) + \cdots + C_{m-L}\delta(t-m+L).$$

Apply both sides to the t-function $\phi(2x-t)$ parameterized by x, we have

$$\sum_{k=0}^{N} a_k \langle T^{(k)}\,,\, \phi(2x-t)\rangle = \sum_{l=m-L}^{m} c_l \phi(2x-l).$$

On the other hand, the left hand side of the above equation is

$$\begin{aligned}
\text{l.h.s} &= \sum_{k=0}^{N} a_k \langle T\,,\, (-1)^k \partial_t^{(k)} \phi(2x-t)\rangle \\
&= \sum_{k=0}^{N} a_k \langle T\,,\, \frac{\partial_x^{(k)}}{2^k}\phi(2x-t)\rangle \\
&= \sum_{k=0}^{N} \frac{a_k}{2^k} D^k \langle T\,,\, \phi(2x-t)\rangle,
\end{aligned}$$

where $D = d/dx$. Since $\phi(x)$ is the solution to the RFE with T, we obtain

$$\text{l.h.s} = \sum_{k=0}^{N} \frac{a_k}{2^k} D^k \phi(x).$$

Hence $\phi(x)$ is the solution to the RDE of the following type $(P(\lambda), H(\lambda))$

$$P(\lambda) = \sum_{k=0}^{N} a_k \left(\frac{\lambda}{2}\right)^k, \qquad H(z) = \frac{1}{2}\sum_{l=m-L}^{m} c_l z^{-l}, \tag{3.14}$$

which obviously has order N and length L.

The converse is proved in a similar way. □

EXAMPLES:

(i) Let T be the following δ–train:

$$T = 2c_0\delta(x) + 2c_1\delta(x-1) + \cdots + 2c_L\delta(x-L).$$

Obviously, T is δ–simple (of order 0 and length L.) The corresponding RFE (3.12) gives directly the refinement equation in wavelet theory.

$$\phi(t) = 2c_0\phi(2t) + 2c_1\phi(2t-1)\cdots + 2c_L\phi(2t-L).$$

(ii) Define

$$T_\alpha = \alpha\exp(-\alpha|x|), \qquad \text{for some positive constant } \alpha.$$

Obviously T is a Schwartz distribution. Moreover,

$$-\frac{T''}{\alpha^2} + T = 2\delta.$$

Hence T is a second order δ–simple distribution of length 0. If $\phi(x) \in \mathcal{S}$ satisfies

$$\phi(x) = \int_R \alpha e^{-\alpha|t|}\phi(2x-t)\,dt,$$

it must be the solution to the following second order RDE of length 0

$$-\frac{\phi''(x)}{4\alpha^2} + \phi(x) = 2\phi(2x).$$

(iii) Let T be the characteristic of interval $[-1,1]$:

$$\langle T\,,\,\phi(x)\rangle = \int_{-1}^{1}\phi(x)\,dx.$$

Then

$$T' = \delta(x+1) - \delta(x-1).$$

Hence the associated RDE is

$$\phi'(x) = 2\phi(2x+1) - 2\phi(2x-1),$$

which is exactly the Rvachev equation.

(iv) Set $T = \mathbf{1}_{x>0}(x) 2 \sin \frac{x}{2}$, which is the solution to the following distribution differential equation:

$$4T'' + T = 2\delta.$$

Hence the associated RDE is given by

$$\phi''(x) + \phi(x) = 2\phi(2x).$$

Remark. As an equation, RFE (3.12) is always well-defined for any Schwartz distribution. However, it may have no solutions in the Schwartz class $\mathcal{S}$. On the other hand, a regular RDE as a point-wise equation can always be solved as we have shown in the previous sections. For instance, in the last example, the associated RDE does have at least one non-trivial C^∞ solution which is periodic (see section 3.3). But a non-zero Schwartz function cannot be periodic.

The second main theorem builds the connection between regular RDE's and a special class of RFE's with δ–simple distribution masks. It provides a new way to interpret and solve generic regular RDE's. In the section below, we show that the two building-block functions-up(x) and kam(x) can be studied successfully in this way.

3.6 Probability Method and Continuous Subdivision Process

Following the discussion in the preceding section, we consider a special class of distribution masks – finite positive distribution masks, or equivalently, by Riesz's representation theorem, probability measure masks. The connection between Rvachev's up function and probability has been pointed out by Rvachev [62, 1990] and Derfel [16, 1989] and Derfel, Dyn and Levin [17, 1995]. In this section, we develop systematically the probabilistic method for RDE's, especially for the two "building block" equations – the Rvachev equation and the kam equation. Derfel's generalized subdivision process is applied to generic RDE's. By "generic equations", we mean regular RDE's of type $(P(\lambda), H(z))$ such that $P(\lambda)$ contains no imaginary roots.

3.6.1 Probability Method

Probability Interpretation of Certain RFE's

When T is a probability distribution of some random variable X, the RFE (3.12) can be rewritten

as

$$\phi(x) = E(2\phi(2x - X)). \tag{3.15}$$

Here E denotes the expectation operator(not the translation operator defined in the abstract.) The factor 2 right after E implies that we are taking $2d\mu$ for T, if $d\mu$ stands for the probability measure of X.

Given a random variable X, we associate it with an "X–averaging" operator $\mathcal{A}_X$, which transforms any random variable Y that is independent of X to a new random variable $\mathcal{A}_X(Y)$ defined by

$$\mathcal{A}_X(Y) = \frac{X+Y}{2}.$$

The "fixed point" of this operator is of the most interesting. A random variable Y independent of X is called a *fixed point* of $\mathcal{A}_X$ if $\mathcal{A}_X(Y)$ has the same distribution as Y.

By recursion, it is not difficult to show the following

Proposition 7 *Let $X, X_n, n = 1, 2, \cdots$ be a sequence of i.i.d. random variables on some probability space. If the following infinite series of random variables converge a.s. to a random variable Y_X*

$$Y_X = \sum_{n=1}^{\infty} \frac{X_n}{2^n},$$

Y_X is the unique(in the sense of distribution) fixed point for $\mathcal{A}_X$.

Qualitative relation between Y_X and X is given by the following lemma.

Lemma 8 *Suppose* $\mathrm{supp}X = [a, b]$ *and* $\mathrm{Prob}(c < X < d) > 0$ *for any c, d: $a < c < d < b$. Then Y_X has the same properties.*

Proof. It is not difficult to see $\mathrm{supp}Y_X \subset [a, b]$ from the infinite summation. Hence we only need to show that Y_X shares the second property. Set $A = \max(|a|, |b|)$. Then

$$\left| \sum_{n>N} \frac{X_n}{2^n} \right| \leq A2^{-N}, \qquad \text{a.s..}$$

Denote $A2^{-N}$ by δ. For any c, d: $a < c < d < b$,

$$\begin{aligned}
&\text{Prob}\Big[c < Y_X < d\Big] \\
\geq &\text{Prob}\Big[c + \delta < \sum_{n=1}^{N} \frac{X_n}{2^n} < d - \delta\Big] \\
\geq &\text{Prob}\Big[(c+\delta)(1-2^{-N}) < X_n < (d-\delta)(1-2^{-N}) : n = 1, 2, \cdots, N\Big].
\end{aligned}$$

Choose N large enough so that $a < c' < d' < b$, where $c' = (c+\delta)(1-2^{-N})$ and $d' = (d-\delta)(1-2^{-N})$. Then

$$\text{Prob}(c < Y_X < d) \geq [\text{Prob}(c' < X < d')]^N > 0.$$

□

By the standard truncation technique, a and/or b can be relaxed to ∞ provided that $E(|X|) < \infty$.

For any random variable Y independent of X, set $Z = \mathcal{A}_X(Y)$. Suppose Y has probability density function (p.d.f.) ρ^Y.

Proposition 8 *If the p.d.f. ρ^Z of Z exists, then*

$$\rho^Z(x) = E(2\rho^Y(2x - X)).$$

Proof. This is because that (a) $E(\rho^Y(x - X))$ is the p.d.f. of $X + Y$ whenever X and Y are independent; (b) $2\rho(2x)$ is the p.d.f. of $X/2$ if X has p.d.f. $\rho(x)$. □

Corollary 11 (Probability Meaning of RFE's) *If the fixed point Y_X of $\mathcal{A}_X$ exists and has p.d.f. $\rho(x)$, then $\phi(x) = \rho(x)$ is the solution to Eq. (3.15).*

Remark. In the above argument, we have left out some technical details about the regularity conditions on the random variables involved. For instance, in Proposition 7, it is not difficult to show, by applying the famous Kolmogrov's Three Series Theorem (see Prakasa Rao [61, 1986] for example), that if $E(|X|) < \infty$, the infinite series of Y_X do converge almost surely. We refer to Derfel, Dyn and Levin [17, 1995] for readers who want to know more on convergence and regularity conditions.

The Uniform Distribution and $\text{up}(x)$

Let $X_u \sim U[-1, 1]$ be a random variable uniformly distributed on $[-1, 1]$. The corresponding T

is denoted by T_u and given by

$$T_u = \mathbf{1}_{[-1,1]}(x),$$

the characteristic of interval $[-1,1]$. T_u is a δ–simple distribution of order 1 and length 1 since

$$T_u' = \delta(x+1) - \delta(x-1).$$

Hence the solution $\phi(x)$ to Eq.(3.15) satisfies the following RDE according to the previous section

$$\phi'(x) = 2\phi(2x+1) - 2\phi(2x-1),$$

which is the Rvachev Equation. Hence up to a multiplicative constant, the solution is $\mathrm{up}(x)$.

Corollary 12 $\mathrm{supp}[up] = [-1,1]$, *and* $\mathrm{up}(x) \geq 0$ *for all* $x \in [-1,1]$.

Proof. Let $X_n, n = 1, 2, \cdots$ be a sequence of i.i.d. random variables of type X_u. Define

$$Y_u = \sum_{n=1}^{\infty} \frac{X_n}{2^n}.$$

By the second Main Theorem and the integral normalization condition, $\mathrm{up}(x) = \rho^Y(x)$. The proof is completed by applying the preceding lemma to the pair (X_u, Y_u). □

The Exponential Distribution and $\mathrm{kam}(x)$

Let X_e be any random variable with mean 2 and exponentially distributed along the positive half-axis. The corresponding T is denoted by T_e and is given by

$$T_e = \exp(-\frac{x}{2})\mathbf{1}_{x\geq 0}.$$

T_e is a δ–simple distribution of first order since it satisfies

$$2T_e' + T_e = 2\delta(x).$$

Therefore the solution $\phi(x)$ to the X-associated RFE equation must be also the solution to the following RDE

$$\phi'(x) + \phi(x) = 2\phi(2x).$$

Up to a multiplicative constant, the only solution is $\mathrm{kam}(x)$.

Corollary 13 $\mathrm{supp}[\mathrm{kam}] = [0, \infty)$, *and* $\mathrm{kam}(x) \geq 0$ *for all* $x > 0$.

Proof. Let $X_n, n = 1, 2, \cdots$ be i.i.d. random variables of exponential type X_e on some probability space. Define

$$Y_e = \sum_{N \geq 1} \frac{X_n}{2^n}.$$

Then Y_e is well defined and its p.d.f is given by

$$\rho^Y(x) = \frac{\mathrm{kam}(x)}{\exp_2(-1)}.$$

A recall of the preceding lemma completes the proof. □

The Normal Distribution and Divisibility

So far, we have discussed two important continuous probability distributions. To be complete, it is natural to ask what is the function $\phi(x)$ that corresponds to the normal distribution. The answer is: normal distribution does not yield new function since it is *divisible*. Normal distribution serves as a famous "fixed point" in the Central Limit Theorem and Fourier transform. So it does here for the refinement process.

For normal distribution, we define the corresponding Schwartz distribution T_n to be

$$T_n = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right).$$

It satisfies the following well-known differential equation

$$T_n' + xT_n = 0,$$

from which it is obvious that T_n is not δ–simple. Hence one should not expect that the solution $\phi(x)$ to Eq.(3.15) can be a solution to an RDE.

On the other hand, let $X_n, n = 1, 2, \cdots$ be a sequence i.i.d. random variables of $N(0, 1)$, and define

$$Y = \sum_{n \geq 1} \frac{X_n}{2^n}.$$

Since normal distribution is divisible, i.e. *the sum of any two independent normal random variables*

is still of normal type, Y must be a normal random variable too! Since

$$E(Y) = \sum_{n\geq 1} \frac{E(X_n)}{2^n} = 0, \qquad \sigma^2(Y) = \sum_{n\geq 1} \frac{\sigma^2(X_n)}{4^n} = \frac{1}{3},$$

we conclude that the solution to Eq.(3.15) for normal distribution is (subject to the integral normalization condition $\langle 1\,,\,\phi\rangle = 1$)

$$\phi(x) = \sqrt{\frac{3}{2\pi}} \exp\left(-\frac{3x^2}{2}\right).$$

Notice the major difference between the solution in this case and those in the previous two cases: the solution here is C^ω!

3.6.2 Continuous Subdivision Scheme for Generic RDE's

The connection between RFE's and RDE's makes it possible to solve generic RDE's using the generalized (continuous) subdivision process (Derfel, Dyn and Levin [17, 1995]).

Let us first recall briefly the role of the subdivision scheme in wavelet theory. For a given refinement equation

$$\phi(x) = 2\sum_n h_n \phi(2x - n),$$

the associated subdivision scheme S is the following "refining operator"

$$(S\mathbf{f})[n] = 2\sum_k h_{n-2k}\mathbf{f}[k], \qquad n = 0, \pm 1, \cdots,$$

which maps an infinite sequence $\mathbf{f}[n]$ to another (refined) sequence $S\mathbf{f}$ (Daubechies [10, 1992].) The subdivision scheme is better viewed as a grid transfer function in Multigrid Method(see Briggs [5, 1987]). The subdivision process (SP) refers to the following iteration process starting with the "impulse signal" $\delta[n]$ (Strang and Nguyen [69, 1996]) :

$$\delta,\ S\delta,\ S^2\delta,\ \cdots.$$

It is said to be *convergent uniformly* if there exists a continuous function $\phi(x)$ such that

$$\lim_{j\to\infty} \|S^j\delta - \phi_j\|_\infty = 0,$$

where sequence ϕ_j is defined by $\phi_j[k] = \phi(k/2^j)$. Obviously if the SP converges, $\phi(x)$ must be unique. Indeed, in the case of finite length filter h_n, $\phi(x)$ is the unique solution to the refinement equation subjecting to the integral normalization condition $\int \phi(x)\, dx = 1$.

SP is very useful in computer graphics for generating continuous objects from discrete data(see Deslauriers and Dubuc [18, 1987] and Dyn and Levin [21, 1990]). It also appears in the analysis of the Picard-Lindelöf iteration in numerical computation of ODE systems (see Nevanlinna [54, 1990]).

For refinement equation with continuous mask, Derfel, Dyn and Levin [17, 1995] generalizes the subdivision process in a natural way. It can be used here to solve RDE's iteratively.

Given a refinement differential equation of type $(P(\lambda),\ H(z))$, suppose T is the δ-simple distribution associated to it. Assume T has a "density" function $\rho(x) \in L_1(R)$ (unnecessary to be non-negative). That is

$$\langle T\ ,\ g(x)\rangle = \int_R \rho(x)g(t)\, dx,$$

for any test function $g(x)$. We require $\int_R \rho(x)\, dx = 2$.

The continuous subdivision scheme is the following operator S_c (subscript c stands for "continuous"):

$$S_c f(x) = \int_R \rho(x - 2t)f(t)\, dt,$$

for any function f with at most a polynomial growth rate at infinity.

Let $\delta = \delta(x)$ be the delta distribution. The generalized SP is the following iteration:

$$\delta, S_c\delta, S_c^2\delta, \cdots .$$

It is said to converge *uniformly* to a continuous function $\phi(x)$, if

$$\lim_{j\to+\infty} \|S_c^j\delta(x) - \phi(2^{-j}x)\|_\infty = 0.$$

and converge *weekly* to a distribution F, if for any test function $g(x)$,

$$\lim_{j\to+\infty} \int S_c^j\delta(x)2^{-j}g(2^{-j}x)\, dx = \langle F\ ,\ g(x)\rangle.$$

The result of Derfel, Dyn and Levin [17, 1995] leads to the following.

Proposition 9 (Derfel, Dyn and Levin [17, 1995], Corollary 15 modified) *If $\rho(x)$ is rapidly*

decreasing, then the generalized SP converges in the weak sense to an infinitely differentiable function in $L_1(R)$, which is the unique solution of the corresponding RFE in $L_1(R)$ subjecting to the integral normalization condition. Furthermore, if $\rho(x)$ consists of finitely many smooth pieces, then the convergence is uniform in any $C^n(R)$.

This leads to an algorithmic approach to all regular generic RDE's.

We illustrate it through the following three examples. Notice that the SP also provides an efficient way to compute convolutions like $f * \phi$ for an arbitrary function $f(x)$ if $\phi(x)$ is an solution to certain RFE: one does not need to know the explicit expression of $\phi(x)$ and just starts the SP with the initial function f, instead of δ.

EXAMPLES:

(i) Rvachev equation: $\phi'(x) = 2\phi(2x+1) - 2\phi(2x-1)$.

In this case, $\rho_u(x) = \mathbf{1}_{[-1,1]}(x)$. The subdivision scheme S_u is given by

$$S_u f(x) = \int_R \rho_u(x-2t) f(t)\, dt = \int_{\frac{x-1}{2}}^{\frac{x+1}{2}} f(t)\, dt.$$

Hence

$$S_u\delta(x) = \rho_u(x) = \mathbf{1}_{[-1,1]}(x),$$

$$S_u^2\delta = \begin{cases} \frac{x+3}{2} & -3 \le x < -1 \\ 1 & -1 \le x < 1 \\ \frac{3-x}{2} & 1 \le x < 3 \\ 0 & \text{the rest} \end{cases}$$

$$\cdots$$

If we look at the rescaled (spline) functions $S_u\delta(2x)$, $S_u\delta(4x)$, $\cdots$, it is easy to observe the following properties (those invariant during the SP) of the limiting function $\mathrm{up}(x)$:

(a) $\mathrm{supp}[\mathrm{up}] = [-1,1]$ and $\mathrm{up}(x) > 0$ for all $x \in (-1,1)$;

(b) $\mathrm{up}(x)$ is infinitely flat at $x = \pm 1$ and $x = 0$, and $\mathrm{up}(0) = 1, \mathrm{up}(\pm 1) = 0$;

(c) $\mathrm{up}\big|_{[0,1]}$ is mirror symmetric around $\frac{1}{2}$: $\mathrm{up}(1-x) = 1 - \mathrm{up}(x)$.

(ii) The kam equation: $\phi'(x) + \phi(x) = 2\phi(2x)$.

Here we have $\rho_e(x) = \exp(-\frac{x}{2})\mathbf{1}_{x>0}$. The subdivision scheme S_e is defined by:

$$S_e f(x) = \int_R \rho_e(x-2t)f(t)\,dt = e^{-\frac{x}{2}} \int_{-\infty}^{\frac{x}{2}} e^t f(t)\,dt.$$

Particularly, if $\mathrm{supp} f \subset [0,\infty)$, then

$$S_e f(x) = \mathbf{1}_{x>0} e^{-\frac{x}{2}} \int_0^{\frac{x}{2}} e^t f(t)\,dt.$$

Hence the first few steps of the SP are given by

$$\begin{aligned} S_e\delta &= \rho_e = \exp(-\frac{x}{2})\mathbf{1}_{x>0}, \\ S_e^2\delta &= [2e^{-\frac{x}{4}} - 2e^{-\frac{x}{2}}]\mathbf{1}_{x>0}, \\ S_e^3\delta &= [\frac{8}{3}e^{-\frac{x}{8}} - 4e^{-\frac{x}{4}} + \frac{4}{3}e^{-\frac{x}{2}}]\mathbf{1}_{x>0}. \\ &\cdots \end{aligned}$$

Generally $S_e^k\delta$ can be obtained by solving a linear system in the following way. Assume

$$S_e^k\delta(x) = \mathbf{1}_{x>0}\sum_{j=1}^{k} c_j e^{-\frac{x}{2^j}}.$$

To determine k coefficients c_j, we impose the following k conditions

$$\int S_e^k\delta(2^k x)dx = 1; \qquad \frac{d^m}{dx^m}\Big|_{x=0} S_e^k\delta(x) = 0, m = 0,1,\cdots k-2.$$

This leads to the following linear system of Vandemonde type:

$$\sum_{j=1}^{k} c_j 2^j = 2^k; \qquad \sum_{j=1}^{k} c_j 2^{-jm} = 0, m = 0,1,\cdots,k-2.$$

Now we have two sequences of functions to approximate $\mathrm{up}(x)$: the scaled SP sequence $\exp_2(-1)S_e^k\delta(2^k x)$ and the k-th partial sum in Eq.(3.5)

$$S_k(x) = \sum_{m=0}^{k-1} \frac{(-2)^m}{(m)_2!} \exp(-2^m x).$$

The advantage of the SP sequence $\exp_2(-1)S_e^k\delta(2^k x)$ is that it gives a better uniform approximation. $S_k(x)$ is only good away from $x = 0$ (see Figure 3-3).

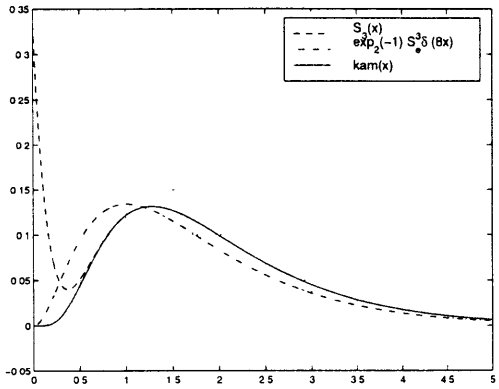

Figure 3-3: $\exp_2(-1)S_e^k\delta(2^kx)$ gives a better uniform approximation to $\mathrm{kam}(x)$ compared with the partial sum sequence $S_k(x)$. $k = 3$ in this plotting.

(iii) $\dfrac{\phi''(x)}{2} + \phi'(x) + \phi(x) = 2\phi(2x)$.

The "density" function for this equation is

$$\rho(x) = \mathbf{1}_{x>0}e^{-\frac{x}{2}}\sin\frac{x}{2}.$$

Hence the resulted subdivision scheme S for any function f supported in the positive half axis is

$$Sf(x) = \mathbf{1}_{x>0}e^{-\frac{x}{2}}\left[\sin\frac{x}{2}\int_0^{\frac{x}{2}} e^t\cos tf(t)dt - \cos\frac{x}{2}\int_0^{\frac{x}{2}} e^t\sin tf(t)dt\right].$$

It is guaranteed by the preceding proposition that $S^k\delta(2^kx), k = 0, 1, \cdots$ converges uniformly to $\sqrt{2}\Phi_{\frac{\pi}{4}}(\sqrt{2}x)$ up to a multiplicative constant.

3.7 Application: Smoothed Wavelets and Quasi-Multiresolution

In this section, we present one application of the previous results in wavelet theory—the construction of smoothed wavelets and quasi-multiresolution.

3.7.1 Classical Wavelets with Compact Support

Wavelets with compact supports are of particular interest in application. The design starts with the following refinement equation

$$\phi(x) = 2[H(E)\phi](2x) = 2\sum_{n=m-L}^{m} h_n\phi(2x-n). \tag{3.16}$$

Once the scaling function has been worked out, the associated (mother) wavelet $\psi(x)$ is obtained through the following wavelet equation:

$$\psi(x) = 2[G(E)\phi](2x) = 2\sum_{k=p}^{p-L} g_k \phi(2x-k). \tag{3.17}$$

Here

$$G(z) = \sum_{k=p}^{p-L} g_k z^{-k}$$

is the companion *highpass filter* of $H(z)$, which must satisfy the highpass condition: $G(1) = 0$.

The scaling function and wavelet generate a multiresolution(MR) in the following way: for any function $f(x) \in L_2(R)$ and each integer j, define a closed subspace by setting

$$V_j(f) = \overline{\text{span}\{f(2^j x - k) \mid k = 0, \pm 1, \cdots\}},$$

MR is the iteration of the following triangular relation in $L_2(R)$:

$$\begin{array}{ccc} V_{j-1}(\phi) & \longrightarrow & V_j(\phi) \\ & \nearrow & \\ \oplus & & \\ V_{j-1}(\psi) & & \end{array}$$

Here $\bigoplus$ is the direct sum of subspaces; and for orthogonal multiresolution, it is the orthogonal direct sum.

Under certain conditions on the two filters, the resulted MR is complete in the sense that $\cup_j V_j(\phi)$ is dense in $L_2(R)$. Further conditions on the filters lead to the orthogonality of the direct sum. One of the unsatisfactory point of classical compact wavelet analysis is the regularity: all scaling functions and wavelets designed through this way are only finitely many times differentiable. The regularity of scaling functions and wavelets is directly related to the vanishing degree of the lowpass filter $H(z)$ at the highest frequency $\omega = \pi$ if $z = e^{i\omega}$.

In certain circumstances, such as applying the Wavelet-Galerkin method to solve differential equations numerically, C^∞ (or piecewise C^∞) basis functions are welcomed (classically, trigonometric functions, splines, orthogonal polynomials and eigenfunctions of a Sturm-Liouville problem.) The naive idea to achieve smoothness is to mollify scaling functions and wavelets in existence. This is indeed what the following differentially perturbed refinement equation (a special RDE) achieves.

3.7.2 Smoothed Wavelets and Quasi-Multiresolution

For an existing system of refinement equation and wavelet equation whose filter pair is given by $(H(z), G(z))$, let us consider the following first order perturbed system:

$$\epsilon\phi_\epsilon'(x) + \phi_\epsilon(x) = 2[H(E)\phi_\epsilon](2x), \tag{3.18}$$

$$\epsilon\psi_\epsilon'(x) + \psi_\epsilon(x) = 2[G(E)\phi_\epsilon](2x), \tag{3.19}$$

where ϵ is a small perturbation parameter. Assume it to be positive. We impose again the normalization condition $\int \phi_\epsilon(x)dx = 1$. By the structure theorem, if the original unperturbed system has the pair of scaling function and wavelet ($\phi(x)$, $\psi(x)$), then the above system has the following solution:

$$\phi_\epsilon(x) = K_\epsilon(x) * \phi(x), \qquad \psi_\epsilon(x) = K_\epsilon(x) * \psi(x),$$

where $K_\epsilon(x)$ is given by

$$K_\epsilon(x) = \frac{\epsilon^{-1}\mathrm{kam}(\epsilon^{-1}x)}{\exp_2(-1)}.$$

The interesting observation is $K_\epsilon(x)$ plays the exact role of a *mollifier* in functional analysis. Hence it is readily seen that

(1) ϕ_ϵ and ψ_ϵ are both C^∞ functions.

(2) ϕ_ϵ and ψ_ϵ converge to ϕ and ψ in $C^\alpha(R)$ and $L_p(R)$ whenever $\phi \in C^\alpha(R)$.

(3) ϕ_ϵ and ψ_ϵ are "weakly" compactly supported, or equivalently, decay faster than any polynomial degree.

If ϵ is small enough, $V_j(\phi_\epsilon)$ and $V_j(\psi_\epsilon)$ are two subspaces very "close" to $V_j(\phi)$ and $V_j(\psi)$ in the sense that Shen and Strang gave in [65, 1996]. Thus it can be expected that the following triangular relation should still hold approximately:

$$\begin{array}{ccc} V_{j-1}(\phi_\epsilon) & \longrightarrow & V_j(\phi_\epsilon) \\ \oplus & \nearrow & \\ V_{j-1}(\psi_\epsilon) & & \end{array}$$

We call the iteration of this approximate triangular relation *a quasi-multiresolution*(QMR).

If the original unperturbed system is orthogonal, that is

$$\int_R \phi(x)\phi(x-n)\,dx = \delta_n, \qquad \int_R \phi(x)\psi(x-n)\,dx = 0$$
$$\int_R \psi(x)\psi(x-n)\,dx = \delta_n,$$

the perturbed system must satisfy the following relations:

$$\int_R \phi_\epsilon(x)\phi_\epsilon(x-n)\,dx = \delta_n + s_n, \qquad \int_R \phi_\epsilon(x)\psi_\epsilon(x-n)\,dx = r_n, \tag{3.20}$$

$$\int_R \psi_\epsilon(x)\psi_\epsilon(x-n)\,dx = \delta_n + w_n, \tag{3.21}$$

where sequences (s_n), (r_n), (w_n) are uniformly (for ϵ) exponentially small for large $|n|$ and the sequence supremum norms are of order $O(\epsilon)$ (in fact $O(\epsilon^2)$ as we show later). This is to say that the resulted QMR is *near orthogonal.*

For example, in Figure 3-4, we have plotted the smoothed Haar scaling function and Daubechies min-phase orthogonal scaling function D_4 (See Daubechies [10, 1992]), both obtained by choosing the perturbation parameter ϵ to be 0.04.

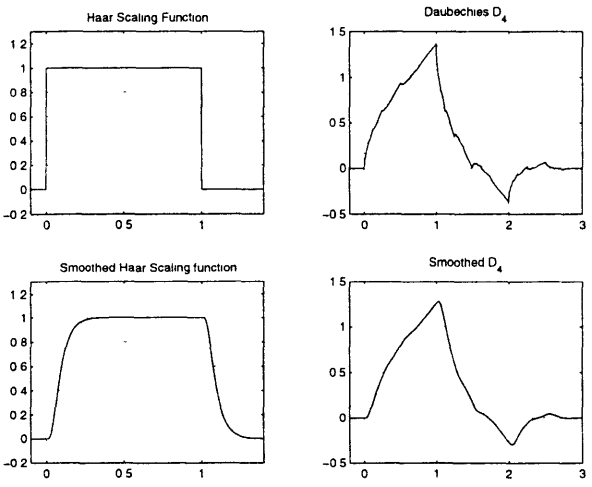

Figure 3-4: First order perturbed Haar scaling function and D_4 ($\epsilon = 0.04$).

3.7.3 Smoothing versus Small Deviation

In this section, we show that the solutions to the differentially perturbed system (3.18) and (3.19) is very close to a small deviation of the original scaling functions and wavelets.

Let us consider the following new system:

$$\bar{\phi}_\epsilon(x+\epsilon) = 2[H(E)\bar{\phi}_\epsilon](2x), \tag{3.22}$$

$$\bar{\psi}_\epsilon(x+\epsilon) = 2[G(E)\bar{\phi}_\epsilon](2x), \tag{3.23}$$

with the same integral normalization condition for the scaling function.

Proposition 10 (Small Deviation) *If $(\phi(x), \psi(x))$ is the scaling function and wavelet pair for the original system, then Eq.(3.22) and (3.23) have solutions*

$$\bar{\phi}_\epsilon(x) = \phi(x-2\epsilon), \qquad \bar{\psi}_\epsilon(x) = \psi(x-2\epsilon).$$

Proof. A direct check using the original refinement and wavelet equation. □

When the lowpass filter $H(z)$ is highly vanishing at $\omega = \pi$ or $z = -1$, $\phi(x)$ and $\psi(x)$ are finitely many times differentiable. So are $\bar{\phi}_\epsilon$ and $\bar{\psi}_\epsilon$ by the above proposition. We can apply Taylor expansion for parameter ϵ:

$$\bar{\phi}_\epsilon(x+\epsilon) = \bar{\phi}(x) + \epsilon\bar{\phi}'(x) + O(\epsilon^2) \qquad \bar{\psi}_\epsilon(x+\epsilon) = \bar{\psi}(x) + \epsilon\bar{\psi}'(x) + O(\epsilon^2).$$

Hence, up to the second order, we have the following

$$\epsilon\bar{\phi}'_\epsilon(x) + \bar{\phi}_\epsilon(x) \simeq 2[H(E)\bar{\phi}_\epsilon](2x),$$
$$\epsilon\bar{\psi}'_\epsilon(x) + \bar{\psi}_\epsilon(x) \simeq 2[G(E)\bar{\phi}_\epsilon](2x).$$

This leads to

Corollary 14 (Smoothing versus Small Deviation) *Suppose the original scaling function $\phi(x)$ is $C^\alpha(R)$ for some $\alpha \geq 2$. Then uniformly for all x,*

$$\phi_\epsilon(x) = \phi(x-2\epsilon) + O(\epsilon^2), \qquad \psi_\epsilon(x) = \psi(x-2\epsilon) + O(\epsilon^2).$$

Proof. We only sketch the proof. Set $\Delta_\epsilon(x) = \bar{\phi}_\epsilon(x) - \phi_\epsilon(x)$. By the regularity condition, we can assume that

$$\Delta_\epsilon(x) = \epsilon\Delta_1(x) + \frac{\epsilon^2}{2}\Delta_2(x) + \cdots,$$

where $\Delta_1(x), \Delta_2(x), \cdots$ are functions independent of ϵ. By the integration normalization condition,

$\int_R \Delta_k(x)\,dx = 0$ for $k = 1, 2, \cdots$. On the other hand, since

$$\bar{\phi}_\epsilon(x+\epsilon) = \bar{\phi}_\epsilon(x) + \epsilon\bar{\phi}'_\epsilon(x) + \frac{\epsilon^2}{2}\bar{\phi}''_\epsilon(x) + \cdots,$$

we have

$$2[H(E)\Delta_\epsilon](2x) - \Delta_\epsilon(x) - \epsilon\Delta'_\epsilon(x) = \frac{\epsilon^2}{2}\bar{\phi}_\epsilon(x) + \text{higher } \epsilon \text{ terms}.$$

Comparing the first order ϵ term, we obtain

$$\Delta_1(x) = 2[H(E)\Delta_1](2x).$$

Hence Δ_1 is a constant multiple of the unperturbed scaling function $\phi(x)$. It must be 0 since $\int_R \Delta_1(x)\,dx = 0$. Therefore,

$$\bar{\phi}_\epsilon(x) - \phi_\epsilon(x) = \Delta_\epsilon(x) = \frac{\epsilon^2}{2}\Delta_2(x) + \cdots = O(\epsilon^2).$$

The uniformality of this relation follows from the fact that both $\phi_\epsilon(x)$ and $\bar{\phi}_\epsilon(x)$ are uniformly weakly compactly supported. The proof for $\psi_\epsilon(x)$ is done in a similar way. □

Therefore visually, $\phi_\epsilon(x)$ accomplishes two things simultaneously: infinitely smoothing the original scaling function $\phi(x)$ (hence also the wavelet) and shifting it (rightward) by a small distance 2ϵ.

Corollary 15 (Linear Orthogonality) *Suppose the original scaling function and wavelet lead to an orthogonal MR, and the scaling function is at least C^2, then the sequences $(s_n), (r_n)$ and (w_n) have order $O(\epsilon^2)$. In such a case, we say that the QMR is* linearly orthogonal.

Finally, we point out that if one considers the following second order perturbed refinement equation (or wavelet equation):

$$\frac{\epsilon^2}{2}\phi''_{\epsilon,2}(x) + \epsilon\phi'_{\epsilon,2}(x) + \phi_{\epsilon,2}(x) = 2[H(E)\phi_{\epsilon,2}](2x), \qquad \int_R \phi_{\epsilon,2}(x)\,dx = 1,$$

then the following results can be established in a similar manner:

(1) $\phi_{\epsilon,2} = K_{\epsilon,2} * \phi$, where

$$K_{\epsilon,2} = \frac{1}{[\exp_2(-1)]^2}\frac{\sqrt{2}}{\epsilon}\Phi_{\frac{\pi}{4}}\left(\frac{\sqrt{2}}{\epsilon}x\right).$$

(2) $\phi_{\epsilon,2}$ is C^∞ and supported in $[0,\ \infty)$ and weakly compactly supported.

(3) (Quadratic Orthogonality) If the original scaling function $\phi(x)$ is C^α for some $\alpha \geq 3$, then

$$\phi_{\epsilon,2}(x) = \phi(x - 2\epsilon) + O(\epsilon^3),$$

$$\psi_{\epsilon.2}(x) = \psi(x - 2\epsilon) + O(\epsilon^3),$$

and the uniform norm of $(s_n), (r_n)$ and (w_n) are all have order $O(\epsilon^3)$ if the original MR is orthogonal. Hence the QMR is "more" orthogonal than the previous case.

Chapter 4

Asymptotics of Optimal Lowpass Filters

Digital filters are polynomials (in terms of the "delay" variable z^{-1}) or trigonometric polynomials (in terms of the "frequency" variable ω with $z = e^{i\omega}$). Filter design almost surely starts with a desired shape (or more generally, constraints) in the frequency domain. This ideal shape is specified by the specific task at hand. An important class of digital filters are the optimal filters: those that are best under certain constraints (for example, those that give the best approximation to a given shape). Inevitably, mathematically, we are led to the analysis and construction of "best" polynomials, a very old yet still surprisingly active field. The Green's function and equilibrium distribution of the underlying domain play a crucial role in the whole analysis.

This chapter consists of two parts, with slightly different motivations and styles. Part I(section 4.1) is aimed to interpreting and improving a very important empirical formula established by Jim Kaiser (Bell Lab) in the beginning of the digital filter age. The presentation meets the taste of signal processing engineers. Part II(section 4.2 and 4.3) studies the properties of the Green's function and the equilibrium distribution of a several-interval domain and their asymptotics. The invention of concepts like "critical polynomials", the discovery of the "square root law", and the by-product application in numerical linear algebra manifest that this part is "more" mathematical.

4.1 Asymptotics of the Error-Length Relation

4.1.1 Introduction

It is a familiar (and happy) fact that the equiripple property of an optimal lowpass filter suggests a good algorithm for designing that filter. This is the Remez-Parks-McClellan algorithm (see Cheney [8, 1966] and Parks and McClellan [58, 1972]), which iteratively pushes down the error at its maximum point. Eventually the error has equal magnitudes and alternating signs at $N+2$ points. Since no polynomial of degree N can have $N+1$ sign changes, this equiripple filter cannot be improved at all $N+2$ points. It is optimal (in the minimax sense). The algorithm is directly available in MATLAB as `remez.m` and is very widely used.

The designer begins with a passband (ending at frequency ω_p) and a stopband (starting at ω_s) and an acceptable error. This section considers first the weight-free case with equal errors in the passband and stopband: $\delta_p = \delta_s = \delta$. The transition bandwidth $\Delta\omega = \omega_s - \omega_p$ is critical to the relation of the filter length $N+1 = 2n+1$ to the distance δ from an ideal one-zero response. A useful formula derived experimentally by Kaiser [40, 1974] suggests an appropriate filter length. There are similar formulas in Rabiner and Gold [60, 1975] and Vaidyanathan [74, 1992]. Kaiser's is the simplest and most characteristic:

$$N \simeq \frac{20\log_{10}\delta^{-1} - 13}{2.324\Delta\omega} \tag{4.1}$$

For this value of N, the Remez algorithm yields the frequency response $H(\omega)$ closest to the ideal "one-zero function" $F(\omega)$ on the union of passband $|\omega| \le \omega_p$ and stopband $|\pi - \omega| \le \pi - \omega_s$. The code outputs the coefficients $h[0], \cdots, h[N]$ of this optimal lowpass filter, for which the error is approximately δ (See Figure 4-1).

Our section analyzes this relation of δ to N (or n). The error decays exponentially, $\delta \approx e^{-n\beta}/\sqrt{n}$, and the key problem is to compute the exponent $\beta = \beta(\omega_p, \omega_s)$. The leading term of β is controlled by $\Delta\omega = \omega_s - \omega_p$ and our asymptotic result is close to Kaiser's experiments for small δ, see Eq.(15):

$$N \simeq \frac{20\log_{10}\delta^{-1} - 10\log_{10}\log_{10}\delta^{-1}}{2.171\Delta\omega}$$

This *asymptotic* result is later modified to the *semi-empirical* formula (4.17), which applies to a wide range of practical parameters and is hence recommended to replace Kaiser's empirical formula.

Kaiser also discovered a nearly optimal family of filters based on the I_0-sinh function. An empirical formula similar to (4.1) was also established by Kaiser [40, 1974] for this family. The constant in the denominator becomes slightly smaller, which increases N. This family was analyzed

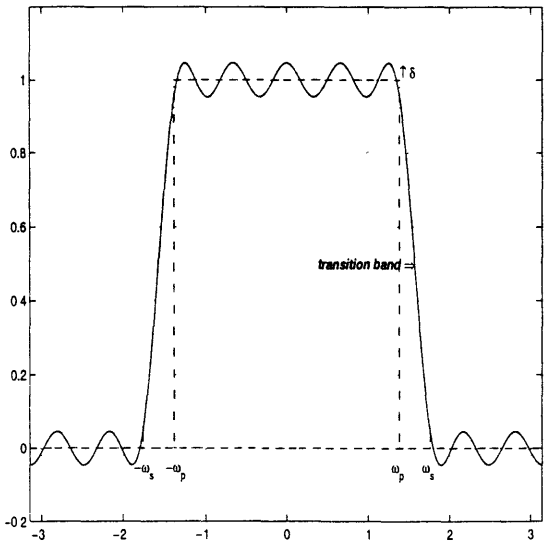

Figure 4-1: The frequency response of the optimal FIR lowpass filter with $N + 1 = 21$ coefficients and $\omega_p = 0.44\pi$, $\omega_s = 0.56\pi$.

theoretically by Fuchs et. al. in [33, 1980] and we return to it in Section 4.1.5.

The fundamental tool in the analysis is the *Green's function* $g(z)$, which solves Laplace's equation on the complement of two intervals (with a pole at infinity). This function has a unique critical point σ, and β is actually $g(\sigma)$. Since our intervals are real, the critical point is also real (and it lies in the transition band). But our problem is emphatically one of *complex* and not *real* analysis. The oscillations of a real polynomial prove that an equiripple filter is optimal, but for more information we must go deeper into the complex plane!

We give references to fundamental work of Walsh [76, 1965] and Widom [80, 1969] and Fuchs [31, 1978]. The virtue of complex analysis is to permit contours of integration to be deformed. Then the leading term in an integral with a large parameter can be computed by the method of steepest descent.

The Green's function has an explicit simple form only in the symmetric case, when $\omega_s + \omega_p = \pi$. Then the critical frequency is $\omega_c = \pi/2$, at the center of the transition band. Our analysis is most complete in this symmetric case. Our task in all other cases (when $g(z)$ becomes an elliptic function) is to recapture the same form, in which $N\Delta\omega$ plays such a key role.

We also present early results on *nearly optimal filters*, for which δ_n is of the same order as the optimal error sequence. Unlike equiripple filters, nearly optimal filters may have closed forms and allow fast algorithms. For the symmetric and weight-free case, we propose an explicit set of interpolation points. This leads to the discovery of the asymptotic behavior of optimal filters *in the transition band.* The frequency response is close to an error function. The limit as $n \to \infty$ is the

ideal brick wall filter with cutoff at the critical frequency ω_c.

This first part of the chapter has been organized as follows. Section 4.1.2 introduces an important result due to Fuchs in approximation theory. The complete asymptotic relation among design parameters in the symmetric case $\omega_p + \omega_s = \pi$ is derived in section 4.1.3. For the non-symmetric case, theoretical results as well as a MATLAB algorithm for the crucial geometric constant β are described in section 4.1.4. Asymptotic analysis is also carried out for the case of narrow transition band. In section 4.1.5, our results are compared with that of Fuchs on Kaiser's window family of filters. In section 4.1.6, we show the numerical comparison between our asymptotic formula and Kaiser's empirical one. Section 4.1.7 describes the asymptotic behavior of optimal filters in the transition band. Some proofs are included in the appendix.

4.1.2 Leading Order For δ_n

Leading Order for General Problem

We now present Fuchs' result on polynomial approximation on several domains in the complex plane. Let K be a compact domain with disjoint simply connected components $K_1, \cdots, K_m$. Our problem is to approximate by polynomials the function $f(z)$ that equals $h_\iota(z)$ on the component K_ι. (The $h_\iota(z)$ are entire functions and not all identical.) The minimum error in the maximum norm is δ_n when the polynomials have degree at most n:

$$\delta_n = \min_{p \in P_n} \max_{z \in K} |f(z) - p(z)|.$$

Here P_n denotes the space of all polynomials of degree not greater than n.

Theorem 19 (Fuchs) *There exist a non-negative integer q, a positive number β, and two positive constants A_- and A_+, such that*

$$A_- n^{q-\frac{1}{2}} \exp(-n\beta) \leq \delta_n \leq A_+ n^{q-\frac{1}{2}} \exp(-n\beta). \tag{4.2}$$

Remark. The nonnegative integer q is determined by the objective function $f(z)$ and domain K together. It is the multiplicity of a particular critical point as a zero of a difference $h_i(z) - h_j(z)$ (Fuchs [31, 1978] gives details). Our case will automatically have $q = 0$, since $h_0(z) = 1$ and $h_1(z) = 0$.

The exponent β is a geometric constant, entirely determined by K. For $m = 2$, β is *Green's logarithmic radius of the unique critical point* of K^c. Its meaning will be explained immediately.

Potential Theory in the Complex Plane

Let $G(z,s)$ be the Green's function for the Laplacian on the complement K^c, which is completely characterized by the following properties:

- $G(z,s)$ is harmonic over K^c except at $z=s$, where $G(\cdot,s)$ behaves like $-\ln|z-s|$ (or $\ln|z|$ when $s=\infty$).
- For any fixed s, $G(z,s)$ goes to zero as z approaches ∂K^c, the boundary of K^c.

We are particularly interested in $g(z)=G(z,\infty)$. For any $z\in K^c$, $g(z)$ is called its *Green's logarithmic radius*, and is denoted by $|z|_K$. The function g has exactly $m-1$ *critical points* ordered by $|\sigma_1|_K \le |\sigma_2|_K \le \cdots \le |\sigma_{m-1}|_K$ inside the domain K^c (Nevanlinna [55, 1970]). A critical point of g (or of K^c) means that the gradient at σ is zero. Geometrically, the level line of g through σ is self-intersected at σ (see Figure 4-2). Then β in Fuchs' theorem is given by

$$\begin{aligned}\beta = &|\sigma_I|_K : f \text{ can be continued analytically on} \\ &|z|_K < |\sigma_I|_K \text{ but not on } |z|_K < |\sigma_{I+1}|_K.\end{aligned} \tag{4.3}$$

Since the h_i are not identical, σ_I does exist. When $m=2$, β must be $|\sigma|_K = g(\sigma)$ at the unique critical point σ.

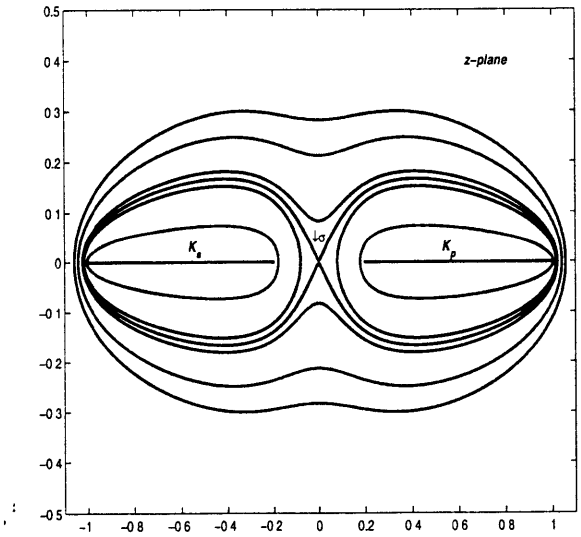

Figure 4-2: The level lines and critical point σ of the Green's function associated to a domain $K=[-1,-b]\cup[a,\ 1]$. Here we show the case $a=b=0.2$. By symmetry $\sigma=0$.

Remark. For optimal polynomial approximation, real analysis yields the famous "Alternation Theorem" (the equiripple property and exchange algorithm, see Cheney [8, 1966] and Rabiner and

Gold [60, 1975]). The deeper asymptotic problems require complex analysis and potential theory. We recommend the classical monographs by Walsh [76, 1965] and Henrici [38, 1986].

Leading Order for δ_n

It is natural to work in the $x = \cos\omega$ domain. Let $x_p = \cos\omega_p$ and $x_s = \cos\omega_s$. The passband and stopband become $K_p = [x_p,\ 1]$, $K_s = [-1,\ x_s]$. Then $K = K_p \cup K_s$ and σ is the unique critical point of K^c.

Lemma 9 *There exist two positive constants A_- and A_+ such that for all n*

$$A_- n^{-\frac{1}{2}} \exp(-n|\sigma|_K) \le \delta_n \le A_+ n^{-\frac{1}{2}} \exp(-n|\sigma|_K) \tag{4.4}$$

Proof. Use Theorem 19 for this special case of $m = 2$. By equation (4.3), $\beta = |\sigma|_K$. On the other hand, $h_0 \equiv 1$ and $h_1 \equiv 0$. Therefore $z = \sigma$ is a zero of order $q = 0$ of $h_0(z) - h_1(z)$. □

Remark. Our K has only two free parameters x_p and x_s (or ω_p and ω_s). So we will also use the function symbol $\beta(x_p, x_s)$ or $\beta(\omega_p, \omega_s)$. To determine the leading order of δ_n, we have to compute β explicitly, which is the task of the next two sections. Lemma 9 leads to the following theorem in terms of logarithms. Its proof has been placed in Appendix A.

Theorem 20 *For long equiripple filters ($n \gg 1$), the asymptotic error satisfies*

$$n \simeq \frac{\ln \delta_n^{-1} - \frac{1}{2} \ln\ln \delta_n^{-1}}{\beta(x_p, x_s)} \tag{4.5}$$

Remark. The empirical formulas (Kaiser [40, 1974], Rabiner and Gold [60, 1975], and Vaidyanathan [74, 1992]) only catch the leading term $\ln \delta_n^{-1}$. They do not capture the correct β or the double logarithm term due to the factor $n^{-1/2}$ (which is overshadowed by the exponential term in all experiments).

4.1.3 The Symmetric Case

In the next section, we shall see that $\beta(x_p, x_s)$ generally has no description by elementary functions. In the symmetric case $x_p + x_s = 0$, or $\omega_p + \omega_s = \pi$, the Green's function simplifies and β can be computed explicitly. Several elementary properties will be useful (referred to as Property 1,2,3 later):

1. (*Unit disk*) The Green's function for the domain $|w| \le 1$ with source $s = 0$ is $-\ln|w|$.

2. (*Conformal equivalence*) Suppose $w = f(z)$ is a conformal mapping from a domain K_z onto a domain K_w. Assume that f is continuous up to the boundary and $f(\partial K_z) \subseteq \partial K_w$. Let z_0 be an

interior point of K_z and $w_0 = f(z_0)$. Suppose $g_0(w)$ is the Green's function for K_w corresponding to source w_0. Then $g_0(f(z))$ is the Green's function of K_z corresponding to source z_0.

3. (*Pullback by covering mapping*) In Property 2, suppose that f is an analytic mapping, but z_0 is the only preimage of w_0 and all the other conditions still hold. Then $g_0(f(z))/d$ is still the corresponding Green's function provided that z_0 is the $(d-1)$-multiple zero or pole of $f'(z)$.

Lemma 10 (Green's function: symmetric case) *Suppose $\omega_p + \omega_s = \pi$. Then $x_p = -x_s = a > 0$. The Green's function $g(z)$ for K^c corresponding to source $s = \infty$ is*

$$-\frac{1}{2}\ln\left|\frac{2}{1-a^2}[\,z^2 - a^2 - \sqrt{(z^2-a^2)(z^2-1)}\,] - 1\right|$$

Here the square root has K as its branch line and takes a positive value at $z = 2$.

Proof. Define

$$\phi(Z) = \frac{2}{1-a^2}[Z - a^2 - \sqrt{(Z-a^2)(Z-1)}] - 1$$

Here $Z = z^2$ folds K into a single interval $I = [a^2,\ 1]$ in the Z–plane. The inverse Joukowski transform $w = \phi(Z)$ maps the complement of I onto the unit disk D_w in the w–plane, and maps $Z = \infty$ to $w = 0$. Let $f(z) = \phi(z^2)$. Then this lemma is a direct conclusion from Properties 1 and 3 with $d = 2$. □

Lemma 11 *Suppose $x_p = -x_s = a$. Then the exponent in the error formula is*

$$\begin{aligned}\beta &= \frac{1}{2}\ln\frac{1+x_p}{1-x_p} \\ &= \frac{1}{2}\ln\frac{1+\cos\omega_p}{1-\cos\omega_p} = \ln\cot\frac{\omega_p}{2}\end{aligned} \tag{4.6}$$

Proof. By symmetry, the unique critical point for K^c must be $\sigma = 0$. Therefore

$$\beta = g(0) = \frac{1}{2}\ln\frac{1+a}{1-a}.$$

□

The combination of (4.5) and (4.6) can be used for design problems when $\omega_p + \omega_s = \pi$. Notice that even in the symmetric case, β is not strictly linear in $\Delta\omega$. However, Kaiser's idea of linear approximation to β as shown in the denominator of his formula (4.1) is good for most applications. The estimated coefficient 2.324 can be improved by our asymptotic analysis. We now look for a theoretical formula in the symmetric case that is similar to Kaiser's.

Suppose $\Delta\omega \ll 1$. Noticing $\omega_p = \pi/2 - \Delta\omega/2$, we have

$$\beta \simeq x_p = \cos(\pi/2 - \Delta\omega/2) \simeq \Delta\omega/2.$$

In Eq.(4.5), we replace β by $\Delta\omega/2$ and rewrite it in terms of decibels by changing the logarithm to base 10. By ignoring the $O(1)$ term (compared with logarithms of δ_n^{-1}), we obtain the following.

Theorem 21 (Asymptotic Relation of N to δ_n) *Assume that $\omega_p + \omega_s = \pi$ and ω_p is close to $\pi/2$. Then the order is related to the ripple height δ_n by*

$$N = 2n \simeq \frac{20\log_{10}\delta_n^{-1} - 10\log_{10}\log_{10}\delta_n^{-1}}{(5\log_{10}e)\,\Delta\omega}. \tag{4.7}$$

Remark.

(a) Numerically $\ln\cot(\omega_p/2)$ is close to $\Delta\omega/2$ except when $\Delta\omega$ is close to π (see Figure 4-3). For most applications, $\Delta\omega$ is small. Hence $\Delta\omega/2$ is a satisfactory approximation to β. In fact, when $\Delta\omega = \pi/4$, the relative error is only $(\beta - \Delta\omega/2)/\beta \simeq 2.6\%$.

(b) Kaiser's linear coefficient 2.324 is larger than our corrected value $5\log_{10}e \simeq 2.171$. The relative error is $(2.324 - 2.171)/2.171 \simeq 7\%$. This slope deviation can be detected in Figures 4-5 and 4-6.

(c) Since the second leading term for n is a double logarithm, the number "13" in Kaiser's formula is not correct theoretically. However, it does reveal the fact that the second leading term changes very slowly. Practically we only deal with δ_n ranging from 10^{-1} to 10^{-16}. Then the double logarithm in (4.7) goes from 0 to 16 (and 13 is inside this range).

4.1.4 The General Case

In the non-symmetric case, $\beta(x_p, x_s)$ is no longer an elementary function. In this section, we first describe the theoretical approach to determine β, and then create a numerical algorithm using MATLAB to compute it.

Conformal Equivalence to Annulus

Lack of symmetry ($\omega_p + \omega_s \neq \pi$) makes K^c a nontrivial doubly connected domain (DCD). Hence one has to turn to the general theory. A famous theorem says that any DCD is conformally equivalent to

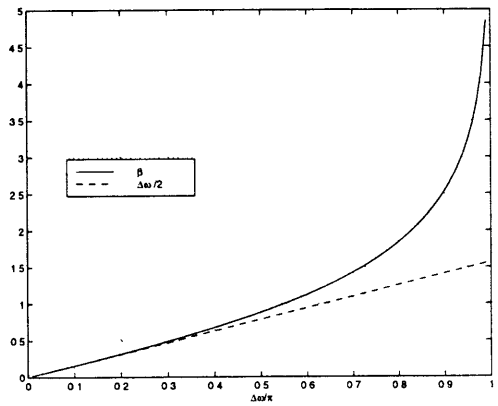

Figure 4-3: $\beta = \ln \cot \frac{\omega_p}{2} \simeq \frac{\Delta\omega}{2}$. The dashed line corresponds to $\beta \simeq \frac{\Delta\omega}{2}$ and the solid line is the true $\beta(\omega_p)$ with $\omega_p = \frac{\pi}{2} - \frac{\Delta\omega}{2}$. The horizontal axis shows $\Delta\omega/\pi$.

an annulus $A_r : r < |w| < 1$ (see Nehari [53, 1952]). The "modulus" r is uniquely determined by the DCD. In our case, this conformal mapping can be obtained in closed form using elliptic functions. The inverse mapping $z = f(w)$ is (Kober [44, 1957])

$$\frac{1+x_p}{2} - \frac{1-x_p}{2} \frac{\mathrm{sn}^2(\frac{K'}{\pi} \ln w; k) + \mathrm{sn}^2(\frac{K'}{\pi} \ln s; k)}{\mathrm{sn}^2(\frac{K'}{\pi} \ln w; k) - \mathrm{sn}^2(\frac{K'}{\pi} \ln s; k)}.$$

Here $0 < s < 1$, and $f(s) = \infty$.

The three parameters r, s, k are given by Freund [30, 1991] as functions of x_p and x_s.

$$k = \sqrt{\frac{2(x_p - x_s)}{(1+x_p)(1-x_s)}} \tag{4.8}$$

$$r = \exp\left(-\frac{\pi K_c(k)}{K'(k)}\right) \tag{4.9}$$

$$s = \exp\left(-\frac{\pi K_i(k)}{K'(k)}\right). \tag{4.10}$$

The elliptic functions $\mathrm{sn}(u; k)$, $K_c(k)$, $K_i(k)$, and $K'(k)$ are defined in Appendix B.

Green's Function and β

Let $g_A(w)$ denote the Green's function for the annulus A_r corresponding to the source s. Then by Property 2, $g(z) = g_A(f^{-1}(z))$ is the Green's function for K^c corresponding to $s = \infty$. Let σ and σ_A denote the unique critical points of $g(z)$ and $g_A(w)$. Then $\sigma = f(\sigma_A)$ since f preserves level lines. Hence $\beta(x_p, x_s) = g(\sigma) = g_A(\sigma_A)$.

Define $\lambda = \ln s / \ln r \in (0, 1)$. For any c inside the unit circle, the symbol $[c] = [c](w)$ denotes

the Möbius transform of the unit disk associated with c:

$$[c](w) = \frac{w - c}{c^* w - 1}.$$

Then $g_A(w)$ is given by Akhiezer [1, 1990] as

$$\lambda \ln|w| - \ln|[s]| - \sum_{j=1}^{\infty} \left(\ln|[r^{2J} s]| + \ln|[\frac{r^{2J}}{s}]| \right).$$

The partial sum from 1 to J of this infinite series converges on A_r with rate $O(r^{2J})$. For small r this is quite satisfactory. However when the transition band is narrow, r defined by (4.8) is close to 1. So the following form of $g_A(w)$ is much better numerically:

$$\lambda \ln|w| - \ln s - \ln|S^+(w)| + \ln|S^-(w)| \tag{4.11}$$

Now the partial sums of S^+ and S^- from $-J$ to J give greater accuracy $O(r^{J^2})$:

$$S^+(w) = \sum_{J=-\infty}^{\infty} r^{J^2} \left(\frac{-w}{rs} \right)^J \tag{4.12}$$

$$S^-(w) = \sum_{j=-\infty}^{\infty} r^{J^2} \left(\frac{-ws}{r} \right)^j. \tag{4.13}$$

Our MATLAB code uses this form for $g_A(w)$.

Theoretically, the unique critical point σ_A can be located as the zero of the gradient vector ∇g_A. This generally requires substantial computation. The following theorem changes it to a one-dimensional optimization problem.

Theorem 22 (β by Optimization) *Consider*

$$g_A(x) = \lambda \ln(-x) - \ln s - \ln S^+(x) + \ln S^-(x) \tag{4.14}$$

for $-1 \le x \le -r$. *Then* $\beta(x_p, x_s) = \max g_A(x)$.

Proof. By definition, $g_A(x) \ge 0$ and $g_A(-1) = g_A(-r) = 0$. Hence $g_A(x)$ reaches its maximum value inside $(-1, -r)$. On the other hand, since $g_A(w)$ is symmetric with respect to y ($w = x + iy$), $\partial g_A / \partial y$ must be zero along $(-1, -r)$. Therefore $\partial g_A(w)/\partial x = 0$ immediately implies a critical point of g_A. Since there is only one critical point σ_A, it must yield the maximum of $g_A(x)$. Hence

$$\beta(x_p, x_s) = g_A(\sigma_A) = \max g_A(x).$$

□

Algorithm and MATLAB Code

The complete elliptic function called `ellipk` in MATLAB can be used to compute K_c and K'. For the incomplete elliptic function K_i, we apply MATLAB integration `quad8` to the function `for_call` , which is simply $1/\sqrt{1-m\sin^2 x}$ with $m = k^2$. Set $\theta = \sin^{-1}\alpha$ (α is defined in Appendix B (iii).) Then

$$\text{“}K_i = \text{quad8('for\_call', 0, } \theta, \text{ } 1e-14, \text{ [ ], } m);\text{”}$$

computes K_i to the precision 10^{-14}. This yields r and s from (4.9) and (4.10) (by `RS.m`). Then `Green.m` uses (4.11)–(4.13) to compute the Green's function g_A on the annulus A_r. Our last program `betak.m` applies the minimization `fmin` to $-g_A(x)$ defined in (4.14) and finally finds β. We distinguish `betak` from MATLAB's `beta`.

Asymptotics for Narrow Transition

When we compute β numerically, we don't know its exact behavior as a function of ω_p and ω_s. To compare with earlier empirical formulas, we apply asymptotic analysis to β when the transition bandwidth is narrow ($\Delta\omega \ll 1$) and fixed. In practice, this narrow transition is preferred. We measure ω_p and ω_s from the mid-frequency $\omega_m = \frac{1}{2}(\omega_p + \omega_s)$:

$$\omega_p = \omega_m - \frac{\Delta\omega}{2} \qquad \text{and} \qquad \omega_s = \omega_m + \frac{\Delta\omega}{2}.$$

Since $\Delta\omega$ is fixed, $\beta(\omega_p, \omega_s)$ becomes a function only of ω_m and is denoted by $\beta(\omega_m)$.

Theorem 23 *The leading term of $\beta(\omega_m)$ is $\beta(\pi/2)$ in the range $\Delta\omega \ll \min(\omega_m,\ \pi - \omega_m)$. Practically, the range can be taken as (see Figure 4-4):*

$$\Delta\omega \ < \ \omega_m \ < \ \pi \ - \ \Delta\omega.$$

The proof is in Appendix C.

It is Kaiser's empirical formula (4.1) that led to our discovery of Theorem 23. In turn, our asymptotic result provides a theoretical support to the *form* of his empirical formula. The transition bandwidth $\Delta\omega$ is crucial and the position ω_m of the transition band has small effect. Our analysis

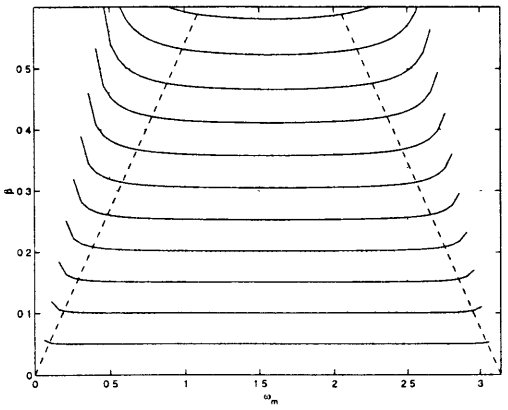

Figure 4-4: Theorem 23: $\beta(\omega_m) \simeq \beta(\pi/2)$. Each solid horizontal line represents $\beta(\omega_m)$ when $\Delta\omega$ is fixed. From the bottom to the top, $\Delta\omega = 0.1 : 0.1 : 1.1$. The segments bounded by the diagonal dashed lines show the practical range inside which $\beta(\omega_m) \simeq \beta(\pi/2)$.

gives the correct constant in the leading term, and also the next term. With the help of Theorem 23, Theorem 22 generalizes to the non-symmetric case.

Theorem 24 *If* $\Delta\omega \ll \min(\omega_m,\ \pi - \omega_m)$, *then*

$$N = 2n \simeq \frac{20\log_{10}\delta_n^{-1} - 10\log_{10}\log_{10}\delta_n^{-1}}{(5\log_{10} e)\ \Delta\omega}. \tag{4.15}$$

In practice, this yields good results for $\Delta\omega \leq \pi/4$ *and* $\omega_m \in [\Delta\omega,\ \pi - \Delta\omega]$ *by the remark of Theorem 21 and Theorem 23.*

4.1.5 Kaiser's Filters Are Near Optimal

Besides the equiripple filters, another popular way of designing FIR filters is the *window method.* The ideal one-zero lowpass filter is IIR. In the frequency domain, we convolve this ideal response with the window response. The frequency response of a window is often a damped wave. The narrowness of the main lobe and side lobes determines the quality of the resulting FIR filter.

Kaiser used the non-linearly scaled zeroth–order modified Bessel function to create a family of windows with good properties. They are of limited duration in the time domain and have most of their energy concentrated at low frequency. (This is the core idea of modern wavelet analysis.) Most important, these filters are *nearly optimal*: the error sequence has the same order as that of the optimal approximation.

First, Kaiser [40, 1974] established an empirical formula for his windows when $\delta < 0.1$:

$$N \simeq \frac{20 \log_{10} \delta^{-1} - 8}{2.285\, \Delta\omega}.$$

(We have converted from Δf to $\Delta\omega = 2\pi\Delta f$.) Fuchs, Kaiser, and Landau [33, 1980] proved that for large window parameter α,

$$\delta \simeq \left[\frac{8}{\pi\Delta\omega}\right]^{\frac{1}{2}} N^{-\frac{1}{2}} \exp\left(-\frac{\Delta\omega}{4}N\right).$$

Comparing with our Eq.(4.2) Kaiser's windows are indeed nearly optimal (but not exactly, since $\beta \simeq \frac{\Delta\omega}{2}$ is only an approximation). Similar to the way we have proved Theorem 20, Fuchs showed that

$$N \simeq \frac{20 \log_{10} \delta^{-1} - 10 \log_{10} \log_{10} \delta^{-1}}{(5 \log_{10} e)\Delta\omega}.$$

This is exactly (4.15).

The Chebyshev optimal filter is completely characterized by the equiripple property. The underlying mechanism of Fuchs' result is that for large α, the side-lobes have approximately the same L_1 norms (same areas). This makes Kaiser's filters near equiripple and hence near optimal.

4.1.6 Numerical Experiments

We use the MATLAB function `remez.m` to compute the minimal error δ_N corresponding to each N. The result is then used to test Kaiser's empirical formula and our asymptotic formula.

Narrow Transition

For narrow transition (this practically extends to $\Delta\omega \leq \pi/4$), Theorem 24 gives the first two leading terms of N. However, to make Eq.(4.15) accurate even for small N, we have to know the constant A_n appearing in the proof of Theorem 20. This means that we have to add a constant term (independent of δ_n) in the numerator of Eq.(4.15). Finding A_n is a mathematically open problem, but our numerical experiments indicate that we can take this constant term as $20 \log_{10} \pi$. Then the following formula applies to all N:

$$N = 2n \simeq \frac{20 \log_{10}(\pi\delta_n)^{-1} - 10 \log_{10} \log_{10} \delta_n^{-1}}{(5 \log_{10} e)\ \Delta\omega}. \tag{4.16}$$

This is very accurate for small $\Delta\omega$. Our experiments have $\Delta\omega = 0.02\pi, 0.04\pi, \cdots, 0.10\pi$ and

$\omega_m = \pi/2$. For each $\Delta\omega$, first we use `remez.m` to compute the N–δ relation exactly. With this result we test the predictions by Kaiser's empirical formula (4.1) and our asymptotic formula (4.16). The test results are plotted in Figure 4-5. It shows that Eq. (4.16) is more accurate.

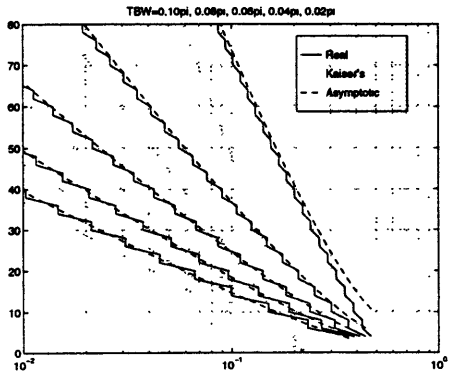

Figure 4-5: Comparison of Kaiser's formula and formula (4.16). There are five sets of curves in the plot, one for each $\Delta\omega$. From right to left, $\Delta\omega = (0.02 : 0.02 : 0.10)\pi$. Each set contains three lines—solid, dotted, and dashed, corresponding to the real N–δ relation, Kaiser's empirical prediction, and the asymptotic prediction by Eq.(4.16).

Modified to Include Wide Transition

For wide transition, say $\Delta\omega \simeq 0.5\pi$, both Kaiser's formula and formula (4.16) assume that β is a linear function of $\Delta\omega$. Generally we need the original $\beta(\omega_m) \simeq \beta(\pi/2) = \ln\cot(\pi - \Delta\omega)/4$ in the denominator. Then the following formula is very accurate even for wide transition:

$$N = 2n \simeq \frac{20\log_{10}(\pi\delta_n)^{-1} \; - 10\log_{10}\log_{10}\delta_n^{-1}}{(10\log_{10} e)\;\ln\cot\frac{\pi-\Delta\omega}{4}}. \tag{4.17}$$

The experiments for wide transition are plotted in Figure 4-6, with δ on the horizontal axis and N on the vertical.

So finally, we would recommend Eq. (4.17) for all design problems with either wide or narrow transition $\Delta\omega$, and symmetric or non-symmetric bands.

One Example

We compare the accuracy of the formulas through a real design problem. Suppose that $\omega_p = .5\pi$, and $\omega_s = .54\pi$. We want an equiripple filter whose passband and stopband errors are $\delta_p = \delta_s = \delta = 0.02$.

By Kaiser's formula (4.1), the filter length should be $N_K = 72$. The exchange algorithm

$$H_K = \text{remez}(N_K, [0\ .5\ .54\ 1],\ [1\ 1\ 0\ 0])$$

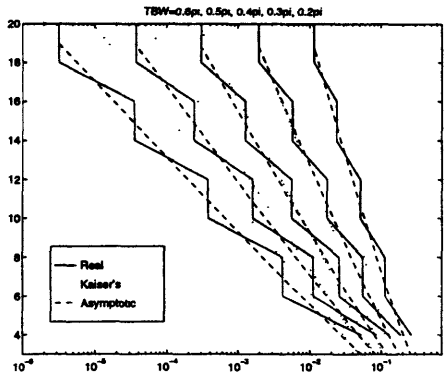

Figure 4-6: Comparison of Kaiser's formula and formula (4.17). The five sets of curves now correspond to wider transitions $\Delta\omega = (0.2 : 0.1 : 0.6)\pi$.

gives the impulse response of the equiripple filter H_K. The actual ripple height is $\delta_K = 0.0255$. Hence the relative design error is

$$r_K = \frac{|\delta - \delta_K|}{\delta} = 27.5\%.$$

The corresponding data using our asymptotic formulas (4.17) or (4.16) are:

$$N_A = 80, \qquad \delta_A = 0.0192, \qquad r_A = 4\%.$$

4.1.7 The Transition Band

This section describes the asymptotic behavior of the equiripple filter response $H_N^{opt}(\omega)$ inside the transition band $\omega_p \leq |\omega| \leq \omega_s$ as the filter length $N + 1 = 2n + 1$ increases. This behavior reveals the convergence of impulse responses to the ideal 0-1 filter with passband $|\omega| \leq \omega_c$. We thank Alan Oppenheim for bringing this problem to our attention.

Nearly Optimal Filters

A family of FIR filters $H_N(\omega)$ of length $N + 1$, is said to be *nearly optimal* if its error sequence

$$e_N = \|H_N(\omega) - I(\omega)\|$$

is of the same order as the optimal error sequence. This means that $e_N \leq C\delta_N$ for a fixed C.

Nearly optimal filters serve two purposes. Unlike equiripple filters, they may have closed forms and allow direct mathematical analysis. Their properties should give an approximation to their

counterparts (the optimal equiripple filters). Second, by relaxing the optimality, we may have a better design algorithm, such as direct interpolation. For the symmetric case $\omega_p + \omega_s = \pi$, we do find such an interpolation scheme.

Theorem 25 *Suppose $\omega_p + \omega_s = \pi$ and $x_p = \cos\omega_p = a > 0$. Define $2k$ points $x_j^{\pm}, j = 1, 2, \cdots, k$ by*

$$x_j^{\pm} = \pm\left[\frac{1+a^2}{2} + \frac{1-a^2}{2}\cos\frac{j-\frac{1}{2}}{k}\pi\right]^{\frac{1}{2}}.$$

Let $p_n(x)$ denote the unique polynomial of degree $n = 2k-1$ interpolating 1 at each x_j^+ and 0 at each x_j^-. Let $N = 2n$ and define

$$H_N(\omega) = p_n(\cos\omega).$$

Then $H_N(\omega)$ is a sequence of nearly optimal filters.

Asymptotics in the Transition Band

With the help of $H_N(\omega)$ just constructed, we find the following asymptotic form of the equiripple filter $H_N^{opt}(\omega)$ in the transition band.

Theorem 26 *Let $\omega_m = \dfrac{\omega_p + \omega_s}{2}$ be the mid frequency in the transition band. For $\Delta\omega = \omega_s - \omega_p \ll 1$, the leading term of $H_N^{opt}(\omega)$ on $\omega_p \le \omega \le \omega_s$ is given by*

$$H_N^{opt}(\omega) \approx \operatorname{erf}\left(\sqrt{\frac{N\beta}{4}}\,\frac{\omega_m - \omega}{\omega_s - \omega_m}\right). \tag{4.18}$$

Here $\beta \approx \Delta\omega/2$ is the geometric constant appearing in previous sections and the error function $\operatorname{erf}(x)$ is defined by

$$\operatorname{erf}(x) = \frac{1}{\pi}\int_{-\infty}^{x} e^{-t^2}\,dt.$$

Practically, this approximation is very satisfactory for a wide range of transition bandwidths. Figure 4-7 shows the case of $\Delta\omega = .1\pi$, for both symmetric and non-symmetric bands.

Computational experiment guided by our error function formula leads to the following *semi-empirical* formula for the weighted case. In minimizing the maximum deviation from the ideal filter, the stopband error is weighted by W. In practice, W can be 100. The optimal filter with heights $\delta_p = W\delta_s$ is still denoted by $H_N^{opt}(\omega)$. Then the leading term approximation in the transition band

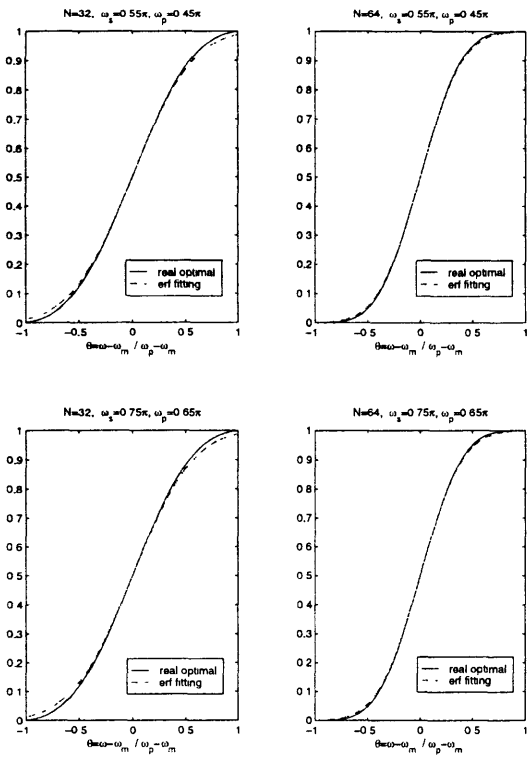

Figure 4-7: Closeness to the error function. The four windows show the scaled transition band: $\theta = \dfrac{\omega - \omega_m}{\omega_p - \omega_m}$. For example, ω_p now corresponds to $\theta = 1$. The solid lines represent the optimal equiripple filters, and the dashed lines show the error function (4.18). For the top two, $\omega_s = .55\pi$ and $\omega_p = .45\pi$ with $N = 32, 64$. For the bottom two, $\omega_s = .75\pi$ and $\omega_p = .65\pi$ with $N = 32, 64$. The fitting improves as the filter length N increases.

is:

$$H_N^{opt}(\omega) \simeq \operatorname{erf}\left(\sqrt{\frac{N\beta}{4}}\,\frac{\omega_m - \omega - S_N(W)}{\omega_s - \omega_m}\right),$$

with

$$S_N(W) = \frac{\ln W + \frac{1}{2}\ln\ln W}{2N}.$$

This experimental expression for the shift $S_N(W)$ has successfully predicted the impulse response of optimal filters in the transition band. They still converge to the ideal filter with cutoff frequency $\omega \approx \omega_m - S_N(W)$ for narrow transition band.

4.1.8 Appendix

A. Proof of Theorem 20 (Section 4.1.2) Suppose $\delta_n = A_n n^{-\frac{1}{2}} \exp(-n\beta)$, with $A_- \leq A_n \leq A_+$ by Lemma 9. Taking the natural logarithm yields

$$\ln \delta_n^{-1} = -\ln A_n + \frac{1}{2} \ln n + n\beta. \tag{4.19}$$

The dominant term on the right is $n\beta$, which must equal the dominant term on the left. Hence $\ln \delta_n^{-1} \simeq n\beta$. This determines the leading term. To find the next term, assume $n = \dfrac{\ln \delta_n^{-1}}{\beta} + \Delta n$. By (4.19),

$$0 = -\ln A_n + \frac{1}{2} \ln n + \beta \Delta n.$$

As $n \gg 1$, the dominant term $\frac{1}{2} \ln n$ can only be balanced by $\beta \Delta n$, since $\ln A_n$ is bounded. Hence

$$\Delta n \simeq -\frac{\frac{1}{2} \ln n}{\beta} \simeq -\frac{\frac{1}{2} \ln \ln \delta_n^{-1}}{\beta}.$$

B. Definitions of elliptic functions (Section 4.1.4)

(i) $v = \mathrm{sn}(u;\ k)$ is the Jacobian elliptic function with modulus $0 < k < 1$, defined by the incomplete elliptic integral:

$$u = \int_0^v \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

(ii) $K_c(k) = \mathrm{sn}^{-1}(1;\ k)$ is a complete elliptic integral:

$$K_c(k) = \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

$K'(k)$ in the expression of f is defined by $K'(k) = K_c(1-k^2)$, also a complete integral.

(iii) $K_i(k) = \mathrm{sn}^{-1}(\alpha;\ k)$ and $\alpha = \sqrt{\frac{1+x_p}{2}}$.

C. Proof of Theorem 23 (Section 4.1.4)

(i) The unique critical point σ of $g(z) = G(z, \infty)$ must lie inside (x_s, x_p). If the mid-frequency $\omega_m = \frac{1}{2}(\omega_p + \omega_s)$ is below $\pi/2$, the stopband K_s is longer than the passband K_p. Hence the

Green's function $g(z)$ grows more slowly near K_s. The maximum of g on $[x_s,\ x_p]$ occurs closer to x_p than to x_s, so that $\sigma > x_m$.

(ii) Define

$$d = \frac{1 + x_p x_s}{x_p + x_s} \qquad \text{and} \qquad c = d - \sqrt{d^2 - 1}.$$

Then

$$c \in [x_s,\ x_p] \qquad \text{and} \qquad c = x_m + O(\Delta x^2). \tag{4.20}$$

The linear fractional transform

$$z' = F(z) = \frac{z - c}{1 - cz}$$

maps $K = K_s \cup K_p$ onto a symmetric domain K' in the z'–plane:

$$K' = K'_s \cup K'_p = [-1,\ x'_s] \cup [x'_p,\ 1].$$

Here $x'_p = F(x_p) = -F(x_s) = -x'_s$.

(iii) Set $s' = F(\infty) = -1/c$ and $\sigma' = F(\sigma)$. Then σ' is the unique critical point of $G'(z',\ s')$. Since the new source s' lies inside $(-\infty, -1)$ and the new domain is symmetric, its Green's function $G'(z', s')$ grows more rapidly near K_s than K_p. Hence the maximum of $G'(\cdot, s')\Big|_{(x'_s, x'_p)}$ must occur closer to x'_s, implying that $\sigma' < 0$. Now denote $F(x_m)$ by x'_m. Since F preserves the critical point as well as the order on $[x_s,\ x_p]$, we have by (i) $x'_m < \sigma' < 0$.

But $\Delta\omega \ll \min(\omega_m, \pi - \omega_m)$ implies that

$$x'_m = \frac{x_m - c}{1 - c x_m} \simeq O(x_m - c) = O(\Delta x^2).$$

So finally we have

$$\sigma' = O(\Delta x^2). \tag{4.21}$$

(iv) For the symmetric case in Lemma 11, when $x'_p \ll 1$,

$$G'(x', \infty) \simeq \sqrt{(x'_p)^2 - (x')^2} + O((x'_p)^2) \tag{4.22}$$

for all x' in $[-x'_p,\ x'_p]$. And since $z' \to x'_p/z'$ maps K' onto itself, Property 2 yields

$$G'(z',\ s') = G'(x'_p/z',\ x'_p/s'). \tag{4.23}$$

Therefore finally,

$$\begin{aligned}
\beta(\omega_m) &= G(\sigma,\ \infty) = G'(\sigma',\ s') \\
&= G'(0,\ s') + O(\Delta x^2) && [(4.21)] \\
&= G'(\infty,\ -x'_p c) + O(\Delta x^2) && [(4.23)] \\
&= G'(-x'_p c,\ \infty) + O(\Delta x^2) \\
&= \sqrt{(x'_p)^2 - (x'_p c)^2} + O(\Delta x^2) && [(4.22)] \\
&= x'_p\sqrt{1-c^2} + O(\Delta x^2) \\
&= \frac{x_p - c}{1 - cx_p}\sqrt{1-c^2} + O(\Delta x^2) \\
&= \frac{x_p - x_m}{1 - x_m^2}\sqrt{1-x_m^2} + O(\Delta x^2) && [(4.20)] \\
&= \frac{\Delta x}{2\sqrt{1-x_m^2}} + O(\Delta x^2) \\
&= \frac{\Delta\omega}{2} + O(\Delta\omega^2) \\
&= \beta(\frac{\pi}{2}) + O(\Delta\omega^2).
\end{aligned}$$

(v) The numerical results displayed in Figure 4-4 show that practically, in the whole range of $[\Delta\omega,\ \pi - \Delta\omega]$, $\beta(\omega_m) \simeq \beta(\pi/2)$ is a satisfactory approximation.

4.2 The Green's Function of Several Intervals and Its Asymptotics

4.2.1 Introduction

There are at least two natural occasions where domains with several intervals arise: the design of optimal bandpass filters (in the previous section, we have considered a special case: lowpass filters) (see also Fuchs, Kaiser and Landau [33, 1980]), and polynomial based matrix iterations (see Eiermann, Niethammer, and Varga [23, 1985], Eiermann, Li, and Varga [22, 1989], Freund [30, 1991], and Wathen, Fischer and Silvester [77, 1995] [78, 1997], for examples). Both cases involve the optimal Chebyshev polynomial approximations on several intervals, with or without constraints.

Their analysis requires the knowledge of the Green's function of the underlying domain (see Fuchs [31, 1978], Walsh [76, 1965], and Widom [80, 1969], for examples).

Design of Digital Bandpass Filters

In most occasions of signal processing and image processing, filters with *linear phases* are preferred (see Strang and Nguyen [69, 1996], and Oppenheim and Schafer [57, 1989]). H has *linear phase* if there exists a real polynomial $p(x)$ such that: $H(e^{\imath\omega}) = e^{\imath L\omega/2}p(\cos\omega)$, or equivalently, the coefficients sequence of $H(z)$ is symmetric around certain index $L/2$ (integer or half integer). For simplicity, assume $L = 0$ from now on. Such a filter is said to have *zero phase.*

An *ideal bandpass filter* D is defined for several given *bands* of interests: $J_1, J_2, \cdots, J_n$, n disjoint (closed) intervals of $[0, \pi]$. D is an even and 2π periodic $0-1$ function, and for $\omega \in [0, \pi]$

$$D(\omega) = \begin{cases} c_k, & \omega \in J_k \\ 0 & \text{else} \end{cases}$$

$c_k = 0$ or 1, depending on whether information contained in band J_k is noise or not (or for some other purposes). In practice, c_k are not all 0's, neither all 1's. The word "ideal" indicates that D cannot be realized by digital filters of finite length.

A natural question is: given a fixed length $N = 2m + 1$, which digital filter of zero phase gives the "best" approximation to the ideal one? The approximation error is usually evaluated by the uniform norm

$$\|D - H\|_J = \max_{\omega \in J} |D(\omega) - H(e^{i\omega})|,$$

where $J = \cup_{k=1}^{n} J_k$. The symmetry assumption eventually leads to the following standard polynomial approximation problem (with $x = \cos\theta$):

$$\text{Minimize } \|D^*(x) - p(x)\|_K \text{ over all polynomials } p(x) \text{ of degree } \leq m.$$

Here $K = \cos(J)$ and $D^*(x) = D(\cos^{-1} x)$. Notice that K is a subset of $[-1, 1]$, consisting of several disjoint intervals.

By Fuchs' result (Theorem 19), the convergence analysis of this approximation requires the Green's function of K. We will go back to this in section 4.2.4.

Polynomial Based Matrix Iteration

Here we introduce two classical examples of polynomial based matrix iteration methods in numerical linear algebra and discuss briefly how the Green's function of a several-interval domain plays an important role in convergence anaysis.

▶ **Semi-iterative method (SIM)**

The simplest and primitive iteration method for solving linear system of algebraic equations is based on the contracting mechanism:

$$\mathbf{x}_{n+1} = T\mathbf{x}_n + \mathbf{b},$$

with the spectral radius $\rho(T) < 1$. It solves $A\mathbf{x} = \mathbf{b}$, with $A = I - T$ usually discretized from some differential equation in a continuous model. The error vector $\mathbf{e}_n = \mathbf{x}_n - \mathbf{x}_*$ satisfies $\mathbf{e}_n = q(T)\mathbf{e}_0$, where $q(T) = T^n$ is a monomial and $\mathbf{x}_*$ is the unique solution.

This primitive iteration is far away from being optimal unless the spectrum $\Lambda(T)$ of T is scattered (almost) everywhere in the disk $|\lambda| \leq \rho(T)$. If the inclusion set K of $\Lambda(T)$ is not a disk, as in the example of Davis and Hageman [13, 1969] where K turns out to be a cross-shaped domain, then we can apply the so called *semi-iterative method* to improve acceleration (see Eiermann, Niethammer, and Varga [23, 1985]). From the signal processing point of view, it is a special **filtering** process: at each step n, with $\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_n$ at hand, we apply a lowpass filtering to them; namely, find a "good" polynomial (or filter) $p(z)$ of degree n, $p(z) = c_0 z^n + \cdots + c_n$, $p(1) = 1$ (the "lowpass" condition), by which, a new vector $\mathbf{y}_n$ is generated from filtering $\mathbf{x}_i$'s:

$$\mathbf{y}_n = c_0\mathbf{x}_n + c_1\mathbf{x}_{n-1} + \cdots + c_n\mathbf{x}_0.$$

Set $\delta_n = \mathbf{y}_n - \mathbf{x}_*$. Then $\boxed{\delta_n = p(T)\mathbf{e}_0}$. This key equation hints that $p(z)$ is "good" if and only if $\|p(T)\|$ can be small. (See Driscoll, Toh and Trefethen [20, 1996] for the most detailed discussion of related questions.) If, as in most cases in practice, we are only able to know that the spectrum is included in some domain K, then the best possible choice for $p(z)$ should be the solution of the following mini-max optimization (though obviously over-constrained):

$$\min_{p\ of degree\ n,\, p(1)=1} \max_{z\in K} |p(z)|.$$

The convergence analysis of SIM therefore inevitably involves the Green's function of K. Domains of several intervals are of particular interest in applications (see ELV [22, 1989]).

▶ **Minimal residual method (MR)**

MR is one example of the Krylov space methods and is also polynomial based. Here we only sketch the main idea for the simple version of MR (without preconditioning). To solve $A\mathbf{x} = \mathbf{b}$ (usually already transformed from the original system), one searches an optimal approximation solution $\mathbf{x}_n$ in the n-th Krylov space generated from $\mathbf{b}$:

$$\{\mathbf{b}, A\mathbf{b}, \cdots, A^{n-1}\mathbf{b}\}.$$

In MR method, the optimality is evaluated by the magnitude of the residue vector $\mathbf{r}_n = \mathbf{b} - A\mathbf{x}_n$. To minimize the norm of $\mathbf{r}_n$, it is equivalently to solve the following optimization problem:

$$\min_{p\ of\ degree\ n,\, p(0)=1} \|p(A)\mathbf{b}\|.$$

If, as in most cases, the spectrum $\Lambda(A)$ is only known to be included in some domain K, then the "best" possible choice of $p(z)$ is naturally the solution to the following mini-max problem:

$$\min_{p\ of\ degree\ n,\, p(0)=1} \max_{x\in K} |p(x)|.$$

General convergence analysis of the MR method depends on analysis of this polynomial optimization problem. If A is discretized from some self-adjoint differential operator, $\Lambda(A)$ is very often contained in the real line. Further information can restrict $\Lambda(A)$ to some intervals (as in the case of Wathen, Fischer, and Silvester [77, 1995]). This is why the Green's function of a several-interval domain can be very important in the convergence analysis of MR like methods.

The Green's Function

The analysis of these polynomial approximations requires the knowledge of the Green's function for a several-interval domain. This has motivated our research. To start, let us formulate the problem in an abstract way, temporarily forgetting those application backgrounds.

Given $2n$ points between -1 and 1:

$$-1 < a_1 < b_1 < a_2 < \cdots < a_n < b_n < 1,$$

we can define $n+1$ intervals:

$$K_1 = [-1, a_1],\ K_2 = [b_1, a_2],\ \cdots,\ K_{n+1} = [b_n, 1];\ \text{with } K = \cup_{j=1}^{n+1} K_j.$$

In between are n "gaps" (or "transition bands" in signal processing)

$$I_1 = (a_1, b_1), I_2 = (a_2, b_2), \cdots, I_n = (a_n, b_n); \text{ with } I = \cup_{k=1}^{n} I_k.$$

Thus $I \cup K = [-1, 1]$. $K^c = \mathbb{C} \backslash K$ denotes the complement of K in the complex plane. The Green's function $g(z)$ is the unique function with these properties: (1) $g(z)$ is harmonic on $K^c \backslash \infty$; (2) near $z = \infty$, $g(z) - \ln |z|$ is finite; (3) $g(z)$ is continuous up to the boundary of K^c (which is K in this case), and (4) $g(z) = 0$ on the boundary K. The main purpose of this paper is to study this function. For simplicity, $g(z)$ is directly called the Green's function of K.

Our main tool is the Schwarz-Christoffel mapping (SCM), which maps the upper half plane onto an arbitrary polygon domain. This SCM idea was first introduced in Trefethen, Embree and Mitchell [73, 1998] and has an important contribution to the study of the Green's function for a several-interval domain. For the sake of comparison, we first make some comments on some existing approaches in this subject. Eiermann, Li and Varga [22, 1989], and Shen and Strang [67, 1998] could only apply some elementary (polynomial) transforms, and study a special two-interval case, which requires some strong symmetry property and can be reduced to a single interval case. Freund [30, 1991] and Fischer [27, 1996] turned to a conformal mapping involving elliptic functions and converted the domain to an annulus, and then studied its Green's function. Wathen, Fischer and Silvester [77, 1995] [78, 1997] avoided the Green's function, but perturbed the two-interval case to a somewhat artificial one, for which a differential equation can be established and asymptotic analysis becomes possible. The SCM method is more elementary, universal, and better suited for asymptotic analysis.

We have emphasized the following two points in our presentation. First, we have singled out the two-interval case because of its importance in applications and its simplicity in analysis. Secondly, domains with narrow gap intervals are welcomed since asymptotic analysis can lead to simple and useful leading terms. We also borrow some ideas from Hilbert space theory and probability theory, which help inspire deeper insights into this analytic problem.

Organization

This part has been organized as follows. Section 4.2.2 studies the Green's function for the two-interval case. Section 4.2.3 discusses the Green's function and related objects for a general domain with several intervals. In Section 4.2.4, we demonstrate two applications in digital filter design and the numerical analysis for the Stokes equation. Our main contribution in this part is the discovery of the so-called "Square Root Law."

4.2.2 The Green's Function for Two Intervals

In this section, we study the Green's function for the two-interval case:

$$K = [-1, a] \cup [b, 1], \qquad -1 < a < b < 1.$$

There are two reasons for singling out this case. First, the two-interval case itself is frequently encountered and important in practice. Secondary, the analysis is relatively simple compared to that for more than two intervals, yet it already contains all the required mechanisms. For more examples and computational results, we refer to Trefethen, Embree and Mitchell [73, 1998].

The SCM and Green's Function

The central idea is to use the symmetry of K in the vertical direction to reduce the doubly connected domain K^c to a simply connected one: the upper-half plane.

First, in the upper-half plane, define a one-parameter SCM family by

$$w = \phi_s(z) = \int_a^z \frac{(s-u)du}{\sqrt{(1-u^2)(b-u)(u-a)}}, \qquad s \in I = (a, b).$$

We take the $\sqrt{}$ branch that is positive for all $u \in I$. Under this choice, the image of the gap I is always a subset of the real line in the w-plane. Moreover, ϕ_s maps the upper-half plane onto a polygon domain in the w-plane, whose vertices are ∞, and the images of -1, a, s, b, and 1, denoted by C, A, S, B, D. By the general theory of SCM, the interior angles of the polygon at C, A, S, B, D are $\pi/2, \pi/2, 2\pi, \pi/2$, and $\pi/2$. With the knowledge that $\phi_s(a) = A = 0$ and $\phi_s((a, s))$ lies inside the positive axis, we conclude that the polygon has the following shape and orientation:

There is a critical parameter $s = \sigma$, which is needed to achieve $B = A = 0$. This amounts to requiring

$$0 = \phi_\sigma(b) = \int_a^b \frac{(\sigma - x)dx}{\sqrt{(1-x^2)(b-x)(x-a)}}.$$

This real integral produces the following probability interpretation of the critical parameter σ.

Proposition 11 (The critical parameter) *Define the number γ by the elliptic-like integral*

$$\gamma = \int_a^b \frac{dx}{\sqrt{(1-x^2)(b-x)(x-a)}}.$$

Let X be a random variable supported in (a, b) and with probability density function given by $\rho(x) =$

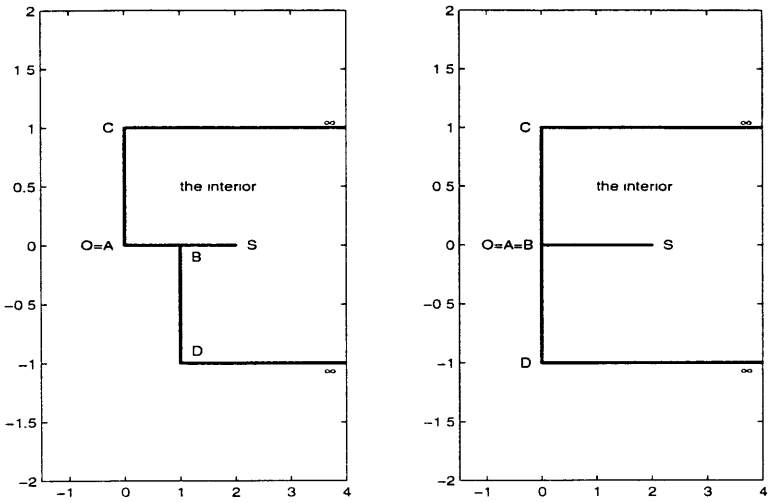

Figure 4-8: The image of the upper-half plane under a general SC mapping ϕ_s (left) and ϕ_σ with the ciritical parameter σ (right).

$[(1-x^2)(b-x)(x-a)]^{-1/2}/\gamma$. *Then* $\sigma = \mathbf{E}\{X\}$, *the mean value (or expectation) of* X.

For this particular σ, denote ϕ_σ simply by ϕ. Define

$$g(z) = \begin{cases} \mathrm{Re}\phi(z) & \mathrm{Im}z \geq 0 \\ g(\overline{z}) & \mathrm{Im}z < 0 \end{cases}.$$

Proposition 12 *$g(z)$ is the Green's function of K.*

Proof. From the definition, $\phi(K)$ is a subset of the imaginary axis (see Figure 4-8). Hence $g(K) = \{0\}$. Since $\phi(\mathbb{R}\backslash K)$ consists of some horizontal open lines, $\phi(z)$ can be analytically continued locally near any $x \in \mathbb{R}\backslash K$, by Schwarz's Reflection Principle. From this it is easy to see that $g(z)$ is harmonic on K^c. Finally, from our choice of the $\sqrt{}$ branch, $\phi(z) = \ln z + c_0 + c_1/z + \cdots$, near $z = \infty$. Hence $g(z) - \ln|z|$ is finite near $z = \infty$. This shows that $g(z)$ is the Green's function for K. □

Recall that $z = z_0$ is a *critical point* of $g(z)$ if the level line through z_0 is self-intersected, or equivalently, the gradient of $g(z)$ vanishes at z_0 .

Corollary 16 (The Green's function on the gap) *For all $x \in I = (a, b)$,*

$$g(x) = \int_a^x \frac{(\sigma - t)dt}{\sqrt{(1-t^2)(b-t)(t-a)}}$$

Especially, σ is the unique critical point *of $g(z)$.*

Proof. For all $x \in I$, $\phi(x)$ is real. Hence $g(x) = \phi(x)$ is given by the above integral. The same integral shows $\partial g/\partial x(\sigma) = 0$. Since $\partial g/\partial y(\sigma) = 0$ holds automatically by the vertical symmetry of $g(z)$, σ is a cirtical point of $g(z)$. Uniqueness follows from the fact that the Green's function for an $n+1$-multiply connected domain has exactly n critical points (see Nevanlinna [55, 1970]). □

Asymptotics for a Small Gap I

Domains with small gap intervals arise from both digital filter design and matrix iterations. Define the midpoint and the half-width:

$$c = \frac{a+b}{2}, \qquad \delta = \frac{b-a}{2}.$$

From now on, we assume that c belongs to a **fixed** compact set of $(-1,\ 1)$, and $\delta \to 0$.

The following change of variable is useful. For any $x \in I$, set $\theta = (x-c)/\delta$. Then for any $f(t)$,

$$\int_a^x \frac{f(t)dt}{\sqrt{(1-t^2)(b-t)(t-a)}} = \int_{-1}^{\theta} \frac{f(c+\delta s)ds}{\sqrt{(1-s^2)(1-(c+\delta s)^2)}}. \tag{4.24}$$

Lemma 12 *The location of the critical point is*

$$\sigma = c + O(\delta^2).$$

A similar result was also proved in Wathen, Fischer and Sylvester [77, 1995], but in a quite different context.

Proof. [Proof of Lemma 12] Following the notation of Proposition 11, and using Eq. (4.24), we have

$$\sigma - c = \mathbf{E}\{X - c\} = \frac{1}{\gamma}\int_{-1}^{1} \frac{ds}{\sqrt{1-s^2}} \frac{\delta s}{\sqrt{1-(c+\delta s)^2}}.$$

Since c is assumed to be in a compact set of $(-1, 1)$, the following infinite series converges uniformly for small ϵ:

$$\frac{\epsilon}{\sqrt{1-(c+\epsilon)^2}} = c_1\epsilon + c_2\epsilon^2 + \cdots$$

Therefore

$$\sigma - c = \delta^2 \frac{c_2}{\gamma} \int_{-1}^{1} \frac{s^2 ds}{\sqrt{1-s^2}} + \cdots .$$

Since γ and c_k are both of order $O(1)$ (from the assumption that c belongs to a compact set of $(-1,1)$), we complete the proof. □

Proposition 13 (The square root law) *Uniformly for all $x \in I = (a,b)$, and c in a compact set of $(-1,1)$,*

$$\begin{aligned} g(x) &= \frac{1}{\sqrt{1-c^2}} \sqrt{(b-x)(x-a)} + O(\delta^2) \\ &= \sqrt{(\omega - \omega_b)(\omega_a - \omega)} + O(\Delta\omega^2). \end{aligned}$$

Here $\omega = \cos^{-1} x$, $\omega_a = \cos^{-1} a$, $\omega_b = \cos^{-1} b$, and $\Delta\omega = \omega_a - \omega_b$ (the "transition bandwidth" in signal processing).

Proof. The second line follows from the first by a change of variables to $x = \cos\omega$.

$$\begin{aligned} g(x) &= \int_{-1}^{\theta} \frac{(\sigma - c) - \delta s}{\sqrt{(1-s^2)(1-(c+\delta s)^2)}}\, ds. \\ &= -\delta \int_{-1}^{\theta} \frac{s ds}{\sqrt{(1-s^2)(1-(c+\delta s)^2)}} + O(\delta^2). \end{aligned}$$

Lemma 12 has been applied to the last step. Suppose

$$\frac{1}{\sqrt{1-(c+\epsilon)^2}} = \frac{1}{\sqrt{1-c^2}} + c_1 \epsilon + \cdots .$$

Then

$$g(x) = -\frac{\delta}{\sqrt{1-c^2}} \int_{-1}^{\theta} \frac{s ds}{\sqrt{1-s^2}} + O(\delta^2) = \frac{\delta}{\sqrt{1-c^2}} \sqrt{1-\theta^2} + O(\delta^2).$$

This completes the proof since $\delta^2(1-\theta^2) = (b-x)(x-a)$. □

In some cases, we have to allow $\epsilon = 1 - c$, or $1 + c$ to be small too. Set $r = \delta/\epsilon$. A modification of the above proof leads to the following stronger version.

Theorem 27 *Suppose $r \ll 1$. Then uniformly for all $x \in (a,b)$,*

$$g(x) = \frac{1}{\sqrt{1-c^2}} \sqrt{(b-x)(x-a)} + O(r^2\sqrt{\epsilon}).$$

4.2.3 The Green's Function for Several Intervals

In this section, we study the Green's function for several intervals. As already introduced in the introduction section, K is the union of $n+1$ disjoint intervals $K_1, K_2, \cdots, K_{n+1}$, and I is the union of all gaps $I_1, I_2, \cdots, I_n$. Thus $K \cup I = [-1, 1]$. Let $g(z)$ be the Green's function for K. Since K^c is $n+1$-multiply connected, $g(z)$ must have n critical points, say $\sigma_1 < \sigma_2 < \cdots < \sigma_n$. The symmetry of the domain implies that all critical points lie along the real axis. It is not hard to see that thereis one critical point in each gap (a_k, b_k).

The configuration polynomial, critical polynomial, and Green's function

The *configuration polynomial* is

$$Q(z) = (z^2 - 1) \prod_{k=1}^{n} (z - a_k)(z - b_k)$$

This is a monic polynomial of degree $2n+2$ and contains all the information of K. One useful property of the polynomial is that Q is positive on all n gaps, and in fact

$$Q(x) > 0 \qquad \text{for all } x \in \mathbb{R} \backslash K.$$

The *critical polynomial* is

$$P(z) = (z - \sigma_1)(z - \sigma_2) \cdots (z - \sigma_n).$$

This is also monic and of degree n. Suppose

$$P(z) = z^n - e_1 z^{n-1} + e_2 z^{n-2} - \cdots + (-1)^{n-1} e_n.$$

Then $e_k = e_k(\sigma_1, \sigma_2, \cdots, \sigma_n)$ is the k-th elementary symmetric function of the σ_k's.

Theorem 28 (The Green's function) *For $x \in I_k = (a_k, b_k)$,*

$$g(x) = (-1)^{n+1-k} \int_{a_k}^{x} \frac{P(t)}{\sqrt{Q(t)}} dt.$$

Proof. The idea is exactly the same as in the two-interval case. Here we only outline the proof.

(a) In the upper-half plane, define

$$w = \phi(z) = (-1)^{n+1-k} \int_{a_k}^{z} \frac{P(u)}{\sqrt{Q(u)}} du.$$

Take the square root component that is positive for all $u \in I_k = (a_k, b_k)$. Then $w = \phi(z)$ is a SCM which maps the upper-half plane onto a polygon domain in the w-plane. The vertices of the polygon are ∞ and the images of ± 1, a_j, b_j, and $\sigma_j, j = 1, \cdots, n$. The interior angles at all the images of the critical points are 2π, and at ∞, 0. All the rest interior angles are $\pi/2$. Since $\phi(a_k) = 0$ and $\phi(x)$ is positive for all $x \in (a_k, \sigma_k)$, the orientation and shape of the polygon domain in the w-plane must look like the right subfigure in Figure 4-8, except this time we have more than one horizontal slits, digging into the interior of the polygon domain.

(b) Define $g(z)$ exactly as we did for the two-interval case. Then $g(K) = \{0\}$; $g(z)$ is harmonic on K^c, and $g(z) - \ln|z|$ is finite near $z = \infty$. Hence $g(z)$ is the Green's function and it is identical with $\phi(x)$ for $x \in I_k$.

□

Critical polynomial: a linear algebra approach

In this subsection, we illustrate one way to compute the critical polynomial $P(x)$ for a given configuration polynomial $Q(x)$. By finding the roots of $P(x)$, we can then find all the critical points of the Green's function, which are very important in some applications. In the next subsection, we provide another geometric way to compute it.

Because the Green's function $g(x)$ vanishes on K, we have

$$\int_{a_k}^{b_k} \frac{P(t)}{\sqrt{Q(t)}} dt = 0, \tag{4.25}$$

for all $k = 1, \cdots, n$. Assume

$$P(x) = x^n - e_1 x^{n-1} + e_2 x^{n-2} + \cdots + (-1)^n e_n.$$

Then we have the following characterization theorem.

Theorem 29 $\mathbf{c} = (e_1, -e_2, \cdots, (-1)^{n-1} e_n)'$ *is the unique solution to the* n *by* n *linear system* $M\mathbf{c} = \mathbf{b}$. *Here the configuration matrix* $M = (M_{jk})$ *and vector* $\mathbf{b} = (\mathbf{b}_j)$ *are defined by*

$$M_{jk} = \int_{I_j} \frac{t^{n-k} dt}{\sqrt{Q(t)}}, \qquad \mathbf{b}_j = \int_{I_j} \frac{t^n dt}{\sqrt{Q(t)}}.$$

Proof. By applying Eq.(4.25) for $k = 1, 2, \cdots, n$, the coefficient vector $\mathbf{c}$ is easily seen to solve $M\mathbf{c} = \mathbf{b}$. Uniqueness follows from the following lemma. □

Lemma 13 *The configuration matrix is non-singular.*

Proof. Otherwise, we can find a non-zero polynomial $q(t)$ of degree no more than $n-1$ such that

$$\int_{I_k} \frac{q(t)dt}{\sqrt{Q(t)}} = 0$$

for $k = 1, 2, \cdots, n$. Therefore $q(t)$ must change its sign on each gap I_k, which implies at least one zero on each gap. But $q(t)$ cannot have n zeros. Contradiction! □

Critical polynomial: a geometric approach

In this section, we compute the critical polynomial in a geometric way based on orthogonalization and projection in a certain L^2 space.

Take the gap set $I = I_1 \cup \cdots \cup I_n$ as the underlying space for a measure. Define the measure $d\mu$ by $d\mu = [Q(t)]^{-1/2}dt$. Then $(I, d\mu)$ is a finite measure space. In what follows, we always work in the Hilbert space $L^2(I, d\mu)$ with inner product $\langle \cdot, \cdot \rangle$.

Let $\chi_k(t)$ be the indicator function of I_k. Consider the following two sets of vectors in $L^2(I, d\mu)$:

$$\{1, t, \cdots, t^{n-1}\} \qquad \text{and} \qquad \{\chi_1, \chi_2, \cdots, \chi_n\}.$$

By the preceding lemma, both are linearly independent sets. The linear space $\mathbb{P}_{n-1}$ spanned by the first set contains all polynomials of degree no more than $n-1$.

Under this setting, the configuration matrix M is given by $M_{jk} = \langle \chi_j,\ t^{n-k} \rangle$. The non-singularity of M implies the existence of a dual basis in $\mathbb{P}_{n-1}$:

Corollary 17 *There exists a unique set of vectors* $\{q_1, \cdots, q_n\}$ *in* $\mathbb{P}_{n-1}$ *that is dual to* $\{\chi_1, \cdots, \chi_n\}$*:*

$$\langle \chi_j,\ q_k \rangle = \delta_{jk}, \qquad 1 \le j, k \le n.$$

With the dual basis, the critical polynomial can be computed explicitly.

Proposition 14 *The critical polynomial is given by*

$$P(t) = t^n - \sum_{k=1}^{n} \langle t^n,\ \chi_k \rangle q_n(t).$$

The critical polynomial can therefore be computed from the dual basis $\{q_1, \cdots, q_n\}$. Let S denote any non-empty subset of $[n] = \{1, 2, \cdots, n\}$. For each subset S of k elements, we shall define a monic polynomial $P_S(t)$ of degree k, subject to

$$\langle P_S,\ \chi_j \rangle = 0 \qquad \text{for any } \ j \in S. \tag{4.26}$$

This is realized by the following inductive projection algorithm.

Step 1 For any subset $S = \{j\}$ of one element, define

$$P_S(t) = t - \frac{\langle t,\ \chi_j \rangle}{\mu(I_j)}.$$

Obviously $\langle P_S,\ \chi_j \rangle = 0$.

Step k Suppose at the end of Step $k-1$, we have defined all polynomials $P_S(t)$ subject to condition (4.26), for a subset S with $k-1$ elements. For any subset S with k elements, define

$$P_S(t) = t^k - \sum_{j \in S} \frac{\langle t^k,\ \chi_j \rangle}{\langle P_{S\setminus j},\ \chi_j \rangle} P_{S\setminus j}(t).$$

This is well-defined since $\langle P_{S\setminus j},\ \chi_j \rangle$ cannot be zero ($P_{S\setminus j}$ has no zero on I_j). Obviously P_S satisfies condition (4.26).

From the above construction, it is easy to see that

Proposition 15 *$P_{[n]}(t)$ is the critical polynomial. After normalization, the dual basis of $\{\chi_1, \cdots, \chi_n\}$ consists of the polynomials*

$$P_{[n]\setminus 1},\ \cdots,\ P_{[n]\setminus n}.$$

Besides its role in characterizing the critical polynomial and the dual basis, this algorithm also works efficiently in practice for a small number of intervals, typically, for $n = 2, 3, 4$. For large n, the algorithm is in no way economic since at least $2^n - 1$ polynomials are to be computed.

Asymptotics for a small gap I_j

The square root law (Proposition 13) for a small gap still holds for several intervals.

Let us fix an index j. Set $c_j = \frac{1}{2}(a_j + b_j)$ and $\delta_j = \frac{1}{2}(b_j - a_j)$. For simplicity, we assume that all the other gaps $I_k : k \neq j$ are fixed (this restriction can be easily relaxed to include other cases),

and c_j belongs to a compact set of (b_{j-1}, a_{j+1}), and $\delta_j \to 0$. Define

$$Q_j(x) = \frac{Q(x)}{(1-x^2)(b_j - x)(x - a_j)}, \qquad P_j(x) = \prod_{k \neq j}(x - \sigma_k).$$

Then we have the following square root law. Its proof is similar to Proposition 13 and has been omitted here.

Proposition 16 (Square Root Law) *Uniformly for all $x \in I_j = (a_j, b_j)$, and c_j in a compact set of (b_{j-1}, a_{j+1}),*

$$g(x) = \frac{|P_j(c_j)|}{\sqrt{Q_j(c_j)}} \sqrt{(\omega - \omega_j^b)(\omega_j^a - \omega)} + O(\delta_j^2).$$

Here $\omega = \cos^{-1} x$, $\omega_j^a = \cos^{-1} a_j$, and $\omega_j^b = \cos^{-1} b_j$.

Similar results can also be established for the delicate cases when c_j approaches b_{j-1} or a_{j+1}.

4.2.4 Applications of the Square Root Law

In this section, we apply our results to two problems: design of optimal (equiripple) lowpass filters, and the convergence analysis of the minimum residual (MR) method for solving the Stokes equation in fluid dynamics. We anticipate more applications in other fields.

Design of equiripple lowpass filters

Our first application is to give a simple proof of Theorem 23.

Recall that an equiripple lowpass filter has only two bands of frequencies on $[0, \pi]$: the passband $[0, \omega_p]$ and the stopband $[\omega_s, \pi]$. The ideal lowpass filter $D(\omega)$ equals 1 on the passband and 0 on the stopband. With $x = \cos\omega$, the polynomial approximation problem for each n is:

$$\text{Minimize} \quad \|D^*(x) - p(x)\|_K \text{ over all polynomials of degree } \leq n.$$

$K = [-1, x_s] \cup [x_p, 1]$, $x_p = \cos(\omega_p)$, and $x_s = \cos(\omega_s)$. The norm is the L^∞ norm. The optimal error is of order $O(n^{-1/2}e^{-n\beta})$, where β is the value of the Green's function of K at the unique critical point σ. Set $\omega_m = \frac{1}{2}(\omega_p + \omega_s)$. $\Delta\omega = \frac{1}{2}(\omega_s - \omega_p)$ (the transition bandwidth). Then β is a function of ω_m and $\Delta\omega$. We denote it by $\beta(\omega_m, \Delta\omega)$. Now we are ready to give a simple proof to Theorem 23:

Theorem 23. *In the range of $\Delta\omega \ll \min(\omega_m, \pi - \omega_m)$, the leading term of $\beta(\omega_m, \Delta\omega)$ is $\beta(\pi/2, \Delta\omega)$.*

Proof. Without loss of generality, assume $\omega_m \in [0, \pi/2]$. From our Theorem 27,

$$g(x) = \sqrt{(\omega - \omega_p)(\omega_s - \omega)} + O(r^2\sqrt{\epsilon}).$$

Since $\Delta\omega = O(\delta/\sqrt{\epsilon})$, and $\omega_m = O(\sqrt{\epsilon})$, we have

$$O(r^2\sqrt{\epsilon}) = O(\frac{\Delta\omega^2}{\omega_m}),$$

which is an order smaller than $O(\Delta\omega)$ when $\Delta\omega \ll \min(\omega_m, \pi - \omega_m)$. Therefore

$$\beta = g(\sigma) = \max_{x_s \le x \le x_p} g(x) = \max_{\omega_p \le \omega \le \omega_s} \sqrt{(\omega - \omega_p)(\omega_s - \omega)} + o(\Delta\omega) = \frac{\Delta\omega}{2} + o(\Delta\omega).$$

Hence the leading term of β is independent of ω_m. Especially one can take ω_m to be $\pi/2$. □

For the symmetric case ($\omega_m = \pi/2$), Shen and Strang showed that

$$\beta(\pi/2, \Delta\omega) = \ln \cot \frac{\pi - \Delta\omega}{4}.$$

For small $\Delta\omega$, it again gives $\Delta\omega/2$ as the leading order. Numerical evidence showed that taking $\ln \cot \frac{\pi - \Delta\omega}{4}$ as an approximation to $\beta(\omega_m, \Delta\omega)$ is better than $\Delta\omega/2$.

Estimation of asymptotic convergence factor

Wathen, Fischer and Silvester [77, 1995] studied the numerical solution of the classical Stokes problem of fluid dynamics:

$$\begin{aligned} -\nabla^2 \mathbf{u} + \text{grad } p &= \mathbf{f} \qquad \text{in} \quad \Omega, \\ \text{div } \mathbf{u} &= 0 \qquad \text{in } \Omega. \end{aligned}$$

With suitable boundary conditions, the equation is usually discretized (by the Finite Element Method, say) and stabilized to a linear system of equations of the form

$$\begin{pmatrix} A & B^T \\ B & -\beta C \end{pmatrix} \cdot \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix},$$

or simply $\mathcal{A}x = b$. The linear system is symmetric and *indefinite*; the matrix $\mathcal{A}$ has both positive and negative eigenvalues. For such an indefinite system, the minimum residual (MR) iteration method is perferred to the conjugate gradient method.

After preconditioning, Wathen and Silvester [79, 1993] showed that the spectrum of the discrete Stokes operator is included in two intervals:

$$K_h = [-a, -bh] \cup [ch^2, d], \qquad a, b, c, d, h > 0.$$

Here h is the mesh size for discretizing the underlying domain Ω. The asymptotic convergence factor ρ is given by:

$$\rho = \exp(-g(0)).$$

Here $g(x)$ is the Green's function of K_h. The main result of Wathen, Fischer and Silvester is the following.

Theorem 30 (Wathen, Fischer and Silvester [77, Theorem 4.1])

$$\rho \leq 1 - \sqrt{bc/ad}\, h^{3/2} + O(h^{5/2}).$$

The proof strategy was based on the equiripple property of the optimal polynomial $p_n(x)$, which is as small as possible on K under the constraint $p_n(0) = 1$. By perturbing the interval a little bit, $p_n(x)$ can have $n+2$ extremal points. This makes it possible to establish a differential equation and carry out some asymptotic analysis.

Here we use our square root law in Section 4.2.2 to give a simple proof of

$$\rho = 1 - \sqrt{bc/ad}\, h^{3/2} + O(h^2). \tag{4.27}$$

Proof. To apply Proposition 13, we normalize the set K_h by introducing

$$z_* = \psi(z) = -1 + 2\frac{z+a}{d+a}.$$

$\psi(z)$ maps K_h to

$$K_* = [-1, a_*] \cup [b_*, 1], \qquad a_* = \psi(-bh), b_* = \psi(ch^2).$$

The gap size is

$$2\delta_* = b_* - a_* = \psi'(0)(ch^2 - (-bh)) + O(h^2) = \frac{2bh}{a+d} + O(h^2),$$

or, $O(\delta_*) = O(h)$. The center point is $c_* = \psi(0) + O(h)$.

Let $g_*(z_*)$ denote the Green's function for the normalized domain. Then

$$\begin{aligned} g(0) = g_*(\psi(0)) &= \frac{1}{\sqrt{1-c_*^2}} \sqrt{[\psi(0) - \psi(-bh)]\,[\psi(ch^2) - \psi(0)]} + O(h^2) \\ &= \frac{1+O(h)}{\sqrt{1-\psi^2(0)}} \sqrt{\psi'(0)bh \cdot \psi'(0)ch^2 + O(h^4)} + O(h^2) \\ &= \frac{\psi'(0)\sqrt{bch^3}}{\sqrt{1-\psi^2(0)}} (1+O(h)) + O(h^2) \\ &= \sqrt{bc/ad}\, h^{3/2} + O(h^2). \end{aligned}$$

Therefore

$$\rho = \exp(-g(0)) = 1 - \sqrt{bc/ad}\, h^{3/2} + O(h^2).$$

□

Similarly, by normalizing the domain and applying Theorem 27, one can give a short proof to another result.

Theorem 31 (Wathen, Fischer and Silvester [78, Theorem 5.1])
If the domain is $K_h = [-ah^L, -bh^l] \cup [ch^r, d]$ *for some positive* a, b, c, d, *with* $L < r$, *and* $L < l$, *then the leading term is*

$$\rho \simeq 1 - \sqrt{bc/da} h^{(r+l-L)/2}.$$

4.3 The Equilibrium Distribution and Asymptotics of Extremal Points

4.3.1 The Potential and Equilibrium Distribution

What is closely related to the Green's function is the *equilibrium distribution* of K. In this section, we give an explicit expression for the equilibrium distribution when K consists of several disjoint compact intervals.

Let μ be any unital distribution (probability measure) on K (built on the Borel algebra). The *potential* generated by μ is

$$V_\mu(z) = \int_K \ln|z-s|\,\mu(ds).$$

The potential must be subharmonic on $\mathbb{C}$ and harmonic on the complement of K.

The *total energy* generated by μ is

$$E_\mu = -\int_K V_\mu(s)\,\mu(ds).$$

The *equilibrium distribution* ν is a unital distribution that minimizes the total energy. The potential associated to the equilibrium distribution is called the *equilibrium potential.*

Such a distribution is not only physically important but also mathematically useful. For a "regular" domain like that in our case, it uniquely exists. Frostman's Theorem gives a characterization of the equilibrium potential.

Theorem 32 (Frostman's Theorem) *Let K be a "regular" compact set of $\mathbb{C}$ and ν the equilibrium distribution of K. Then*

(i) $V_\nu(z) \geq -E_\nu$ *on* $\mathbb{C}$.

(ii) $V_\nu(z) \equiv -E_\nu$ *on* K.

Conversely, a subharmonic function $V(z)$ with the following two properties must be the equilibrium potential.

(i) $V(z)$ is harmonic on the complement of K and $V(z) - \ln|z| = o(1)$ near $z = \infty$.

(ii) $V(z) = -E$ for all $z \in K$ and a certain constant E.

The inverse problem is solved by the generalized Laplacian: $\nu = \Delta V/2\pi$ (in the sense of generalized functions).

The Green's function $g(z)$ and the equilibrium potential $V_\nu(z)$ are almost identical:

$$g(z) = V_\nu(z) + E_\nu.$$

Our main result of this section is the following theorem.

Theorem 33 *Following the notation of section 4.2.3, let $Q(z)$ and $P(z)$ denote the configuration polynomial and critical polynomial. Then the equilibrium distribution ν is supported on K and given by*

$$\nu(dx) = C\frac{|P(x)|}{\sqrt{|Q(x)|}}dx,$$

where the positive constant C normalizes ν to be unital. Especially, in terms of the Schwarz-Christoffel mapping $\phi(z)$, ν is the pullback (by $\phi(z)$) of the uniform (unital) distribution on $\phi(K)$ (the purely imaginary edge of the polygon domain (see Figure 4-8).

Proof. Let us prove the two-interval case. The general case is exactly the same.

Suppose $K = [-1, a] \cup [b, 1]$ and the equilibrium distribution is given by

$$\nu(dx) = \rho(x)dx.$$

For all x in the gap $I = (a, b)$,

$$V_\nu(x) = \int_K \ln|x - t|\, \rho(t)dt,$$

and the Green's function

$$g(x) = \int_a^x \frac{\sigma - t}{\sqrt{Q(t)}}\, dt.$$

Differentiating $g(x) = V_\nu(x) + E_\nu$ yields

$$\frac{\sigma - x}{\sqrt{Q(x)}} = \int_K \frac{\rho(t)}{x - t}\, dt$$

for all $x \in (a, b)$. Now define two analytic functions on $\mathbb{C}\backslash K$:

$$\Phi_1(z) = \frac{\sigma - z}{\sqrt{Q(z)}}, \qquad \Phi_2(z) = \int_K \frac{\rho(t)}{z - t}\, dt.$$

Take $\sqrt{Q}$ to be positive on $I = (a, b)$. Since $\Phi_1(z) = \Phi_2(z)$ on I, $\Phi_1(z) \equiv \Phi_2(z)$ for all $z \in \mathbb{C}\backslash K$. Notice that $\Phi_2(z)$ is a Cauchy integral. Therefore, for any $x \in K$ (excluding the end points),

$$\rho(x) = \frac{i}{2\pi}\Big(\Phi_2(x^+) - \Phi_2(x^-)\Big),$$

where $\Phi_2(x^+) = \lim_{\delta \to 0^+} \Phi_2(x + i\delta)$. Hence

$$\rho(x) = \frac{i}{2\pi}\Big(\Phi_1(x^+) - \Phi_1(x^-)\Big) = \frac{1}{\pi}\frac{|\sigma - x|}{\sqrt{|Q(x)|}}.$$

This completes the proof of the first part. For the remaining part, simply notice that

$$d\phi = \frac{\sigma - z}{\sqrt{Q(z)}}, \qquad \text{and so} \qquad \nu(dx) = C|d\phi|.$$

□

Remark. We have noticed that Peherstorfer [59, 1990] also obtained the first part of the theorem based on Widom's formula [80, 1969] for the complex Green's function. The proof presented here avoids the multivalue problem caused by the multi-connectivity of K, by first considering the restriction of the Green's function on a gap. The first part can also be obtained from Geronimo and Van Assche's result on polynomial mappings [35, 1988]. The second part gives for the first time a clear geometric meaning to the equilibrium measure. □

An asymptotic result for the two-interval case can thus be established based on this theorem and Lemma 12 (on the location of σ).

Corollary 18 *Suppose* $K = [-1, a] \cup [b, 1]$, *and* a, b *are contained in a fixed compact set of* $(-1, 1)$. *Let* $\delta = (b - a)/2 \ll 1$. *Then for any* c, d*:*

$$-1 \le c < d < a \qquad or \qquad b < c < d \le 1,$$

with $\epsilon = \min\{|a - d|, |c - b|\} \gg \delta$,

$$\nu[c, d] = (\omega_c - \omega_d)/\pi + O((\delta/\epsilon)^2),$$

where $\omega_c = \cos^{-1} c$ *and* $\omega_d = \cos^{-1} d$.

4.3.2 Asymptotics of Extremal Points and Its Applications

The convergence analysis of the matrix iteration problem has little to do with the equilibrium distribution. But the filter design problem has a lot.

The current design of optimal lowpass filters or bandpass filters is realized by the Remez-Parks-McClellan exchange algorithm. The algorithm needs improvement in at least two aspects. First, it has no recursive structure, which forces one to rerun the program to get a filter of length 65 even when the optimal one of length 33 has already been available. Secondly, the efficiency of the algorithm can be improved if one starts it with an initially guessed set of extremal points that are very close to the real ones.

These two problems are closely connected, guiding us to investigate the distribution pattern of

the extremal points, or equivalently, the zeros of the error function $r_n(x) = D(x) - p_n(x)$, where $p_n(x)$ is the optimal polynomial of degree n and $D(x)$ the ideal lowpass filter. Suppose one knows enough information about the pattern, then it is possible to design an approximately optimal filter by a once-and-for-all interpolation. It is also possible to give a good set of initial points for the exchange algorithm and lessen the correction work in the algorithm.

This has been the major motivation of Fuchs paper [32, 1980] and the current section.

Let $F_e(x)$ denote the cumulative distribution function(c.d.f.) of the equilibrium distribution:

$$F(x) = \nu(-\infty, x].$$

Let $y_0, y_1, \cdots, y_{n+1}$ denote the $n+2$ extremal alternating points of the optimal error $r_n(x)$. These are a set of points on K satisfying:

$$r_n(y_i) = \pm\|r_n\|, \qquad r_n(y_i)r_n(y_{i+1}) < 0.$$

By assigning each point y_i a measure $1/(n+2)$, we can define another c.d.f.:

$$F_n(x) = \sum_{i:y_i \le x} \frac{1}{n+2}.$$

Fuchs' main result is

Theorem 34 (Fuchs [32, Theorem 3]) *Uniformly for all $x \in \mathbb{R}$,*

$$F_n(x) - F_e(x) = O(n^{-1/5}), \qquad as \ \ n \to \infty.$$

The following two results are obvious from it.

Corollary 19 *Let $Z_n(x)$ denote the c.d.f for the zeros of $r_n(x)$ (assigning each zero a measure $1/n$). Then uniformly for all $x \in \mathbb{R}$,*

$$Z_n(x) - F_e(x) = O(n^{-1/5}), \qquad as \ \ n \to \infty.$$

This is because between each pair of y_i and y_{i+1} (excluding one i), there is exactly one zero.

Corollary 20 *Let $H = [c, d]$ be any interval contained in K. Then the portion of the extremal points (or zeros) on H is $\nu[c, d] + O(n^{-1/5})$.*

Especially, for the two-interval case, as $n \to \infty$,

$$\#\text{ of zeros on } [-1, a] \ : \ \#\text{ of zeros on } [b, 1] \to \int_{-1}^{a} \frac{\sigma - x}{\sqrt{|Q(x)|}} dx \ : \ \int_{b}^{1} \frac{x - \sigma}{\sqrt{|Q(x)|}} dx.$$

Furthermore, if the gap (a, b) is narrow, under the condition of Corollary 18, we have

Proposition 17 *The portion of extremal points (or zeros) on* $[c, d]$ *is*

$$(\omega_c - \omega_d)/\pi + O\Big((\delta/\epsilon)^2 + n^{-1/5}\Big).$$

Under the Schwarz-Christoffel mapping $\phi(z)$, the equilibrium distribution of K becomes the uniform distribution on the imaginary edge $\phi(K)$. Therefore, the zeros of $r_n(x)$ can be approximated by the preimages of any n points equidistributed along $\phi(K)$. The resulted interpolation leads to *nearly optimal filters* in the sense of section 4.1.7. For numerical examples and more discussions, see Trefethen, Embree and Mitchell [73, 1998].

4.3.3 Summary

The work presented in section 4.2 and 4.3 is an extension of that in Trefethen, Embree and Mitchell [73, 1998]. Based on the SCM idea first appearing in Widom [80, 1969] and later rediscovered in the above paper, we study closely the properties of the Green's function for a real several-interval domain. The undetermined parameters in the mapping happen to be the critical points of the Green's function. By introducing the critical polynomial and configuration polynomial, we can determine those unknown parameters either by solving a linear system of equations or an inductive projection process.

On a narrow gap interval, the Green's function behaves like the square root of a quadratic polynomial. This "square root law" is applied to a problem from digital filter design and a problem from computational fluid dynamics.

We have also studied the equilibrium distribution and its properties. Most of the research has been motivated by the optimal design of digital filters.

Bibliography

[1] N.I. Akhiezer. *Elements of the theory of elliptic functions.* AMS, Providence, 1990.

[2] N. Anderson, E. B. Saff, and R. S. Varga. On the Eneström–Kakeya theorem and its sharpness. *Linear Alg. Appl.*, 28:5–16, 1979.

[3] G. Battle. Phase space localization theorem for ondelettes. *J. Math. Phys.*, 30:2195–2196, 1989.

[4] C. Bender and S. Orszag. *Advanced Mathematical Methods for Scientists and Engineers.* McGraw-Hill, New York, 1978.

[5] W. L. Briggs. *A multigrid tutorial.* SIAM, Philadelphia, 1987.

[6] G. Brown and W. Moran. A dichotomy for infinite products of discrete measures. *Proc. Cambridge Philos. Soc.*, 73:307–316, 1973.

[7] T. Cai and J. Shen. Boundedness is redundant in a theorem of Daubechies. To appear in *Appl. Comp. Harm. Anal.*, 1998.

[8] E.W. Cheney. *Introduction to approximation theory.* McGraw-Hill, New York, 1966.

[9] I. Daubechies. Orthogonal bases of compactly supported wavelets. *Comm. Pure. Appl. Math.*, 41:909–996, 1988.

[10] I. Daubechies. *Ten lectures on wavelets.* SIAM, Philadelphia, 1992.

[11] I. Daubechies and J. Lagarias. Two scale difference equations I. *SIAM J. Math. Anal.*, 22:1388–1410, 1991.

[12] I. Daubechies and J. Lagarias. Two scale difference equations II. *SIAM J. Math. Anal.*, 23:1031–1079, 1992.

[13] J. A. Davis and L. A. Hageman. An iterative method for solving the neutron transport equation in x-y geometry. *SIAM J. Appl. Math.*, 17:149–161, 1969.

[14] N. G. de Bruijn. The difference-differential equation $F'(x) = e^{\alpha x+\beta}F(x-1)$. *I. II. Nederl. Akad. Wetensch. Proc. Ser. A* **56**=*Indag. Math.* **15**, 449-464, 1953.

[15] J. A. de Reina Martinez. Definicio y estudio de una funcion indefinidamente diferenciable de soporto compacto. *Rev. Real Cienc. Acad. Exact. Fis. Natur. Madrid*, 76:21–38, 1982.

[16] G. Derfel. Probabilistic method for a class of functional-differential equations. *Ukrain. Math. J.*, 41:1117–1234, 1989.

[17] G. Derfel, N. Dyn, and D. Levin. Generalized refinement equations and subdivision processes. *J. Approx. Theory*, 80:272–297, 1995.

[18] G. Deslauriers and S. Dubuc. Interpolation dyadique. In: *Fractals, dimensions non entiès et applications,* Masson, Paris, 44-55, 1987.

[19] G. Deslauriers and S. Dubuc. Symmetric iterative interpolation processes. *Constr. Approx.*, 5:49–68, 1989.

[20] T. A. Driscoll, K. C. Toh, and L. N. Trefethen. Matrix iterations: the six gaps beteen potential theory and convergence. Preprint, 1996.

[21] N. Dyn and D. Levin. Interpolating subdivision schemes for the generation of curves and surfaces. *Internat. Ser. Numer. Math.*, 94:91–106, 1990.

[22] M. Eiermann, X. Li, and R. S. Varga. On hybrid semi-iterative methods. *SIAM J. Numer. Anal.*, 26:152–168, 1989.

[23] M. Eiermann, W. Niethammer, and R. S. Varga. A study of semiiterative methods for nonsymmetric systems of linear equations. *Numer. Math.*, 47:505–533, 1985.

[24] P. Erdös. On a family of symmetric Bernoulli convolutions. *Amer. J. Math.*, 61:974–976, 1939.

[25] P. Erdös. On the smoothness properties of a family of symmetric Bernoulli convolutions. *Amer. J. Math.*, 62:180–186, 1940.

[26] H. E. Fettis, J. C. Caslin, and K. R. Craer. Complex zeros of the error function and of the complementary error function. *Math. Comp.*, 27:401–407, 1973.

[27] B. Fischer. *Polynomial based iteration methods for symmetric linear systems.* Wiley & Sons and Teubner, 1996.

[28] L. Fox and D. F. Mayers. On a functional differential equation. *J. Inst. Math. Appl.*, 8:271–307, 1971.

[29] P. O. Frederickson. Dirichlet series solutions for certain functional differential equations. in *Lecture Notes in Mathematics*, 243:249–254, 1971.

[30] R. W. Freund. On polynomial preconditioning and asymptotic convergence factors for indefinite hermitian matrices. *Linear Alg. Appl.*, 154-156:259–288, 1991.

[31] W. H. J. Fuchs. On the degree of chebyshev approximation on sets with several components. *Izv. Akad. Nauk Armyan. SSR*, 13:396–404, 1978.

[32] W. H. J. Fuchs. On Chebyshev approximation on sets with several components. In D. A. Brannan and J. G. Clunie, editors, *Aspects of Contemporary Complex Analysis*, pages 399–408. Academic Press, 1980.

[33] W. H. J. Fuchs, J. F. Kaiser, and H. J. Landau. Asymptotic behavior of a family of window functions used in non–recursive digital filter design. Technical report, Bell Laboratories, 1980.

[34] A. Garsia. Arithmetic properties of Bernoulli convolutions. *Trans. Amer. Math. Soc.*, 102:409–432, 1962.

[35] J. S. Geronimo and W. Van Assche. Orthogonal polynomials on several intervals via a polynomial mapping. *Trans. Amer. Math. Soc.*, 308(2):559–581, 1988.

[36] J. Goldman and G.-C. Rota. On the foundations of combinatorial theory iv: finite vector spaces and Eulerian generating functions. *Stud. Appl. Math.*, 49:239–258.

[37] J. Goldman and G.-C. Rota. The number of subspaces of a vector space. in: W. Tutte Ed., *Recent progress in Combinatorics,* Academic Press, New York, 1969.

[38] P. Henrici. *Applied and computational complex analysis.* Wiley, New York, 1986.

[39] B. Jessen and A. Wintner. Distribution functions and the Riemann zeta function. *Trans. Amer. Math. Soc.*, 38:48–88, 1935.

[40] J. F. Kaiser. Nonrecursive digital filter design using the i_0–sinh window function. *Proc. 1974 IEEE Symp. Circuits and Syst.*, pages 20–23, 1974.

[41] D. Kateb and P. G. Lemarié. Asymptotic behavior of the Daubechies filters. *Appl. Comp. Harm. Anal.*, 2:398–399, 1995.

[42] T. Kato and J. B. McLeod. The functional-differential equation: $y'(x) = ay(\lambda x) + by(x)$. *Bull. Amer. Math. Soc.*, 77:891–937, 1971.

[43] G. Kh. Kirov and G. A. Totkov. Distributions of zeros of derivatives of the function $\lambda(x)$. *Differential Equations and Applications,* I, II, Tech. Univ., Ruse 1982, 341-344, 1982.

[44] H Kober. *Dictionary of conformal representations.* Dover, New York, 1957.

[45] M. Lang and B. C. Frenzel. Polynomial root finding. Preprint, Rice University, 1994.

[46] A. Aldroubi M. Unser and M. Eden. On the asymptotic convergence of b-spline wavelets to Gabor functions. *IEEE Trans. Inform. Theory*, 38:864–872, 1992.

[47] K. Mahler. On a special functional equation. *J. London Math. Soc.*, 15:115–123, 1940.

[48] S. Mallat. Multiresolution approximation and wavelets. *Trans. Amer. Math. Soc.*, 315:69–88, 1989.

[49] M. Marden. *Geometry of Polynomials.* AMS, Providence, 1966.

[50] Y. Meyer. *Wavelets and Operators.* Cambridge University Press, 1992.

[51] J. Murray. *Asymptotic Analysis.* Clarendon Press, Oxford, 1974.

[52] A. D. Myshkis. On certain problems in the theory of differential equations with deviating argument. *Russian Math. Surveys*, 32:181–213, 1977.

[53] Z. Nehari. *Conformal mapping.* McGraw–Hill, New York, 1952.

[54] O. Nevanlinna. Power bounded prolongations and Picard-Lindelöf iteration. *Numer. Math.*, 58:479–501, 1990.

[55] R. Nevanlinna. *Analytic functions.* Springer-Verlag, New York, 1970.

[56] F.W.J. Olver. *Asymptotics and Special Functions.* Academic Press, 1974.

[57] A. V. Oppenheim and R. W. Schafer. *Discrete-time signal processing.* Prentice Hall, New Jersey, 1989.

[58] T. W. Parks and J. H. McClellan. Chebyshev approximation for nonrecursive digital filters with linear phase. *IEEE Trans. on Circuit Theory*, CT-19, 1972.

[59] F. Peherstorfer. Gauss-Tchebycheff quadrature formulas. *Numer. Math.*, 58:273–286, 1990.

[60] L. R. Rabiner and B. Gold. *Theory and application of digital signal processing.* Prentice-Hall, 1975.

[61] B. L. S. Prakasa Rao. *Asymptotic theory of statistical inference.* John Wiley & Sons, 1986.

[62] V. A. Rvachev. Compactly supported solutions of functional-differential equations and their applications. *Russian Math. Surveys*, 45:87–120, 1990.

[63] V. L. Rvachev and V. A. Rvachev. On a function with compact support. *Dopov. Akad. Nauk. URSR*, 8:705–707, 1971.

[64] J. Shen. Refinement differential equations and wavelets. Submitted to *Methods Appl. Anal.*, 1997.

[65] J. Shen and G. Strang. Asymptotic analysis of Daubechies polynomials. *Proc. Amer. Math. Soc.*, 124:3819–3833, 1996.

[66] J. Shen and G. Strang. Asymptotics of Daubechies filters, scaling functions and wavelets. To appear in *Appl. Comp. Harm. Anal.*, 1998.

[67] J. Shen and G. Strang. The asymptotics of optimal (equiripple) filters. To appear in *IEEE Trans. Sig. Proc.*, 1998.

[68] J. Shen, G. Strang, L. N. Trefethen, and A. Wathen. The Green's function of several intervals and its applications. Preprint, 1998.

[69] G. Strang and T. Nguyen. *Wavelets and filter banks.* Wellesley-Cambridge Press, Wellesley, MA, 1996.

[70] R. Strichartz. *A guide to distribution theory and Fourier transform.* CRC Press, Florida, 1993.

[71] G. Szegö. über eine eigenschaft der exponentialreihe. *Sitzungsber. Berlin Math. Ges.*, 23:50–64, 1924.

[72] N. M. Temme. Asymptotics and numerics of zeros of polynomials that are related to Daubechies wavelets. Technical report, Report AM–R9613, CWI, Amsterdam, 1996.

[73] L. N. Trefethen, M. Embree, and S. Mitchell. Weighted Green's functions and polynomial approximation on multiple intervals via conformal mapping. Preprint, Oxford Computing Lab., 1998.

[74] P. P. Vaidyanathan. *Multirate systems and filter banks.* Prentice-Hall, 1992.

[75] R. S. Varga. *Scientific Computation on Mathematical Problems and Conjectures.* SIAM, Philadelphia, 1992.

[76] J. L. Walsh. *Interpolation and approximation by rational functions in the complex domain.* AMS, Providence, 1965.

[77] A. Wathen, B. Fischer, and D. J. Silvester. The convergence rate of minimal residual method for the Stokes problem. *Numer. Math.*, 17:121–134, 1995.

[78] A. Wathen, B. Fischer, and D. J. Silvester. The convergence of iterative solution methods for symmetric and indefinite linear system. Technical Report 16, Oxford University Computing Laboratory, 1997.

[79] A. Wathen and D. J. Silvester. Fast iterative solution of stabilised Stokes systems Part i: using simple diagonal preconditioners. *SIAM J. Numer. Ana.*, 30:630–649, 1993.

[80] H. Widom. Extremal polynomials associated with a system of curves in the complex plane. *Advances in Mathematics*, 3:127–232, 1969.

[81] R. Wong. *Asymptotic Approximations of Integrals.* Academic Press, 1989.