Estimating the Knot Locations of Noisy Splines

by

Ilya Polyak

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degrees of

Bachelor of Science in Electrical Science and Engineering

and Master of Engineering in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

January 1995

Author . . . . . .
Department of Electrical Engineering and Computer Science
January 27, 1995

Certified by . . . . . .
Alan S. Willsky
Professor
Thesis Supervisor

Certified by . . . . . .
William C. Karl
Research Scientist
Thesis Supervisor

Accepted by . . . . . .
Frederic R. Morgenthaler
Chairman, Department Committee on Graduate Theses

Estimating the Knot Locations of Noisy Splines

by

Ilya Polyak

Submitted to the Department of Electrical Engineering and Computer Science
on January 27, 1995, in Partial Fulfillment of the Requirements for the Degrees of
Bachelor of Science in Electrical Science and Engineering
and Master of Engineering in Electrical Engineering and Computer Science

Abstract

In [10], [11], Mallat, Hwang and Zhong describe a novel wavelet-based noise removal algorithm. The algorithm uses extrema of the continuous wavelet transform (CWT) to estimate the Lipschitz exponent at each singular point of a noisy signal. By looking at the Lipschitz exponents, it is determined which CWT extrema correspond to the useful signal and which ones correspond to noise. The latter extrema are discarded, and an approximation of the signal is constructed from the remaining extrema. The nature of the algorithm leads to the conjecture that it is fundamentally robust to the detailed statistical structure of the noise and that it is well-suited to the task of spline approximation, but a number of interesting and important questions remain. In order to compute the Lipschitz exponent at a singular point of a noisy signal, it is necessary to determine which extremum is due to that point at every scale of the CWT. Since calculating the CWT for a dense set of scales is computationally inefficient, the authors of [10], [11] use the CWT only at the scales which are integer powers of 2 (so-called "dyadic scales"). In this case, the extrema matching problem becomes highly nontrivial. The authors of [10], [11] use a simple algorithm based on the comparison of the values and positions of extrema at a pair of consecutive scales. This algorithm cannot work equally well for all input signals, because the trajectories of CWT extrema depend on the type of the corresponding singularity as well as on the wavelet. In this thesis, we restrict the approximation class of signals and consider only those input signals which are noisy splines. In this case, a more sophisticated and reliable extrema matching algorithm is possible. We arrive at such an algorithm by treating the extrema matching problem as a multi-target tracking problem. The final result is a robust spline approximation algorithm which we test using Cauchy-contaminated Gaussian noise.

Thesis Supervisor: Alan S. Willsky

Title: Professor

Thesis Supervisor: William C. Karl

Title: Research Scientist

Acknowledgments

I thank my parents Tamara Skomorovskaya and Mikhail Polyak whose strong will, determination, and many sacrifices made it possible for me to live in the freest country in the world and attend its best engineering college.

I am greatly indebted to my thesis advisor, Professor Alan Willsky, for giving me the opportunity to work under his guidance. Most of what I learned during my years at MIT came either directly from him or as a result of working in his group. I would also like to thank Clem Karl for helpful discussions and for his suggestions related to this thesis.

I also gratefully acknowledge the generous support of the National Science Foundation Fellowship which I have held for the past semester.

In addition, I thank my officemates Rachel, Seema (or is it Seeema?), Paul, Lori, François, Eric and Mickey for providing a friendly working atmosphere.

Contents

List of Figures

Chapter 1

STATEMENT OF THE PROBLEM, BACKGROUND AND MOTIVATION

1.1 Introduction

This thesis is devoted to the description of a new algorithm for robust spline approximation. This algorithm emerged on the basis of the results presented in [10], [11] by Mallat, Hwang, and Zhong who used the extrema of the continuous wavelet transform (CWT) of a noisy signal to reconstruct the signal. Their method does not constrain the probabilistic structure of the noise; the sole condition which must hold in order for it to work successfully is distinguishability of the Lipschitz exponents of the useful signal from those of the noise realizations. This circumstance allows one to hope that the method is robust to the detailed statistical nature of the noise, although no statistical analysis of the method has been performed. We shall see later that several other features of this method also contribute to its potential effectiveness in spline approximation. We shall also see that major changes must be made in order to turn this method into a good spline estimation algorithm. Particularly,

the method of [10], [11] involves estimation of the extrema trajectories of CWT across scale, which is done in [10] by comparing the locations and magnitudes of the extrema at consecutive scales. Since we are only interested in reconstructing splines, we are able to develop a better extrema-matching algorithm by casting this matching as a multi-target tracking problem. We then demonstrate the robustness of the resulting algorithm using several noise processes with "heavy tails".

This introductory chapter contains a brief discussion of spline approximation in general, as well as of those applications that motivated this thesis. At the end of this chapter, we formulate the particular problem addressed in this thesis.

1.2 Background on Spline Functions and Applications of Splines

We start this section with the formal definition of spline functions and then discuss the importance of spline approximation.

Definition 1.2.1 *Given a strictly increasing sequence of real numbers* $t_1, t_2, ..., t_n$, *a* **spline function** $f(t)$ *of* **order** m *with* **the knots** $t_1, t_2, ..., t_n$ *is a function defined on the entire real line having the following two properties:*
(1) For every interval (t_i, t_{i+1}), $i = 0, 1, ..., n$ *(where* $t_0 = -\infty$ *and* $t_{n+1} = \infty$*),* $f(t) = P_{m,i}(t)$, $\forall t \in (t_i, t_{i+1})$, *where* $P_{m,i}(t)$ *is a polynomial of degree* m *or less.*
(2) If $m \geq 1$, $f(t)$ *is continuous everywhere; if* $m \geq 2$, *the derivatives of* $f(t)$ *of orders* $1, ..., m-1$ *are everywhere continuous as well.*

Thus, a spline function is a piecewise polynomial function satisfying certain conditions regarding continuity of the function and its derivatives. The knots are defined as the places where the smooth polynomial pieces are joined together. When $m = 0$, condition (2) is not operative, and a spline function of order zero with n knots is a staircase function consisting of n steps. For $m > 0$, a spline function of order m could equally well be defined as an $(m-1)$-times continuously differentiable function whose m-th derivative is a staircase function. Even more concisely, a spline

function of order m is any m-th order indefinite integral of a staircase function.

We shall neither discuss nor need any spline theory beyond this. The books by de Boor [4], Greville [5], and Schumaker [18] are all excellent references.

Splines are very useful for solving approximation problems, because they possess a number of attractive properties. It is easily seen from their definition that splines are easy to store and manipulate on a digital computer; they are relatively smooth; their derivatives and anti-derivatives are splines; every continuous function on the interval $[a, b]$ can be approximated arbitrarily well by polynomial splines with the order m fixed.

Spline approximation is both a nice theoretical problem and has many applications. Some examples are initial value problems [5], nonlinear boundary value problems [5], eigenvalue problems [5], and detection and estimation of abrupt changes in signals [14], [1]. The latter problem is the one in which we are interested. Namely, we would like to approximate a given function by a spline, placing the knots at the points of abrupt changes of the function and approximating the pieces between these points by polynomials.

As the name "detection of abrupt changes" suggests, the estimation of the knot locations is a critical part of the problem. Many practical applications can be used to illustrate this. For example, in analyzing electrocardiograms [14], it is very important to know the times when a heartbeat starts and ends. In tomographic reconstruction of a polygonal object from its projections [2], it is crucial to know the locations of the knots in the projections, because the knots correspond to the vertices of the polygon. In seismic signal processing [15], an important problem is to estimate the time of arrival of a seismic wave or the change of the type of a seismic wave.

1.3 Background of the Problem

There are various least-squares-based methods for spline approximation. When the knot locations are known, the problem becomes that of solving a linear

system of equations. In the case of variable knots, we are led to nonlinear problems (such as nonlinear least squares) which arise in estimating the knot locations. For example, Mier Muth and Willsky [14] proposed an algorithm which is based on the generalized likelihood ratio (GLR) method for detection and estimation of abrupt changes in dynamic systems described by Willsky and Jones in [21].

It is easy to prove that least-squares methods are optimal for estimation in noise if the noise is Gaussian [20]. However, it is also easy to show that they are not robust to heavy-tailed noise [7]. From our point of view, it is unrealistic to always assume Gaussianity of noise or even count on the absence of outliers in observed signals. Therefore, a robust estimation method is needed.

As a starting point for the development of such a method, we choose the wavelet-based noise removal algorithm developed by Mallat, Hwang, and Zhong [10], [11]. The intuition behind this choice is based on two observations. First, Mallat's algorithm does not require the noise to be Gaussian. In fact, incredibly few assumptions about the noise are made. This circumstance allows one to anticipate robustness of the algorithm. Its second attractive feature is that it is based on finding singularities of functions. A particular case of this is precisely what we would like to solve, because the locations of the singularities of a spline are its knots.

We are now ready to summarize everything said above and formulate the objectives of this thesis.

1.4 Problem Statement and Organization.

In this thesis, we would like to adapt Mallat's noise removal algorithm to robust spline approximation. We would like to achieve an algorithm which is robust to outliers and test it on various kinds of heavy-tailed noise. For simplicity, we limit ourselves to linear splines.

We proceed as follows. In Chapter 2, Mallat's algorithm and the theory behind it are described, and the explanation as to why the algorithm in its original form is poorly suited to our problem is given. The end of Chapter 2 and Chapter 3

are devoted to the discussion of the changes that one needs to make in the algorithm in order to adapt it to robust spline approximation. The performance of the modified algorithm for the case of a linear spline with one knot is described in Chapter 4, in comparison with the performance of a least-squares estimator. The last section of that chapter is devoted to estimating the order of a spline. Chapter 5 contains concluding remarks and the discussion of the possibilities for future work.

Chapter 2

A WAVELET-BASED NOISE REMOVAL ALGORITHM

Most of this chapter is a review of the connections between the local regularity of a function and its wavelet transform. These connections lead to a remarkable noise removal algorithm discovered by Mallat and his students [10], [11]. The algorithm is described in the last section of the chapter. Most of the theoretical results described here are due to Holschneider and Tchamitchian [6], and Mallat [10], [11]. The core of what is presented in this chapter can also be found in the excellent book on wavelets by Daubechies [3]. However, in order to find everything presented here, one has to go through a considerable amount of literature. This circumstance motivated the decision to write this tutorial chapter. Therefore, in spite of the abundance of sources, this chapter is self-contained; it should not be necessary to consult any of the references in order to understand it.

The structure of this chapter is as follows. In Section 2.1, which is the basis of all the following sections, we define the wavelet transform and prove its invertibility. In Section 2.2, we define a certain quantity called "the Lipschitz exponent of a function $f(t)$ at a point t_0" which characterizes local smoothness of $f(t)$. We need to do that, because the computation of Lipschitz exponents is the critical part of Mallat's algorithm. We then describe how to compute the Lipschitz exponents of a function from the modulus maxima of its continuous wavelet transform. Unfortu-

nately, the continuous wavelet transform is computationally expensive. Therefore, in applications we are forced to use the dyadic wavelet transform, which can be computed much faster. Section 2.3 defines the dyadic wavelet transform, gives a sufficient condition for its invertibility, and explains how to implement it efficiently. The following section then describes an algorithm for reconstructing an approximation of a signal from its dyadic wavelet transform modulus maxima. Finally, Section 2.5 ties Sections 2.2 and 2.4 together, describing a noise removal algorithm which utilizes our ability to estimate the Lipschitz exponents of a signal from its dyadic wavelet transform.

2.1 Continuous Wavelet Transform (CWT)

Definition 2.1.1 *The* **Fourier transform** *of a function $f(t) \in L^2(R)$ is given by*

$$\mathcal{F}[f(t)] = \hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt$$

Definition 2.1.2 *A function $\psi(t) \in L^2(R)$ is called a* **wavelet** *if its Fourier transform $\hat{\psi}(\omega)$ satisfies the following condition:*

$$C_\psi \stackrel{\text{def}}{=} \int_0^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty \tag{2.1}$$

If $\hat{\psi}(\omega)$ is continuous, (2.1) can only be satisfied if $\hat{\psi}(0) = 0$, which means that $\int_{-\infty}^{\infty} \psi(t)dt = 0$.

Definition 2.1.3 *The* **continuous wavelet transform (CWT)** *of a function $f(t)$ with respect to a wavelet $\psi(t)$ is given by*

$$W_\psi f(t,s) = \int_{-\infty}^{\infty} f(\tau)\frac{1}{s}\psi(\frac{t-\tau}{s})d\tau, s > 0 \tag{2.2}$$

The variable s is called **the scale**. *Sometimes we shall use the notation $\psi_s(t) = \frac{1}{s}\psi(\frac{t}{s})$. Using this notation, $W_\psi f(t,s) = f * \psi_s(t)$, for $s > 0$.*

The following theorem shows that the CWT is invertible, i.e., any function $f(t)$ can be recovered from its wavelet transform.

Theorem 2.1.1 *For any two functions $f, g \in L^2(R)$,*

$$\frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^\infty W_\psi f(t,s) W_\psi g(t,s)\frac{1}{s}\,dt\,ds = <f,g>, \tag{2.3}$$

where C_ψ is as defined by (2.1).

Proof. Using Parseval's relation, we get:

$$\begin{aligned}
W_\psi f(t,s) &= \int_{-\infty}^\infty f(\tau)\frac{1}{s}\psi(\frac{t-\tau}{s})d\tau \\
&= \int_{-\infty}^\infty f(\tau)[\frac{1}{s}\psi(\frac{t-\tau}{s})]^* d\tau \\
&= \frac{1}{2\pi}\int_{-\infty}^\infty \hat{f}(\omega)[e^{-j\omega t}\hat{\psi}(s\omega)]^* d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^\infty \hat{f}(\omega)\hat{\psi}^*(s\omega)e^{j\omega t} d\omega \\
&= \mathcal{F}[F(\omega)]^*,
\end{aligned}$$

where $F(\omega) = \frac{1}{2\pi}\hat{f}^*(\omega)\hat{\psi}(s\omega)$ and the star denotes complex conjugation. Similarly,

$$\begin{aligned}
W_\psi g(t,s) &= \int_{-\infty}^\infty g(\tau)\frac{1}{s}\psi(\frac{t-\tau}{s})d\tau \\
&= \int_{-\infty}^\infty g^*(\tau)\frac{1}{s}\psi(\frac{t-\tau}{s})d\tau \\
&= \frac{1}{2\pi}\int_{-\infty}^\infty \hat{g}^*(\omega)e^{-j\omega t}\hat{\psi}(s\omega)d\omega \\
&= \mathcal{F}[G(\omega)],
\end{aligned}$$

where $G(\omega) = \frac{1}{2\pi}\hat{g}^*(\omega)\hat{\psi}(s\omega)$. Therefore,

$$\begin{aligned}
\int_0^\infty \int_{-\infty}^\infty W_\psi f(t,s) W_\psi g(t,s)\frac{1}{s}\,dt\,ds &= \int_0^\infty \int_{-\infty}^\infty \mathcal{F}[F(\omega)]^*\mathcal{F}[G(\omega)]\frac{1}{s}\,dt\,ds \\
&= \int_0^\infty 2\pi\left[\int_{-\infty}^\infty F^*(\omega)G(\omega)\,d\omega\right]\frac{ds}{s} \\
&= \int_0^\infty \int_{-\infty}^\infty \frac{1}{2\pi}\hat{f}(\omega)\hat{\psi}^*(s\omega)\hat{g}^*(\omega)\hat{\psi}(s\omega)d\omega\,\frac{ds}{s} \\
&= \int_{-\infty}^\infty \frac{1}{2\pi}\hat{f}(\omega)\hat{g}^*(\omega)\left[\int_0^\infty \frac{|\hat{\psi}(s\omega)|^2}{s}\,ds\right]d\omega \\
&= C_\psi \int_{-\infty}^\infty \frac{1}{2\pi}\hat{f}(\omega)\hat{g}^*(\omega)\,d\omega \\
&= C_\psi <f,g>,
\end{aligned}$$

where we have used Parseval's theorem two more times. Interchanging the order of integration was allowed by Fubini's theorem. □

So, if $\{e_i\}_{i=-\infty}^{\infty}$ is an orthonormal basis of the space $L^2(R)$, (2.3) can be used to compute $\{<f, e_i>\}_{i=-\infty}^{\infty}$, which are the coefficients of f with respect to this basis. This means that the coefficients of f with respect to an orthonormal basis can be recovered from its wavelet transform. Therefore, f itself can be recovered from its wavelet transform.

Since (2.3) is actually

$$\frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty W_\psi f(t,s)\frac{1}{s}\int_{-\infty}^\infty g(\tau)\frac{1}{s}\psi(\frac{t-\tau}{s})\,d\tau\,dt\,ds = \int_{-\infty}^\infty g(\tau)f(\tau)\,d\tau,$$

it is very tempting to identify the parts of the left-hand side and right-hand side that operate on $g(\tau)$ and to interpret this equation as:

$$f(\tau) = \frac{1}{C_\psi}\int_0^\infty\int_{-\infty}^\infty W_\psi f(t,s)\frac{1}{s^2}\psi(\frac{t-\tau}{s})\,dt\,ds \tag{2.4}$$

Unfortunately, the double integral in (2.4) is not guaranteed to converge. It turns out, however, that if $\psi(t)$ is compactly supported and differentiable, (2.4) does hold for bounded continuous $f(\tau)$. (The proof of a generalization of this result can be found in Chapter 2 of [3].)

2.2 Using the CWT of a Signal to Characterize Its Lipschitz Regularity

Holschneider and Tchamitchian showed in [6] that the wavelet transform is a very useful tool for studying regularity properties of functions. In order to discuss these results, we are going to need the definition of the Lipschitz regularity of a function.

Definition 2.2.1 *Let $n \in Z^+ \cup \{0\}$, and $n \leq \alpha \leq n+1$. A function $f(t)$ is said to be* **Lipschitz (or Hölder)** *α at a point t_0 if there exist an n-th degree polynomial $P_{t_0,n}(h)$ and two positive constants C and h_0 such that for any $h \in [-h_0, h_0]$, the following inequality holds:*

$$|f(t_0+h) - P_{t_0,n}(h)| \leq C|h|^\alpha \tag{2.5}$$

The quantity $\sup\{\alpha : f(t)$ *is Lipschitz* α *at* $t_0\}$ *is called the* **Lipschitz regularity** *of* $f(t)$ *at* t_0.

We can also define uniform Lipschitz regularity, i.e., Lipschitz regularity on an interval:

Definition 2.2.2 *A function* $f(t)$ *is* **uniformly Lipschitz** α *on an interval* (a, b) *for* $\alpha \in (0, 1]$ *if*

$$|f(t_1) - f(t_2)| \leq C|t_1 - t_2|^\alpha, \quad \forall t_1, t_2 \in (a, b) \tag{2.6}$$

$f(t)$ *is* **uniformly Lipschitz** α *on* (a, b) *for* $\alpha \in (n, n+1]$ *if* $f(t)$ *is* n *times continuously differentiable on* (a, b), *and* $f^{(n)}(t)$ *is uniformly Lipschitz* $\alpha - n$ *on* (a, b). *The* **uniform Lipschitz regularity** *of* $f(t)$ *on* (a, b) *is* $\sup\{\alpha : f(t)$ *is uniformly Lipschitz* α *on* $(a, b)\}$.

The Lipschitz regularity of a function $f(t)$ at a point is a measure of the smoothness of $f(t)$ at that point. For example, it follows immediately from the definition that the unit step function, the unit ramp function, and $f_1(t) = -|t|^{\frac{1}{3}}$ are Lipschitz 0, 1, and $\frac{1}{3}$ at $t = 0$, respectively. The unit step is discontinuous; the derivative of the unit ramp is the unit step; and $f_1(t)$ is in between: even though it is continuous, it has a peak at the origin which is sharper than any peak of any linear spline. It also follows that $f_2(t) = t^2$, which is smooth everywhere, is Lipschitz α at $t = 0$ for any real α. In this thesis, we are interested mostly in linear splines. The Lipschitz regularity of a linear spline at a knot is 1. (It follows then that, for an m-th order spline, the Lipschitz regularity at a knot is m.)

The following several theorems link the Lipschitz exponents of a function with its wavelet transform.

Theorem 2.2.1 *Suppose that* $\psi(t)$ *is a wavelet and* $\int_{-\infty}^{\infty} |\psi(t)|(1 + |t|)\, dt < \infty$. *If a bounded function* $f(t)$ *is uniformly Lipschitz* α *on* R *with* $0 < \alpha \leq 1$, *its wavelet transform satisfies*

$$|W_\psi f(t, s)| \leq A s^\alpha \tag{2.7}$$

for some constant A.

Proof. Let us define μ and I by: $\mu = \frac{t-\tau}{s}$, and $I = \int_{-\infty}^{\infty} f(t)\frac{1}{s}\psi(\frac{t-\tau}{s})\,d\tau$. Observe that $I = f(t)\int_{-\infty}^{\infty}\psi(\mu)\,d\mu \equiv 0$, and therefore

$$|W_\psi f(t,s)| = |W_\psi f(t,s) - I| = \left|\int_{-\infty}^{\infty}(f(\tau)-f(t))\frac{1}{s}\psi(\frac{t-\tau}{s})\,d\tau\right| \leq$$

$$\leq \int_{-\infty}^{\infty}|f(\tau)-f(t)|\frac{1}{s}\left|\psi(\frac{t-\tau}{s})\right|\,d\tau \leq \int_{-\infty}^{\infty}C|\tau-t|^\alpha\frac{1}{s}\left|\psi(\frac{t-\tau}{s})\right|\,d\tau =$$

$$Cs^\alpha\int_{-\infty}^{\infty}|\psi(\mu)|\cdot|\mu|^\alpha\,d\mu = As^\alpha \quad \square$$

Theorem 2.2.2 *Suppose that $\psi(t)$ is a compactly supported, continuously differentiable wavelet, and $f \in L^2(R)$ is bounded and continuous. If, for some $\alpha \in (0,1)$,*

$$|W_\psi f(t,s)| \leq As^\alpha, \tag{2.8}$$

then f is uniformly Lipschitz α on R.

Proof: see Theorem 2.2.2a below. $\square$

The two preceding theorems show that the Lipschitz exponents of a function that are between zero and one can be characterized by the decay in s of the absolute value of the CWT of the function. The generalizations of these theorems to the case of $\alpha \in (n, n+1)$ are easily obtained by induction. For that case, however, we need an n times continuously differentiable $f(t)$ and a wavelet $\psi(t)$ with n vanishing moments.

The integral $\int_{-\infty}^{\infty}t^p g(t)\,dt$ is called the **p-th moment** of the function $g(t)$. We say that a wavelet $\psi(t)$ has n vanishing moments if $\int_{-\infty}^{\infty}t^p\psi(t)\,dt = 0$ for $p = 0, 1, ..., n-1$.

Inductive Lemma. *Suppose that $\psi(t)$ is a compactly supported wavelet with n vanishing moments. Also suppose that $f(t)$ is n times continuously differentiable. Then $\psi^{(-n)}(t)$ is a compactly supported wavelet and*

$$W_\psi f(t,s) = s^n W_{\psi^{(-n)}} f^{(n)}(t,s), \tag{2.9}$$

where superscript (n) means "the n-th derivative", and superscript $(-n)$ means "the n-th anti-derivative defined by $\psi^{(-n)}(t) = \int_{-\infty}^{t}\psi^{(-(n-1))}(\rho)\,d\rho$".

Proof. Let us prove by induction that $\psi^{(-m)}(t)$ is a compactly supported wavelet with $n-m$ vanishing moments for $0 \leq m \leq n$, i.e., that

$$\int_{-\infty}^{\infty} t^p \psi^{(-m)}(t)\, dt = 0,$$

for $0 \leq m \leq n$ and $0 \leq p < n-m$. When $m = 0$, the statement is true because $\psi(t)$ is a wavelet. Suppose that it is true for $m = k$. Then, since $\psi^{(-k)}(t)$ is compactly supported and its integral from $-\infty$ to ∞ is zero, it follows from the definition of $\psi^{(-(k+1))}(t)$ that it, too, is compactly supported. We have:

$$\int_{-\infty}^{\infty} t^p \psi^{(-(k+1))}(t)\, dt = \frac{t^{p+1}}{p+1}\psi^{(-(k+1))}(t)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{t^{p+1}}{p+1}\psi^{(-k)}(t)\, dt$$

The first term is equal to zero because $\psi^{(-(k+1))}$ is compactly supported. The second term is a constant times the $(p+1)$-st moment of $\psi^{(-k)}$, and therefore, by the induction assumption, that term is also zero as long as $0 \leq p+1 < n-k$. So, we proved that

$$\int_{-\infty}^{\infty} t^p \psi^{(-(k+1))}(t)\, dt = 0,$$

for $p = 0, 1, ..., n-(k+2)$. The induction is done.

Now observe that $s\frac{d}{dt}[\psi_s^{(-1)}(t)] = s\frac{d}{dt}[\frac{1}{s}\psi^{(-1)}(\frac{t}{s})] = \frac{d}{dt}[\psi^{(-1)}(\frac{t}{s})] = \frac{1}{s}\psi(\frac{t}{s}) = \psi_s(t)$, and similarly, $s^n \frac{d^n}{dt^n}[\psi_s^{(-n)}(t)] = \psi_s(t)$. Therefore, $W_\psi f(t,s) = f(t) * \psi_s(t) = f^{(n)}(t) * s^n \psi_s^{(-n)}(t) = s^n W_{\psi^{(-n)}} f^{(n)}(t,s)$. The lemma is proved. □

An immediate consequence of the Lemma is that

$$|W_\psi f(t,s)| \leq As^\alpha \quad \Leftrightarrow \quad |W_{\psi^{(-n)}} f^{(n)}(t,s)| \leq As^{\alpha-n} \tag{2.10}$$

This allows us to generalize Theorems 2.2.1 and 2.2.2. Combining their general cases, we obtain the following:

Theorem 2.2.3 *Given a number $\alpha \in (n, n+1)$ and a compactly supported wavelet $\psi(t)$ with at least n vanishing moments, the following two statements are equivalent:*
1) $f(t)$ is n times continuously differentiable, with all the $f^{(m)}$, $m = 0, 1, ..., n$ bounded and square-integrable, and $f(t)$ is uniformly Lipschitz α on R.
2) $|W_\psi f(t,s)| \leq As^\alpha$. □

If we use a wavelet with compact support, there is going to be a region in (t, s)-plane where the CWT of $f(\tau)$ is going to be influenced only by the values of $f(\tau)$ in the interval $[a, b]$. Specifically, suppose that the support of $\psi(t)$ is $[-r, r]$, and we want $W_\psi f(t, s_0)$ to be equal to $\int_a^b f(\tau)\psi_{s_0}(t-\tau)\, d\tau$ for $c \leq t \leq d$. In order for that to hold, the support of $\psi\left(\frac{t-\tau}{s_0}\right)$ must be inside of $[a, b]$, for any t between c and d. In other words, $[t - rs_0, t + rs_0] \subset [a, b]$, $\forall t \in [c, d] \Rightarrow c = a + rs_0$, $d = b - rs_0$. Since $c \leq d$, we also have the condition $s_0 \leq \frac{b-a}{2r}$. So, the piece of $f(t)$ from $t = a$ to $t = b$ determines $W_\psi f(t, s)$ in the triangle with vertices $(t = a,\ s = 0)$, $(t = b,\ s = 0)$, and $(t = \frac{b+a}{2},\ \frac{b-a}{2r})$ (see Fig. 2.1). We shall call this triangle **the triangle of sovereignty** of the interval $[a, b]$. Theorem 2.2.1 suggests that perhaps if ψ were compactly supported, we could use Lipschitz exponents of $f(t)$ for $a < t < b$ to characterize the decay in s of $|W_\psi f(t, s)|$ in this triangle.

Theorem 2.2.4 (A more local version of Theorem 2.2.1.) *Let $\psi(t)$ be a compactly supported wavelet, with $supp[\psi(t)] = [-r, r]$. If a bounded function $f(t)$ is uniformly Lipschitz α on (a, b) with $0 < \alpha \leq 1$, its wavelet transform satisfies*

$$|W_\psi f(t, s)| \leq As^\alpha,$$

for all interior points of the triangle of sovereignty of $[a, b]$. □

The proof is the same as the proof of Theorem 2.2.1, with integration limits changed from $(-\infty, \infty)$ to $[a, b]$ when appropriate.

Similarly, if we use a wavelet with compact support, there will be a region in the (t, s)-plane where the CWT of $f(\tau)$ will not be influenced by the values of $f(\tau)$ on the interval $[a, b]$. Let us suppose that the support of $\psi(t)$ is $[-r, r]$. Then, for any s_0, we can determine the interval $c \leq t \leq d$ on which $f(\tau)$, $a \leq \tau \leq b$ contributes to $W_\psi f(t, s_0)$. To find d, we consider the extreme case when the leftmost point of $supp\left[\psi\left(\frac{d-\tau}{s_0}\right)\right]$ is b: $d - rs_0 = b \Rightarrow d = b + rs_0$. To find c, we consider the other extreme case when the rightmost point of $supp\left[\psi\left(\frac{c-\tau}{s_0}\right)\right]$ is a: $c + rs_0 = a \Rightarrow c = a - rs_0$. So, the infinite trapezoid in (t, s)-plane defined by the lines $t_1(s) = a - rs$, $t_2(s) = b + rs$ is the region where $f(\tau)$ for $a \leq \tau \leq b$ influences

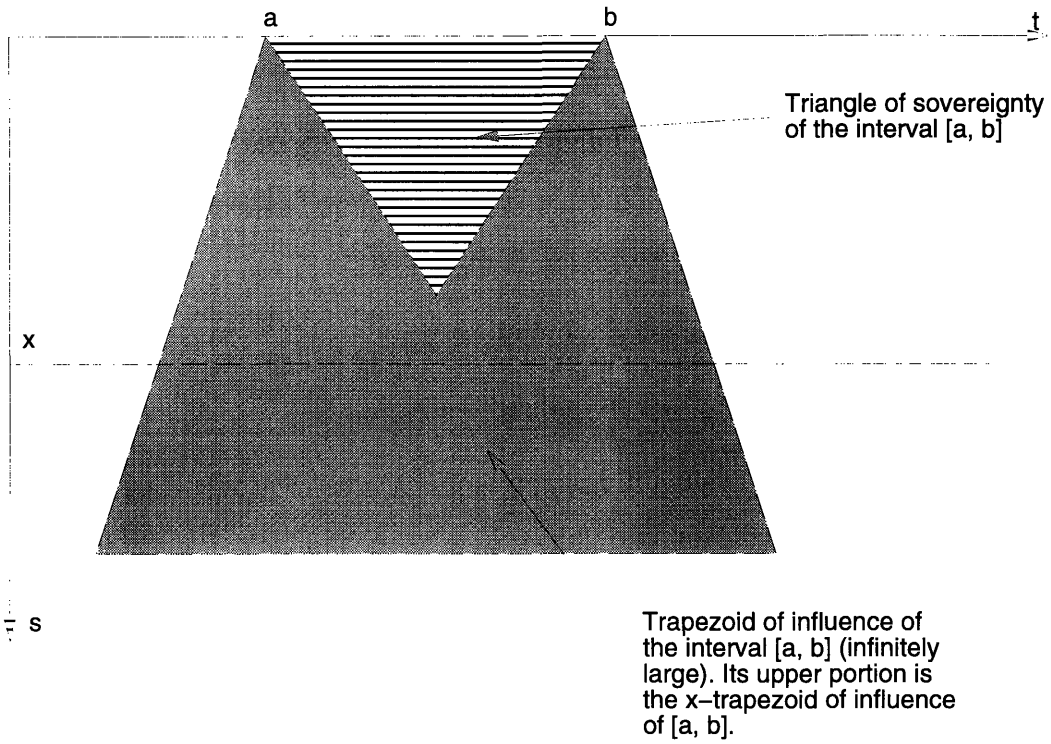

Figure 2.1: The triangle of sovereignty and the trapezoid of influence of $[a, b]$

$W_\psi f(t, s)$. We shall call this region **the trapezoid of influence of** $[a, b]$. We shall call the portion of this trapezoid defined by $s < s_0$ **the s_0-trapezoid of influence of $[a, b]$.** (See Fig. 2.1).

So, we should be able to characterize the uniform Lipschitz regularity on finite intervals using compactly supported wavelets. We have the following, slightly stronger, version of the Theorem 2.2.2:

Theorem 2.2.5 *Suppose that $supp[\psi(t)] = [-r, r]$, and that $\psi(t)$ is a continuously differentiable wavelet. Suppose further that $f \in L^2(R)$ is bounded and continuous. If, for some $\alpha \in (0, 1)$ and $s_0 > 0$, the CWT of $f(t)$ satisfies*

$$|W_\psi f(t, s)| \leq A s^\alpha, \tag{2.11}$$

inside the s_0-trapezoid of influence of $[a, b]$, then f is uniformly Lipschitz α on (a, b).

Proof. Throughout this proof, we are going to be interested in $f(\tau)$ for $a \leq \tau \leq b$. Therefore, whenever we talk about $f(\tau)$, it is assumed that $a \leq \tau \leq b$.

In order to prove that some function $h(\tau)$ is uniformly Lipschitz α on (a,b), we could prove that for all $\varepsilon \in (0,\varepsilon_0]$, $|h(\tau+\varepsilon)-h(\tau)| \leq A\varepsilon^\alpha$. Then , if $b-a$ is finite, the result immediately follows, because we can partition any subinterval of (a,b) into $\lceil \frac{b-a}{\varepsilon_0} \rceil$ intervals of length ε or less. However, if a or b is infinite, we also need $h(\tau)$ to be bounded, i.e., $|h(\tau)| \leq A_1$. Then,

$$\begin{aligned} |h(\tau+\varepsilon)-h(\tau)| &\leq 2A_1 \leq 2A_1\frac{\varepsilon^\alpha}{\varepsilon_0^\alpha} = \frac{2A_1}{\varepsilon_0^\alpha}\varepsilon^\alpha, \ \ \varepsilon \geq \varepsilon_0 \\ |h(\tau+\varepsilon)-h(\tau)| &\leq A\varepsilon^\alpha, \ \ \varepsilon \leq \varepsilon_0, \end{aligned}$$

which means that $|h(\tau+\varepsilon)-h(\tau)| \leq \max\left[A, \frac{2A_1}{\varepsilon_0^\alpha}\right]\varepsilon^\alpha$. Consequently, we need to bound both $|h(\tau+\varepsilon)-h(\tau)|$ (by $A\varepsilon^\alpha$) and $|h(\tau)|$ (by A_1).

Our plan for proving that $f(\tau)$ is uniformly Lipschitz α on (a,b) is as follows. We represent $f(\tau)$ as $f(\tau) = I_1(\tau) + I_2(\tau)$, and prove, according to the procedure outlined in the previous paragraph, that both $I_1(\tau)$ and $I_2(\tau)$ are Lipschitz α.

Without loss of generality, we can assume that $\psi(t)$ is normalized in such a way that $C_\psi = 1$. Since $\psi(t)$ is compactly supported, the inversion formula (2.4) holds:

$$f(\tau) = \int_0^\infty \int_{-\infty}^\infty W_\psi f(t,s)\frac{1}{s^2}\psi(\frac{t-\tau}{s})\,dt\,ds$$

Moreover, since $supp[\psi(t)] = [-r,r]$, we can replace the limits of integration with respect to t by $\tau+rs$ and $\tau-rs$:

$$f(\tau) = \int_0^\infty \int_{\tau-rs}^{\tau+rs} W_\psi f(t,s)\frac{1}{s^2}\psi(\frac{t-\tau}{s})\,dt\,ds$$

Let us break this integral into two: $f(\tau) = I_1(\tau) + I_2(\tau)$, where the first integral is over small scales, and the second one is over large scales:

$$\begin{aligned} I_1(\tau) &= \int_0^{s_0} \int_{\tau-rs}^{\tau+rs} W_\psi f(t,s)\frac{1}{s^2}\psi(\frac{t-\tau}{s})\,dt\,ds \\ I_2(\tau) &= \int_{s_0}^\infty \int_{\tau-rs}^{\tau+rs} W_\psi f(t,s)\frac{1}{s^2}\psi(\frac{t-\tau}{s})\,dt\,ds \end{aligned}$$

Our task now is to come up with some bounds for $|I_1(\tau)|$, $|I_2(\tau)|$, $|I_1(\tau+\varepsilon)-I_1(\tau)|$, and $|I_2(\tau+\varepsilon)-I_2(\tau)|$ given (2.11). Notice that since $I_1(\tau)$ is an integral over some subset of the s_0-trapezoid of influence of $[a,b]$, we can use (2.11) during the process of bounding it.

For $I_2(\tau)$, we have:

$$\begin{aligned} |I_2(\tau)| &\leq \int_{s_0}^{\infty}\int_{\tau-rs}^{\tau+rs} |W_\psi f(t,s)|\frac{1}{s^2}\left|\psi(\frac{t-\tau}{s})\right| dt\,ds \\ &\leq \int_{s_0}^{\infty}\int_{\tau-rs}^{\tau+rs} \|f\|_{L^2}\frac{1}{s}\|\psi\|_{L^2}\left|\psi(\frac{t-\tau}{s})\right|\frac{1}{s^2} dt\,ds \\ &= \|f\|_{L^2}\cdot\|\psi\|_{L^2}\cdot\|\psi\|_{L^1}\int_{s_0}^{\infty}\frac{ds}{s^2} = const < \infty, \end{aligned}$$

where the first transition was made because $|\int f| \leq \int |f|$; the second transition was true because $|W_\psi f(t,s)| = |<f,\psi_s>| \leq \|f\|_{L^2}\cdot\|\psi_s\|_{L^2} = \|f\|_{L^2}\cdot\left(\int_{-\infty}^{\infty}\frac{1}{s^2}\psi^2(\frac{t}{s})\,dt\right)^{\frac{1}{2}} = \frac{1}{s}\|f\|_{L^2}\cdot\|\psi\|_{L^2}$.

A bound for $I_1(\tau)$ can be found as follows:

$$\begin{aligned} |I_1(\tau)| &\leq \int_{0}^{s_0}\int_{\tau-rs}^{\tau+rs} |W_\psi f(t,s)|\frac{1}{s^2}\left|\psi(\frac{t-\tau}{s})\right| dt\,ds \\ &\leq \int_{0}^{s_0}\int_{\tau-rs}^{\tau+rs} As^{\alpha}\frac{1}{s^2}\left|\psi(\frac{t-\tau}{s})\right| dt\,ds = \|\psi\|_{L^1}\int_{0}^{s_0} As^{\alpha-1}\,ds = C_1 < \infty, \end{aligned}$$

where the second transition was obtained by using (2.11). So, both $I_1(\tau)$ and $I_2(\tau)$ are bounded.

Now we must estimate ΔI_2 and ΔI_1.

$$\begin{aligned} &|I_2(\tau+\varepsilon) - I_2(\tau)| \leq \\ &\leq \int_{s_0}^{\infty}\int_{\tau-rs}^{\tau+rs+\varepsilon}\left[\int_{t-rs}^{t+rs}|f(\mu)|\frac{1}{s}\left|\psi(\frac{t-\mu}{s})\right| d\mu\right]\cdot\left|\psi(\frac{t-\tau-\varepsilon}{s}) - \psi(\frac{t-\tau}{s})\right|\frac{1}{s^2} dt\,ds \\ &\leq \int_{s_0}^{\infty}\int_{\tau-rs}^{\tau+rs+\varepsilon}\left[\int_{t-rs}^{t+rs}|f(\mu)|\frac{1}{s}C_2\,d\mu\right] C_3\frac{\varepsilon}{s}\frac{1}{s^2} dt\,ds \\ &\leq C_4\int_{s_0}^{\infty}\int_{\tau-rs}^{\tau+rs+\varepsilon}\left[\int_{t-rs}^{t+rs}|f(\mu)|\,d\mu\right]\frac{\varepsilon}{s^4} dt\,ds \\ &\leq C_4\int_{s_0}^{\infty}\int_{\tau-rs}^{\tau+rs+\varepsilon} C_5\cdot 2rs\,\frac{\varepsilon}{s^4} dt\,ds \\ &= C_6\int_{s_0}^{\infty}\frac{2rs(2rs+\varepsilon)\varepsilon}{s^4} ds \\ &= C_7\varepsilon + C_8\varepsilon^2 \\ &\leq C_9\varepsilon, \end{aligned}$$

for all $\tau+\varepsilon,\ \tau\in(a,b) \Rightarrow I_2(\tau)$ is uniformly Lipschitz 1 $\Rightarrow I_2(\tau)$ is uniformly Lipschitz α, because $\alpha \leq 1$. (The second transition in the last chain of inequalities was obtained using the differentiability of $\psi(\tau)$, i.e, the fact that $\left|\psi(\frac{t-\tau-\varepsilon}{s}) - \psi(\frac{t-\tau}{s})\right| \leq C_3\left|\frac{\varepsilon}{s}\right|$.)

Finally,

$$
\begin{aligned}
&|I_1(\tau+\varepsilon)-I_1(\tau)| \leq \\
&\leq \left[\int_0^{\varepsilon}+\int_{\varepsilon}^{s_0}\right]\int_{\tau-rs}^{\tau+rs+\varepsilon}|W_\psi f(t,s)|\frac{1}{s^2}\left|\psi(\frac{t-\tau-\varepsilon}{s})-\psi(\frac{t-\tau}{s})\right| dt\,ds \\
&\leq \int_0^{\varepsilon}\int_{\tau-rs}^{\tau+rs+\varepsilon} As^{\alpha}\frac{1}{s^2}\left(\left|\psi(\frac{t-\tau-\varepsilon}{s})\right|+\left|\psi(\frac{t-\tau}{s})\right|\right) dt\,ds \\
&+ \int_{\varepsilon}^{s_0}\int_{\tau-rs}^{\tau+rs+\varepsilon} s^{\alpha}\frac{1}{s^2}C_9\frac{\varepsilon}{s}\,dt\,ds \\
&\leq C_{10}\|\psi\|_{L^1}\int_0^{\varepsilon}s^{\alpha-1}\,ds+C_9\varepsilon\int_{\varepsilon}^{s_0}s^{\alpha-3}\cdot(2rs+\varepsilon)\,ds \leq C_{11}\varepsilon^{\alpha}
\end{aligned}
$$

We are finished, because $|f(\tau+\varepsilon)-f(\tau)| = |I_1(\tau+\varepsilon)+I_2(\tau+\varepsilon)-I_1(\tau)-I_2(\tau)| \leq |I_1(\tau+\varepsilon)-I_1(\tau)|+|I_2(\tau+\varepsilon)-I_2(\tau)| \leq A\varepsilon^{\alpha} \Rightarrow f(\tau)$ is uniformly Lipschitz α on (a,b). The proof of the Theorem 2.2.5 is completed. □

At the beginning of this section, we had Theorems 2.2.1 and 2.2.2 linking the global uniform Lipschitz exponents and the decay in s of $|W_\psi f(t,s)|$ everywhere. We combined those theorems together, and obtained an "if and only if" statement in Theorem 2.2.3. Now we would like to combine Theorems 2.2.4 and 2.2.5 and get an "if and only if" theorem for the uniform Lipschitz exponents on intervals. We are faced with a problem here, because Theorems 2.2.4 and 2.2.5 are not converses of each other: Theorem 2.2.5 requires apparently stricter conditions on the decay of CWT than Theorem 2.2.4 is willing to give. In order to repair this inconsistency, we consider an interval $(a+\varepsilon, b-\varepsilon)$, for $0 < \varepsilon < \frac{a+b}{2}$. Since ε is strictly positive, there exists a q-trapezoid of influence of $(a+\varepsilon, b-\varepsilon)$ which lies inside the triangle of sovereignty of (a,b). Moreover, for any γ which is between zero and ε, there exists a p-trapezoid of influence of $(a+\varepsilon, b-\varepsilon)$ which lies inside the triangle of sovereignty of $(a+\gamma, b-\gamma)$ (see Fig. 2.2)

Therefore, we can combine Theorems 2.2.4 and 2.2.5 in the following way:

Theorem 2.2.6 *Given a compactly supported wavelet $\psi(t)$ with n vanishing moments and a number $\alpha \in (n, n+1)$, the following two statements are equivalent:*
1) $f(t)$ is n times continuously differentiable on (a,b), with all the $f^{(m)}$, $m = 0,1,...,n$ bounded and square-integrable on (a,b), and $f(t)$ is uniformly Lipschitz α on $(a+\varepsilon, b-\varepsilon)$, $\forall\varepsilon$ such that $0 < \varepsilon < \frac{a+b}{2}$.

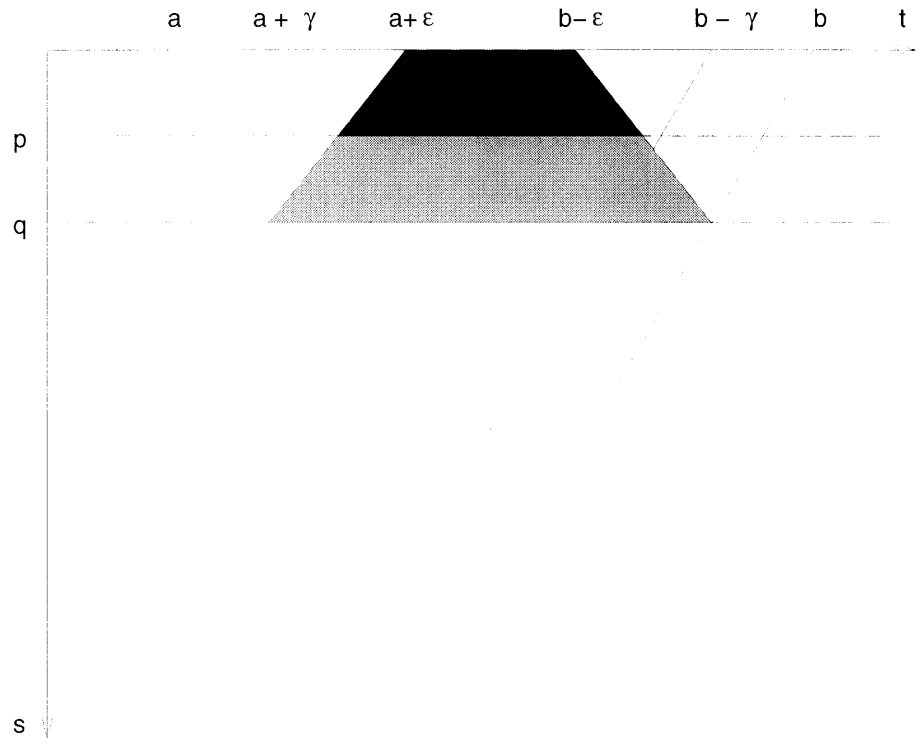

Figure 2.2: Why Theorem 2.2.6 follows from Theorems 2.2.4 and 2.2.5. (The trapezoid shaded in black is the p-trapezoid of influence of the interval $(a + \varepsilon, b - \varepsilon)$; its union with the trapezoid shaded in gray is the q-trapezoid of influence of the same interval.)

2) $|W_\psi f(t,s)| \leq A_\varepsilon s^\alpha$ *at every interior point of the triangle of sovereignty of* $(a+\varepsilon, b-\varepsilon)$, $\forall\varepsilon$ *such that* $0 < \varepsilon < \frac{a+b}{2}$. *(Here* A_ε *is a constant depending only on* ε*.)* □

(Again, here we have taken advantage of the fact that the case of $\alpha \in (0,1)$ can be extended to $\alpha \in (n, n+1)$ with the help of the Inductive Lemma.)

So far, we have used CWT to compute Lipschitz regularity on an interval. CWT can also be used to characterize the regularity of a function at a point.

Theorem 2.2.7 *Suppose that* $\psi(t)$ *is a compactly supported wavelet. If a bounded function* $f(t)$ *is Lipschitz* α *at* t_0, *with* $\alpha \in (0,1]$, *then there exists a scale* s_0 *such that*

$$|W_\psi f(t,s)| \leq A(s^\alpha + |t-t_0|^\alpha), \tag{2.12}$$

for all t *in some neighborhood of* t_0 *and for all* $s < s_0$.

Proof. See Appendix B. (The proofs of all theorems in this section consist of coming up with clever bounds for various integrals, which was illustrated by the proofs of Theorems 2.2.1 and 2.2.5. The proofs of all other theorems discussed in this section are in the references [3] and [10].) □

There is a converse theorem:

Theorem 2.2.8 *Suppose that* $\psi(t)$ *is a compactly supported wavelet. Suppose also that* $f \in L^2(R)$ *is bounded and continuous. Then* $f(t)$ *is Lipschitz* α *at* t_0 *if, for some* $\gamma > 0$ *and* $\alpha \in (0,1)$, *the following two conditions hold for all* t *in some neighborhood of* t_0 *and for all* $s < s_0$*:*

$$|W_\psi f(t,s)| \leq As^\gamma \tag{2.13}$$

$$|W_\psi f(t,s)| \leq B\left(s^\alpha + \frac{|t-t_0|^\alpha}{|log|t-t_0||}\right) \tag{2.14}$$

Proof. *See [3], page 49.* □

From the Inductive Lemma, similar theorems for higher order local regularity immediately follow.

It turns out that we do not need the whole CWT in order to compute the Lipschitz regularity of a function. The next three theorems show that we can estimate the Lipschitz exponents of a function using only its CWT modulus maxima (which we shall abbreviate by CWTMM).

Definition 2.2.3 *A* **modulus maximum** *of a wavelet transform $W_\psi f(t,s)$ is any point (t_0, s_0) such that*

$$\left.\frac{\partial |W_\psi f(t,s)|}{\partial t}\right|_{t=t_0, s=s_0} = 0 \tag{2.15}$$

The following theorems can be proved using the results discussed earlier in this section:

Theorem 2.2.9 *Suppose that $\psi(t)$ is a compactly supported wavelet with n vanishing moments. Let $f(t) \in L^1([a,b])$. If there exists such a scale s_0 that for $s < s_0$ and $t \in (a,b)$, $W_\psi f(t,s)$ has no modulus maxima, then for any $\varepsilon > 0$ and $\alpha < n$, $f(t)$ is uniformly Lipschitz α on $(a+\varepsilon, b-\varepsilon)$ (provided, of course, that $\varepsilon < \frac{a+b}{2}$).*

Proof: see [10], pages 639-641. □

If we define "a singularity" as a place where the Lipschitz regularity of a function is less than n, then this theorem says that it is useless to look for singularities at the places where there are no CWTMM. This is nice, but now we would like to know where exactly to look for singularities and how to compute the corresponding Lipschitz exponents.

Theorem 2.2.10 *Suppose that $\psi(t)$ is a compactly supported wavelet with n vanishing moments. Suppose also that all CWTMM that are inside of the rectangle defined by*

$$\begin{aligned} s &< s_0 \\ a < t &< b \end{aligned}$$

are actually inside of the triangle defined by

$$s < s_0 \tag{2.16}$$

$$t_0 - Cs \leq t \leq t_0 + Cs \tag{2.17}$$

for some $t_0 \in (a,b)$ and some constant $C > 0$. Then each point of the interval (a,b), except t_0, has a neighborhood where $f(t)$ is uniformly Lipschitz n. Let $\alpha < n$ be a non-integer. The function $f(t)$ is Lipschitz α at t_0 if and only if, at each modulus maximum (t,s) in the triangle defined by (2.16) and (2.17),

$$|W_\psi f(t,s)| \leq As^\alpha, \tag{2.18}$$

for some constant $A > 0$.

Proof: see [10], page 641. □

This is almost what we want. This theorem says that, in order to compute the Lipschitz regularity of $f(t)$ at a point t_0, we need to trace all maximum lines of the modulus of CWT of $f(t)$ which go to t_0. The next theorem will allow us to trace just one line of maxima.

Theorem 2.2.11 *Suppose that $\psi(t)$ is a wavelet with n vanishing moments. Suppose that supp$[\psi(t)] = [-r,r]$. Let a, t_0, and b be three constants, $a < t_0 < b$. Let t=X(s) be a curve in the scale space (t,s) lying entirely inside the trapezoid of influence of the point $t_0 \in (a,b)$. Suppose also that for all points (t,s) of the rectangle defined by*

$$\begin{aligned} s &< s_0 \\ a < t &< b, \end{aligned}$$

$W_\psi f(t,s)$ has a constant sign and

$$|W_\psi f(t,s)| \leq As^\nu, \tag{2.19}$$

for some $\nu > 0$. If

$$|W_\psi f(X(s),s)| \leq As^\gamma, \tag{2.20}$$

with $0 \leq \gamma \leq n$, then $f(t)$ is Lipschitz α at t_0, for any $\alpha < \gamma$.

Proof: see [10], pages 641-642. □

It has already been mentioned that Mallat's noise removal algorithm utilizes the computation of Lipschitz exponents of a function. Theorems 2.2.9, 2.2.10, and

2.2.11 tell us that we can compute the Lipschitz regularity at a point t_0 from the decay in s of the CWTMM along the curve of maxima converging to t_0. Now, in order to be able to estimate Lipschitz regularity of a function from its samples, all we need is to implement the wavelet transform on a digital computer. The discussion of this implementation is the content of the next section.

2.3 Dyadic Wavelet Transform (DWT)

Computation of the CWT for a dense set of scales and times is extremely inefficient. In order to be able to perform fast computations, we restrict our attention to the scales which are integer powers of two. The wavelet transform taken along those scales is called the dyadic wavelet transform:

Definition 2.3.1 *The* **dyadic wavelet transform (DWT)** *of a function $f(t)$, $\mathbf{W}_\psi f$, is defined as the following sequence of functions:*

$$\mathbf{W}_\psi f \stackrel{\text{def}}{=} \{W_\psi f(t, 2^j)\}_{j \in Z}$$

Note that $\mathbf{W}_\psi$ is a linear operator from L^2 to the space of all countable sequences of L^2 functions.

The following theorem gives a sufficient condition for invertibility of the operator $\mathbf{W}_\psi$.

Theorem 2.3.1 *Suppose that the wavelet $\psi(t)$ is an even function (i.e., that $\hat{\psi}(\omega)$ is real.) If there exist two positive constants A_1 and B_1 such that, for any $\omega \in R$,*

$$A_1 \leq \sum_{j=-\infty}^{\infty} |\hat{\psi}(2^j \omega)|^2 \leq B_1, \tag{2.21}$$

then any $f \in L^2(R)$ can be recovered from its dyadic wavelet transform.

Proof. From (2.21), we have:

$$\sum_{j=-\infty}^{\infty} \hat{\psi}(2^j \omega) \frac{\hat{\psi}(2^j \omega)}{B_1} \leq 1, \ \forall \omega \in R \tag{2.22}$$

$$\sum_{j=-\infty}^{\infty} \hat{\psi}(2^j \omega) \frac{\hat{\psi}(2^j \omega)}{A_1} \geq 1, \ \forall \omega \in R \tag{2.23}$$

Define a subspace D of $L^2(R)$ by $D = \{\mu \in L^2(R) | 0 < \sum_{j=-\infty}^{\infty} \mu^2(2^j\omega) < \infty\}$. Define a functional $\mathbf{K}$ on D by $\mathbf{K}\mu = \sum_{j=-\infty}^{\infty} \hat{\psi}(2^j\omega)\mu(2^j\omega)$. Note that $\mathbf{K}$ is continuous. Now, (2.22) and (2.23) say that $\mathbf{K}\left(\frac{\hat{\psi}(2^j\omega)}{B_1}\right) \leq 1$, and $\mathbf{K}\left(\frac{\hat{\psi}(2^j\omega)}{A_1}\right) \geq 1$. Since $\mathbf{K}$ is continuous, the Intermediate Value Theorem [17] says that there exists a (not necessarily unique) function $\hat{\eta}$ such that $\mathbf{K}\hat{\eta} = 1$. I.e.,

$$\sum_{j=-\infty}^{\infty} \hat{\psi}(2^j\omega)\hat{\eta}(2^j\omega) = 1 \tag{2.24}$$

Let us now use the fact that $W_\psi f(t, 2^j) = f * \psi_{2^j}(t)$, where $\psi_{2^j}(t) = \frac{1}{2^j}\psi(\frac{t}{2^j})$. Since the Fourier transform converts convolution to multiplication, and since $\mathcal{F}[\psi_{2^j}(t)] = \hat{\psi}(2^j\omega)$, we can write the Fourier transform of the wavelet transform of $f(t)$ as:

$$\hat{W}_\psi f(\omega, 2^j) = \hat{f}(\omega)\hat{\psi}(2^j\omega) \tag{2.25}$$

Now notice that

$$\sum_{j=-\infty}^{\infty} \hat{W}_\psi f(\omega, 2^j)\hat{\eta}(2^j\omega) = \sum_{j=-\infty}^{\infty} \hat{f}(\omega)\hat{\psi}(2^j\omega)\hat{\eta}(2^j\omega) = \hat{f}(\omega) \sum_{j=-\infty}^{\infty} \hat{\psi}(2^j\omega)\hat{\eta}(2^j\omega) = \hat{f}(\omega),$$

where we used (2.25) to make the first transition and (2.24) to make the last one. Taking the inverse Fourier transform of the last equation, we get:

$$f(t) = \sum_{j=-\infty}^{\infty} W_\psi f(t, 2^j) * \eta_{2^j}(t) \tag{2.26}$$

This completes the proof. □

We denote by $\mathbf{W}_\psi^{-1}$ the operator from the space of all countable sequences of L^2 functions to L^2 defined by

$$\mathbf{W}_\psi^{-1}(\{g_j(t)\}_{j\in Z}) = \sum_{j=-\infty}^{\infty} g_j * \eta_{2^j}(t) \tag{2.27}$$

It is important to realize that $\mathbf{W}_\psi^{-1}$ is *not* the inverse of $\mathbf{W}_\psi$, because a sequence $\{g_j(t)\}_{j\in Z}$ is not necessarily the DWT of an L^2 function. Rather, $\mathbf{W}_\psi^{-1}$ is a pseudo-inverse, because $\mathbf{W}_\psi^{-1} \circ \mathbf{W}_\psi = Identity$ on L^2.

We are not ready yet for the numerical implementation of the wavelet transform, because the computer cannot deal with infinite range of scales. We are going to avoid this problem by fixing the finest scale at 1 and computing the set

$\{W_\psi f(t, 2^j)\}_{1 \le j \le J}$, for an arbitrary J. We will show that all coarse-scale information can be accumulated in one function.

Assume that $\hat{\psi}(\omega)\hat{\eta}(\omega)$ is real, even, and positive. Let $\hat{\phi}(\omega) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \hat{\psi}(2^j\omega)\hat{\eta}(2^j\omega)$. Then we have:

$$\sum_{j=1}^{J} \hat{W}_\psi f(\omega, 2^j)\hat{\eta}(2^j\omega) = \sum_{j=1}^{J} \hat{f}(\omega)\hat{\psi}(2^j\omega)\hat{\eta}(2^j\omega) = \hat{f}(\omega)(\hat{\phi}(\omega) - \hat{\phi}(2^J\omega))$$

Now define $S_\phi f(t, 2^j) = f * \phi_{2^j}(t)$ and notice that $\hat{S}_\phi f(\omega, 1) - \hat{S}_\phi f(\omega, 2^J) = \hat{f}(\omega)(\hat{\phi}(\omega) - \hat{\phi}(2^J\omega))$. Therefore,

$$\begin{aligned} \hat{S}_\phi f(\omega, 1) &= \hat{S}_\phi f(\omega, 2^J) + \sum_{j=1}^{J} \hat{W}_\psi f(\omega, 2^j)\hat{\eta}(2^j\omega) \Rightarrow \\ S_\phi f(t, 1) &= S_\phi f(t, 2^J) + \sum_{j=1}^{J} W_\psi f(t, 2^j) * \eta_{2^j}(t) \end{aligned}$$

That is, $S_\phi f(t, 1)$ can be recovered from $\{W_\psi f(t, 2^j)\}_{j=1}^{J}$ and $S_\phi f(t, 2^J)$. It turns out [12] that any finite-energy discrete signal $d[n]$ can be represented as $d[n] = S_\phi f(n, 1)$ for some $f(t)$. There are many wavelets for which the discrete sequences $W_\psi f(n, 2^{j+1})$ and $S_\phi f(n, 2^{j+1})$ can be quickly computed from $S_\phi f(n, 2^j)$ and vice versa. This means that we can compute $\{W_\psi f(n, 2^j)\}_{j=1}^{J}$ and $S_\phi f(n, 2^J)$ given any discrete sequence $d[n] = S_\phi f(n, 1)$. Conversely, we can also reconstruct $d[n]$ from $\{W_\psi f(n, 2^j)\}_{j=1}^{J}$ and $S_\phi f(n, 2^J)$. One class of such wavelets is described in [11]. The wavelet that was used in this thesis is described in Chapter 4. We finish this section by defining the discrete dyadic wavelet transform of a signal up to the scale 2^J, which is the portion of the CWT that we use in computer simulations.

Definition 2.3.2 *The* **discrete dyadic wavelet transform (DDWT)** *up to the scale 2^J of a discrete signal $d[n] = S_\phi f(n, 1)$ (or continuous signal $f(t)$) is the following collection of $J + 1$ discrete sequences:*

$$(\{W_\psi f(n + w, 2^j)\}_{j=1}^{J},\ S_\phi f(n + w, 2^J)), \tag{2.28}$$

where w is a sampling shift that depends only on the wavelet $\psi(t)$.

2.4 An Algorithm for Reconstructing a Close Approximation of a Signal from Its Dyadic Wavelet Transform Modulus Maxima

The three preceding sections dealt with implementing the wavelet transform on a digital computer and using it to estimate the Lipschitz regularity of a function. The only building block of Mallat's noise removal algorithm that we have not discussed is an algorithm for approximating a signal given its Dyadic Wavelet Transform Modulus Maxima (DWTMM). This algorithm was developed by Mallat and Zhong. In this section, which is almost entirely borrowed from [11], principal features of the algorithm are described.

In general, it is impossible to reconstruct a signal from its DWTMM, because several (in fact, uncountably many) signals can have the same DWTMM [13]. However, it seems that the L^2 norm of the difference of any two functions having the same DWTMM is small compared to the norms of the functions themselves. (To the best of our knowledge, this conjecture has been neither proven nor even quantified.) We assume that it is true. In that case, it makes sense to try to reconstruct *some* function from a given set of DWTMM, because this reconstruction then should be a good approximation to any other function possessing the same DWTMM.

We shall now examine the problem of reconstructing a function from its DWTMM. We start by characterizing the set of all functions $h(t)$ such that the DWTMM of $h(t)$ is the same as the DWTMM of some fixed function $f(t)$. We suppose that the wavelet $\psi(t)$ is continuously differentiable. Since $W_\psi f(t, 2^j)$ is the convolution of $f(t)$ with $\psi_{2^j}(t)$, $W_\psi f(t, 2^j)$ is also continuously differentiable and has, at most, a countable number of modulus maxima.

Let $\{t_n^j\}_{n\in Z}$ be the set of all locations where $|W_\psi f(t, 2^j)|$ is locally maximum. Then the constraints on $h(t)$ can be decomposed into two conditions:

(1) $W_\psi h(t_n^j, 2^j) = W_\psi f(t_n^j, 2^j), \ \ \forall n, j \in Z$, and

(2) DWTMM of $h(t)$ are located at the points $\{t_n^j\}_{n,j\in Z}$.

Let us examine these two conditions. Condition (1) can be rewritten in the following way:

$$\int_{-\infty}^{\infty} f(t)\psi_{2^j}(t_n^j - t)\,dt = \int_{-\infty}^{\infty} h(t)\psi_{2^j}(t_n^j - t)\,dt, \quad \forall n, j \in Z \Leftrightarrow$$
$$\Leftrightarrow \; < f(t),\, \psi_{2^j}(t_n^j - t) > = < h(t),\, \psi_{2^j}(t_n^j - t) >, \quad \forall n, j \in Z,$$

which is equivalent to $h(t)$ and $f(t)$ having the same orthogonal projections on the space $U = Cl(Sp\{\psi_{2^j}(t_n^j - t)\}_{n,j\in Z})$. (Here Cl means "closure" and Sp means "span".) In other words, condition (1) is equivalent to: $h = f + g$ with $g \in U^{\perp}$, which holds if and only if $h \in f + U^{\perp}$.

Condition (2) is more difficult to analyze, because it is non-convex. We approximate it with another constraint. Namely, we impose that $|W_\psi h(t, 2^j)|^2$ be as small as possible on average, at every scale. This, together with condition (1), generally creates modulus maxima at or near the positions $\{t_n^j\}_{n,j\in Z}$. To have as few additional modulus maxima as possible, we also minimize the energy of the derivative of $W_\psi h(t, 2^j)$ – again, at every scale. Notice that $\frac{dW_\psi h(t,2^j)}{dt} = h(t) * \frac{d}{dt}(\psi(\frac{t}{2^j})\frac{1}{2^j}) = h(t) * \psi'(\frac{t}{2^j})\frac{1}{2^j}\frac{1}{2^j}$ – i.e., the integral of the absolute value of the function with which we convolve $h(t)$ decreases by a factor of 2 every dyadic scale. Since we would like the terms at different scales to carry the same weight in the quantity that we are going to minimize, we must weigh $\left\|\frac{dW_\psi h(t,2^j)}{dt}\right\|$ by a factor of 2^j. This means that we minimize the following quantity:

$$|||h|||^2 \stackrel{\text{def}}{=} \sum_{j=-\infty}^{\infty} (\|W_\psi h(t, 2^j)\|^2 + 2^{2j} \left\|\frac{dW_\psi h(t, 2^j)}{dt}\right\|^2) \tag{2.29}$$

over all $h \in f + U^{\perp}$, where 2^{2j} is a normalizing factor expressing the fact that $W_\psi h(t, 2^j)$ is smoother at larger scales. Using (2.21) for both $\psi(t)$ and $\psi'(t) = \frac{d\psi}{dt}$, we get:

$$A \leq \sum_{j=-\infty}^{\infty} |\hat{\psi}(2^j\omega)|^2 + \sum_{j=-\infty}^{\infty} |\hat{\psi}'(2^j\omega)|^2 \leq B$$

Multiplying this inequality by $|h(\omega)|^2$, integrating the result from $\omega = -\infty$ to $\omega = \infty$, and using Parseval's relation, we get:

$$A\|h\|^2 \leq \sum_{j=-\infty}^{\infty} [(h(t) * \psi_{2^j}(t))^2 + (h(t) * \psi'_{2^j}(t))^2] \leq B\|h\|^2 \quad \Leftrightarrow$$

$$A\|h\|^2 \leq |||h|||^2 \leq B\|h\|^2 \tag{2.30}$$

Hence, $|||\cdot|||$ is a norm over $L^2(R)$, which is equivalent to $\|\cdot\|$. So, we replaced condition (2) with a minimization of a norm. We showed previously that condition (1) restricts h to the set $f + U^{\perp}$. This is a convex set, since $U^{\perp}$ is a linear space. This means that we have converted our problem into minimization of a norm over a convex set. The new problem has a unique solution which, we hope, is a good approximation to any solution of the original problem.

Now we shall outline the algorithm that computes the solution of our minimization problem. Instead of computing the solution directly, we first compute its wavelet transform with an algorithm based on alternating projections.

Let $\mathbf{K}$ be the space of all sequences of functions $\{g_j(t)\}_{j\in Z}$ such that

$$|\{g_j(t)\}_{j\in Z}|^2_{\mathbf{K}} \stackrel{\text{def}}{=} \sum_{j=-\infty}^{\infty} (\|g_j\|^2 + 2^{2j}\left\|\frac{dg_j}{dt}\right\|^2) < \infty \tag{2.31}$$

Let $\mathbf{V}$ be the space of all DWT's of $L^2(R)$ functions. Equations (2.29) and (2.30) then say that $\mathbf{V} \subset \mathbf{K}$. Now let $\mathbf{\Gamma}$ be the set of all sequences of functions $g_j(t)_{j\in Z} \in \mathbf{K}$ such that for all maxima positions t_n^j,

$$g_j(t_n^j) = W_\psi f(t_n^j, 2^j)$$

We must therefore find the element of $\mathbf{\Lambda} = \mathbf{V} \cap \mathbf{\Gamma}$ whose norm defined by (2.31) is minimum. In other words, we need to find the orthogonal projection of the zero element of $\mathbf{K}$ onto $\mathbf{\Lambda}$. This is done by alternating projections on $\mathbf{V}$ and $\mathbf{\Gamma}$. We already know the projection operator onto $\mathbf{V}$: it is $\mathbf{W}_\psi \circ \mathbf{W}_\psi^{-1}$ (see Definition 2.3.1 and (2.27)). It turns out [11] that this operator is orthogonal if the wavelet is antisymmetric (as in [10] and [11]) or symmetric (as in this thesis – see Chapter 4.) The orthogonal projector $\mathbf{P}_{\mathbf{\Gamma}}$ on $\mathbf{\Gamma}$ is described in [11]. Now it follows from [22] that repeated application of $\mathbf{W}_\psi \circ \mathbf{W}_\psi^{-1} \circ \mathbf{P}_{\mathbf{\Gamma}}$ to any element x of $\mathbf{K}$ converges to $\mathbf{P}_{\mathbf{\Lambda}} x$, where $\mathbf{P}_{\mathbf{\Lambda}}$ is the orthogonal projector on $\mathbf{\Lambda}$.

2.5 A Robust Noise Removal Algorithm of Mallat

We saw in Section 2.2 that we could compute the Lipschitz regularity of a function at any point t_0 from the wavelet transform modulus maxima. Section 2.4 provided us with an algorithm to estimate a signal from its wavelet transform modulus maxima. Combining the two results, we have the following noise removal algorithm.

Problem. Suppose that we observe samples of the following signal:

$$y(t) = f(t) + \nu(t),$$

where $\nu(t)$ is a noise process whose realizations have Lipschitz regularity less than k, and $f(t)$ is the useful signal which is Lipschitz at least k at every point.

Procedure. Compute the wavelet transform of $y(t)$ and compute its modulus maxima. From their evolution across scales, determine local Lipschitz regularities. Remove all those maxima corresponding to the Lipschitz regularities which are less than k. From the rest of the maxima, reconstruct a signal according to the procedure described in Section 2.4. (We translate everything into discrete domain using the procedure described in Section 2.3.)

For example, the realization of a white Gaussian noise is a distribution whose uniform Lipschitz regularity is -0.5. Suppose that the worst singularity of the signal that we want to recover is Lipschitz 1. Then we would set the threshold k to be anywhere between -0.5 and 1. Notice, however, that the algorithm does not require the noise to be Gaussian. The only requirement is that the Lipschitz regularities of the noise and the signal be able to be differentiated.

Unfortunately, we cannot apply this algorithm directly to the problem posed in this thesis. The problem is that the algorithm does not allow us to use any additional information that is known about the signal (besides its Lipschitz exponents.) The reconstruction depicted in the Figure 2.3 is extremely good if we have no information about the original signal. However, if we know that it is a linear spline with one knot, we would like a better estimate! Besides, in an application such as the

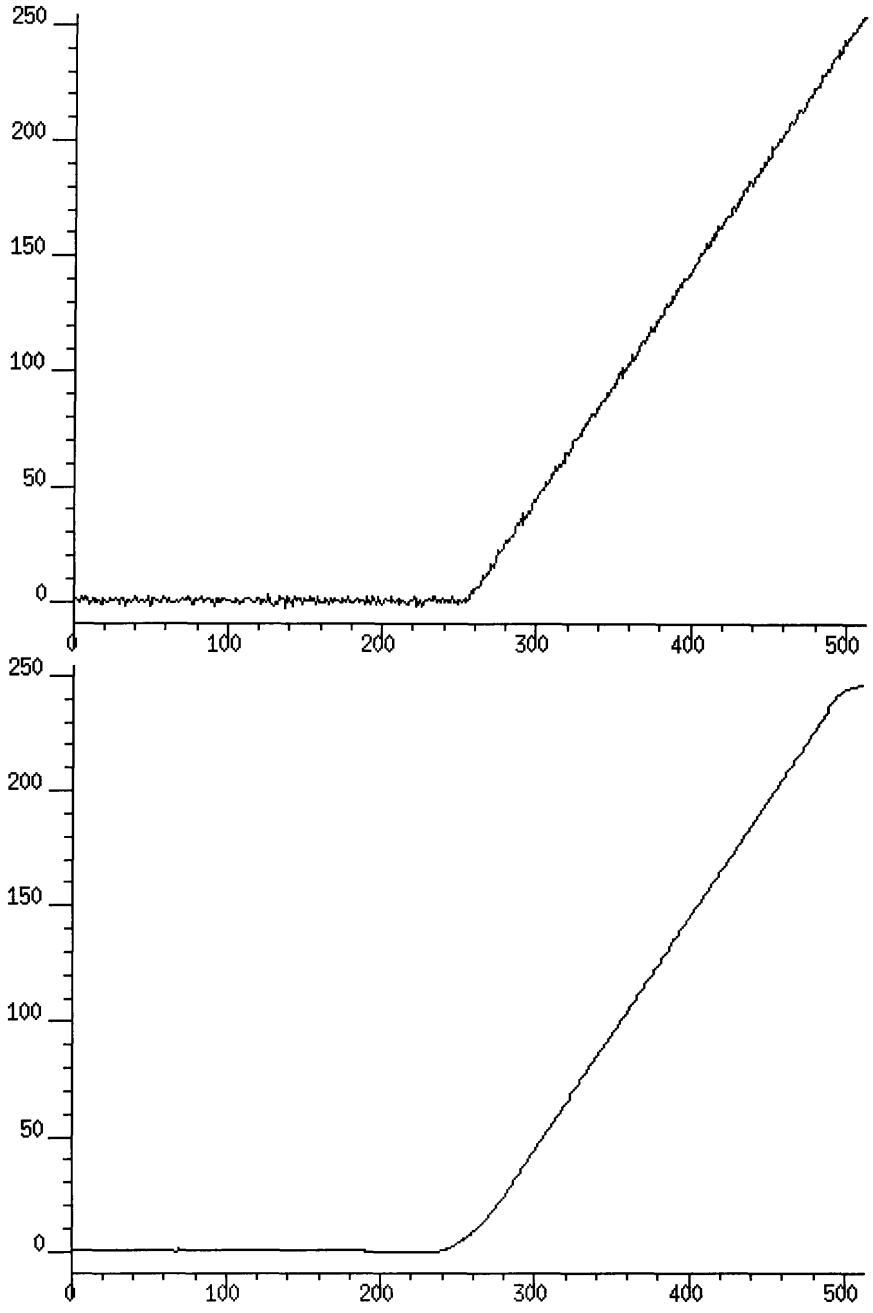

Figure 2.3: A noisy linear spline (top) and the result of putting it through the noise-removal procedure of Mallat (bottom)

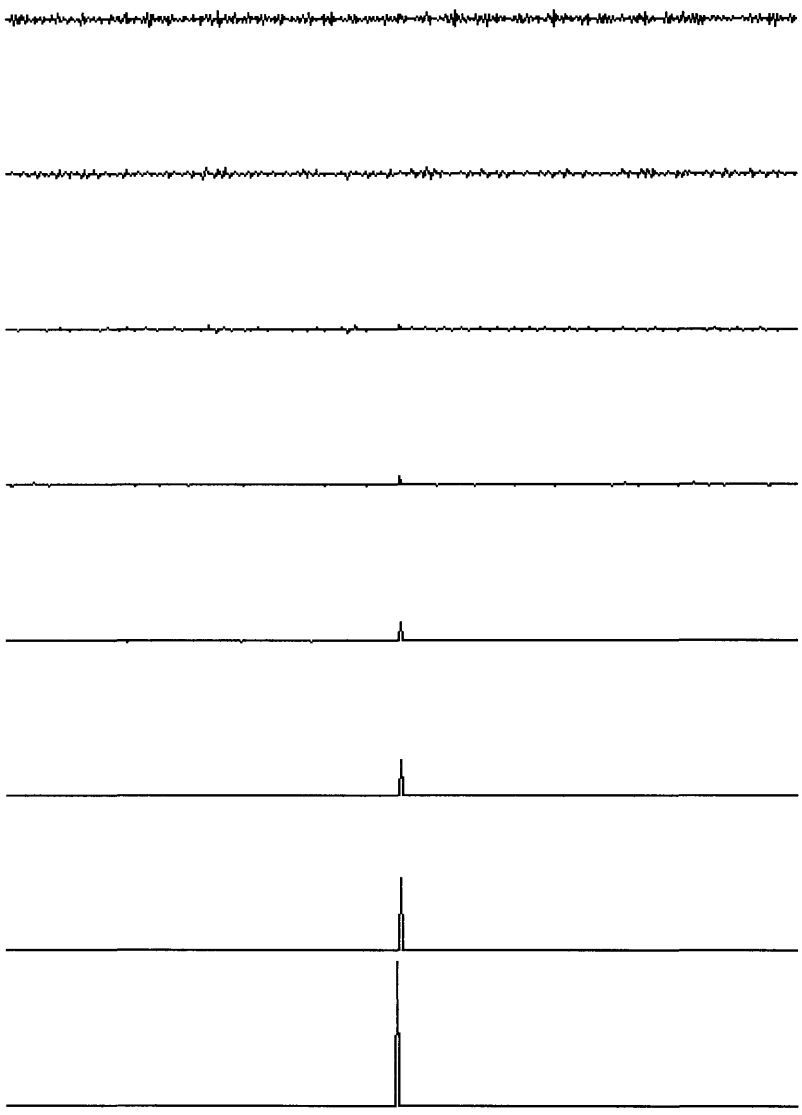

Figure 2.4: The extrema at the first 8 dyadic scales of the wavelet transform of the signal depicted in the top portion of Figure 2.3

detection of abrupt changes, the location of the knot and the corresponding slope change is precisely the information we want. Our goal is to modify this procedure so as to be able to incorporate our knowledge of the fact that the original signal is a spline. What we need to do is to replace the part that reconstructs the signal from the wavelet transform modulus maxima by some parameter estimation algorithm. For example, if we have a linear spline with one knot, we would like to know three parameters: the location of the knot and the slopes of the linear pieces. We know that the wavelet transform of such a signal has just one maxima line (which corresponds to the knot.) So, our task is to choose, among many possible maxima lines of the wavelet transform of noisy observations, the one which matches most closely a singularity with Lipschitz regularity equal to 1. Figure 2.4 shows the wavelet transform modulus maxima of the signal depicted in Figure 2.3, at the first 8 dyadic scales. We can see that the extrema corresponding to noise are absent from the coarse (bottom) scales, and start to appear at finer scales. Their amplitudes are generally less at lower scales. On the other hand, the extrema corresponding to the knot are strongest at the coarser scales, with geometrically decreasing magnitude at finer scales (i.e., towards the top). These observations suggest that perhaps we could formulate the problem of estimating the movement of the extrema across scale as a target tracking problem, with extrema corresponding to noise modeled as false targets whose density and amplitude decrease at coarser scales. This is the motivation for the brief discussion of a general multi-target tracking problem presented in the next chapter.

Chapter 3

A MULTI-TARGET TRACKING ALGORITHM

3.1 Introduction

This chapter is a brief review of the multi-target tracking problem. Most of the results described here are taken from [16], [9] and [19]. It is not our goal to present the subject in great depth or generality; for broader and deeper treatment, the reader can consult [16] and [9], as well as many excellent references cited in those two papers.

3.2 Exact Algorithm

We assume that at each time t, we have $N_{TT}(t)$ targets each of which is described by a vector $\mathbf{x}_i(t)$ of n parameters. The number of targets $N_{TT}(t)$ may change with time. We model the evolution in time of every state vector $\mathbf{x}_i(t)$ by the equation

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + G\mathbf{w}(t), \tag{3.1}$$

where A and G are known matrices and $\mathbf{w}(t)$ is a discrete white Gaussian noise process with zero mean and covariance Q. We also assume that we make $M(t)$

measurements $\{\mathbf{y}_m(t)\}_{m=1}^{M(t)}$ at time t, where $M(t)$ may vary with time. Some of the measurements are due to targets and the others are false alarms. (Note that the number of measurements that are due to targets is not necessarily $N_{TT}(t)$, because some targets may not be detected. The model for the probability of a missed detection P_D and the probability of a false alarm P_F depends on the particular application.) We assume that the following relation holds between the state vector $\mathbf{x}(t)$ of a target and the measurement vector $\mathbf{y}(t)$ which came from this target:

$$\mathbf{y}(t) = C\mathbf{x}(t) + \mathbf{v}(t), \tag{3.2}$$

where C is a known measurement matrix and $\mathbf{v}(t)$ is a zero-mean white Gaussian noise with covariance R.

If every measurement could be associated with a particular, unique target, then the optimal estimate of every $\mathbf{x}_i(t)$ is given by the Kalman filter [8]. After a measurement $\mathbf{y}(t)$ is received, the best estimate of the corresponding $\mathbf{x}(t)$ based on $\mathbf{y}(\tau)$ up to time t is

$$\hat{\mathbf{x}}(t|t) = \hat{\mathbf{x}}(t|t-1) + P(t|t)C^T R^{-1}(\mathbf{y}(t) - C\hat{\mathbf{x}}(t|t-1)), \tag{3.3}$$

where the corresponding error covariance $P(t|t) = E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t))(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t))^T]$ is given by

$$P(t|t) = P(t|t-1) - P(t|t-1)C^T[CP(t|t-1)C^T + R]^{-1}CP(t|t-1) \tag{3.4}$$

Now we can predict $\mathbf{x}(t+1)$ based on $\mathbf{y}(\tau)$ up to time t:

$$\hat{\mathbf{x}}(t+1|t) = A\hat{\mathbf{x}}(t|t) \tag{3.5}$$

$$P(t+1|t) = AP(t|t)A^T + GQG^T \tag{3.6}$$

However, before we can use the Kalman filter, we must solve the data association problem. Namely, when we receive a set of measurements, we must know how to associate each measurement with a target.

Let $Y(t) \stackrel{\text{def}}{=} \{\mathbf{y}_m(t),\ i = 1, 2, ..., M(t)\}$ be the set of all measurement vectors at time t. Let $Y_{cum}(t) \stackrel{\text{def}}{=} \{Y(\tau),\ \tau = 1, 2, ..., t\}$ be the set of all measurements for

times $1, 2, ..., t$. A **global hypothesis** $\Omega_l(t)$ at time t is a rule which associates each measurement vector in $Y_{cum}(t)$ with either a particular target or a false alarm. Usually, one measurement vector $\mathbf{y}_i(t)$ is not allowed to be associated with more than one target under one hypothesis. Conversely, two measurement vectors are not allowed to be associated with the same target. Let $\Omega(t) \stackrel{\text{def}}{=} \{\Omega_l(t),\ l = 1, 2, ..., L_t\}$ be the set of all possible global hypotheses at time t. An **association hypothesis** $\kappa_g(t)$ for a data set $Y(t)$ at any time t is a rule which associates each measurement vector in $Y(t)$ with either a particular target or a false alarm. Thus, we may view $\Omega_l(t)$ as the joint hypothesis formed from the prior hypothesis $\Omega_{l_1}(t-1)$ and the current association hypothesis $\kappa_g(t)$, and therefore

$$\begin{aligned} Pr(\Omega_l(t)|Y_{cum}(t)) &= Pr(\Omega_{l_1}(t-1), \kappa_g(t)|Y_{cum}(t)) = \\ & p(Y(t)|\Omega_{l_1}(t-1), \kappa_g(t), Y_{cum}(t-1)) \cdot Pr(\kappa_g(t)|\Omega_{l_1}(t-1), Y_{cum}(t-1)) \cdot \\ \cdot\ & Pr(\Omega_{l_1}(t-1)|Y_{cum}(t-1))\frac{1}{c}, \end{aligned} \tag{3.7}$$

where c is a normalization constant found by summing the rest of the right-hand side over all possible values of l_1 and g. The symbol "Pr" is used to denote probabilities and "p" is used for probability densities.

Let us now show how to evaluate the right-hand side of (3.7). Since in all practical applications the range of a sensor is limited, all measurements come from some bounded set. Let V be the volume of that set. We assume that the measurement from clutter or any target whose existence is not implied by the prior global hypothesis $\Omega_{l_1}(t-1)$ is uniformly distributed – i.e., it is equally likely to come from anywhere in the region covered by the sensor. Equation (3.2) implies that, for a target whose existence is implied by the prior global hypothesis, a measurement at time t conditioned on the measurements up to time $t-1$ is a normally distributed random vector with mean $C\hat{\mathbf{x}}(t|t-1)$ and covariance $CP(t|t-1)C^T + R$. Then

$$p(Y(t)|\Omega_{l_1}(t-1), \kappa_g(t-1), Y_{cum}(t-1)) = \prod_{m=1}^{M(t)} f_m, \tag{3.8}$$

where:

$f_m = \frac{1}{V}$ if the m-th measurement is from clutter or a target whose existence is not

implied by the prior global hypothesis, and
$f_m = \mathcal{N}(\mathbf{y}_m(t) - C\hat{\mathbf{x}}_m(t|t-1); 0, CP(t|t-1)C^T + R)$ if the measurement is from a target whose existence is implied by the prior global hypothesis $\Omega_{l_1}(t-1)$. Here, $\mathcal{N}(\mathbf{z}; \mathbf{z}_\mu, P)$ denotes the normal probability density function $\frac{1}{\sqrt{(2\pi)^n |P|}} \exp[-\frac{1}{2}(\mathbf{z} - \mathbf{z}_m)^T P^{-1}(\mathbf{z} - \mathbf{z}_m)]$

The second term on the right-hand side of (3.7) is the probability of a current association hypothesis $\kappa_g(t)$ given the prior global hypothesis $\Omega_{l_1}(t-1)$. Each association hypothesis $\kappa_g(t)$ includes the following information:

Number: The number of measurements associated with the prior targets $N_{DT,g}(t)$, the number of measurements associated with false targets $N_{FT,g}(t)$ and the number of measurements associated with newly detected targets $N_{NT,g}(t)$.

Configuration: Those measurements which are from previously known targets, those measurements which are from false targets, and those measurements which are from new targets.

Assignment: The specific source of each measurement which came from a previously known target.

The prior global hypothesis $\Omega_{l_1}(t-1)$ includes information as to the total number of the previously known targets $N_{TT,l_1}(t-1)$.

In order to compute the second term on the right-hand side of (3.7), we must have a probabilistic model for $N_{DT,g}(t)$,$N_{FT,g}(t)$ and $N_{NT,g}(t)$ which depends on the particular application under study. (Section 4.2 describes this model for the application considered in this thesis.) Once we know how to calculate $Pr(\Omega_m(t)|Y_{cum}(t))$ from $Pr(\Omega_{l_1}(t-1)|Y_{cum}(t-1))$, we proceed as follows. When a new set of measurements $Y(t)$ comes in, we calculate $Pr(\Omega_l(t)|Y_{cum}(t)) = Pr(\Omega_{l_1}(t-1), \kappa_g(t)|Y_{cum}(t))$ for all possible $\Omega_{l_1}(t-1)$ and $\kappa_g(t)$. We then update the Kalman filters for every possible hypothesis $\Omega_l(t)$. At the end, we choose the estimates of the targets resulting from the Kalman filter corresponding to the likeliest hypothesis.

3.3 Reducing the Number of the Hypotheses

The optimal filter discussed in the previous section requires large and exponentially growing memory and computational resources even if the number of measurements is relatively small. This is due to the fact that the number of hypotheses is large and exponentially increasing. This means that, for a practical implementation, we need to limit the number of hypotheses. The most natural thing to do is to avoid hypothesizing an association between a measurement and a target whose estimated tracks are far from the measurement. In other words, a target whose state vector is $\mathbf{x}_i(t)$ is associated with a measurement $\mathbf{y}_m(t)$ only if

$$(\mathbf{y}_m(t) - C\hat{\mathbf{x}}_i(t|t-1))^T \cdot (CP_k(t|t-1)C^T + R)^{-1} \cdot (\mathbf{y}_m(t) - C\hat{\mathbf{x}}_i(t|t-1)) \leq a^2, \quad (3.9)$$

where a is generally chosen between 1 and 3.

Once the probabilities of all feasible (in the sense of (3.9)) global hypotheses are computed, we retain a fixed number of most likely hypotheses. A slightly more sophisticated method (called an N-scan filter) is to prune the set of hypotheses $\Omega(t-N)$ at time t. That is, at any time $t \geq N$, all remaining global hypotheses $\Omega_l(t)$ have some hypothesis $\Omega_{l_1}(t-N)$ in common.

Chapter 4

ESTIMATION OF A LINEAR SPLINE WITH ONE KNOT

In this chapter, we show how to combine the algorithms presented in Chapters 2 and 3 to produce an algorithm for estimating the knot location and the slope change of a linear spline with one knot. Namely, we take the dyadic wavelet transform of an input signal and compute the local maxima of its absolute value. Then we use the target tracker described in the preceding chapter in order to estimate the knot location and the magnitude of the slope change.

4.1 An Appropriate Class of Wavelets

We start by describing a class of wavelets well suited to our application. Since we are looking to find a singularity whose Lipschitz exponent is 1, Theorem 2.2.11 tells us that we need a wavelet with at least 2 vanishing moments. An easy way to obtain such a wavelet is by differentiating a smoothing function 2 or more times. In general, if $\psi(t) = \frac{d^p\theta(t)}{dt^p}$, then

$$\begin{aligned}\int_{-\infty}^{+\infty} t^k \psi(t) dt &= \int_{-\infty}^{+\infty} t^k \frac{d^p\theta(t)}{dt^p} dt \\ &= \mathcal{F}\left[t^k \frac{d^p\theta(t)}{dt^p}\right](0)\end{aligned}$$

$$
\begin{aligned}
&= i^k \frac{d^k}{d\omega^k} \left(\mathcal{F} \left[\frac{d^p \theta(t)}{dt^p} \right] \right) (0) \\
&= i^k \frac{d^k}{d\omega^k} \left((i\omega)^p \Theta(\omega) \right) (0) \\
&= 0, \quad k = 0, ...p-1,
\end{aligned}
$$

provided that all derivatives that were used in this calculation exist, and that the Fourier transforms are continuous at zero. That is, if $\psi(t)$ is the p-th derivative of a function $\theta(t)$ which is sufficiently smooth and whose Fourier transform is sufficiently smooth, then $\psi(t)$ has p vanishing moments. A standard procedure for obtaining such a function $\theta(t)$ is convolving several box functions. It is easy to check that the convolution of $2k+3$ identical box functions is a bell-shaped, non-negative, compactly supported spline of order $2k+2$ (see Figure 4.1). We take this convolution as our $\theta(t)$ and set $\psi(t)$ to be equal to its second derivative. Therefore, $\psi(t)$ is the 2k-th order, compactly supported spline which looks like an upside-down Mexican hat (see Figure 4.2).

The wavelet transform of any function $f(t)$ at a fixed scale 2^j is then the second derivative of $f(t)$ smoothed by $\theta(t)$. Particularly, if $f(t)$ is a linear spline with one knot, $W_\psi f(t, 2^j)$ is a smoothed impulse whose maximum is at the knot location (see Figure 4.3). Thus, for a noiseless linear spline with one knot, there is just one extrema chain consisting of extrema whose location is constant across scale, which makes the corresponding state equation of our target very simple (see the next section.) If we were to make $\psi(t)$ a higher-order derivative of $\theta(t)$, multiple extrema tracks of different shapes would result (see Figure 4.4). This would make the model more complicated, without enhancing the performance of the tracker. (Indeed, the energy of the CWT would be spread over these several tracks, rather than being concentrated along one extrema track, thereby making noise removal more difficult.)

A fast recursive algorithm for computing the wavelet transform at dyadic scales (taken from [11]) is described in Appendix A. The particular wavelet used in computations resulted from setting $k = 1$. In other words, our smoothing function $\theta(t)$ is the convolution of five box functions. The corresponding quadratic spline

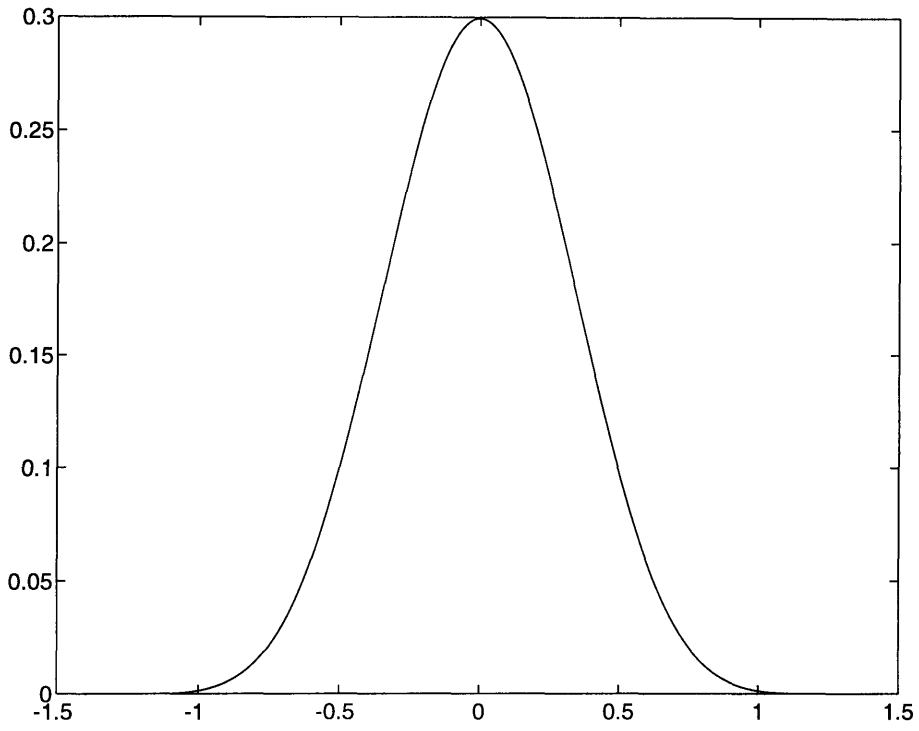

Figure 4.1: The shape of the smoothing function whose second derivative is $\psi(t)$

wavelet is $\psi(t) = r(t - \frac{1}{2})$, where

$$r(t) = \begin{cases} (2t + \frac{5}{2})^2 & \text{if } t \in [-\frac{5}{4}; -\frac{3}{4}] \\ -16t^2 - 20t - 5 & \text{if } t \in [-\frac{3}{4}; -\frac{1}{4}] \\ 24t^2 - \frac{5}{2} & \text{if } t \in [-\frac{1}{4}; \frac{1}{4}] \\ -16t^2 + 20t - 5 & \text{if } t \in [\frac{1}{4}; \frac{3}{4}] \\ (-2t + \frac{5}{2})^2 & \text{if } t \in [\frac{3}{4}; \frac{5}{4}] \\ 0 & \text{otherwise} \end{cases}$$

4.2 The Tracker of Maxima

Once we have computed the wavelet transform of the input sequence $d[n]$, we use the target tracker described in Chapter 3 to track the maxima of its absolute value across scale. For every maximum, we are interested in two parameters: its location and amplitude, which we therefore choose as our state variables x_1 and x_2. We showed at the end of the preceding section that, for a linear spline with one

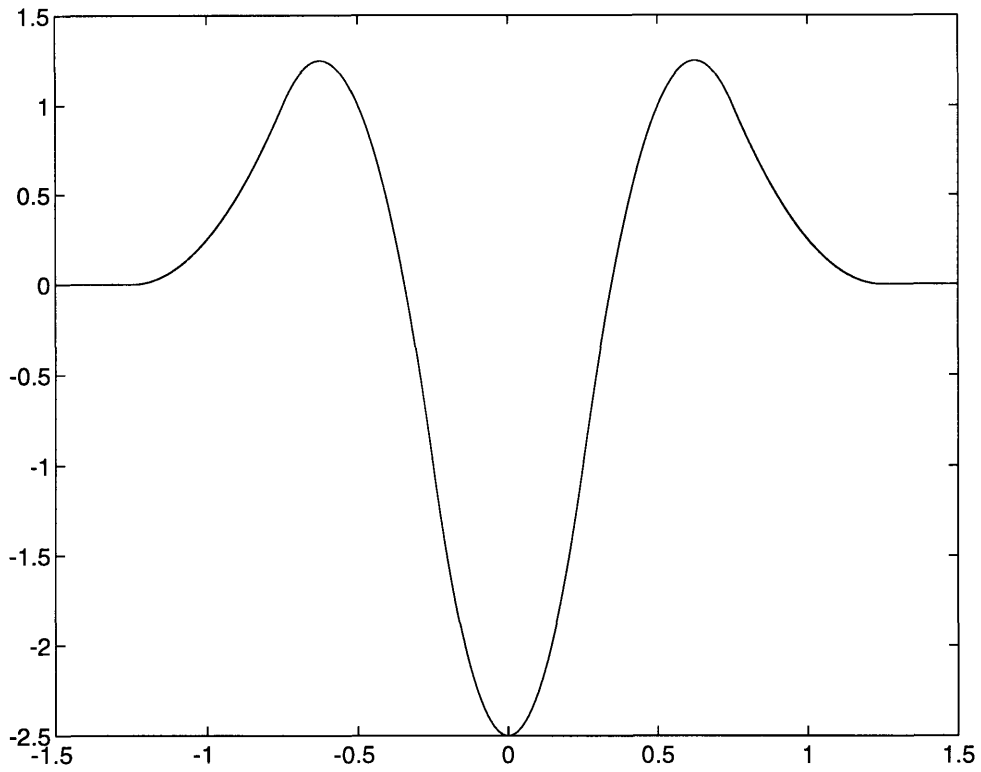

Figure 4.2: The wavelet $\psi(t)$ defined by (4.6), with $k = 1$

knot, the location of the maximum corresponding to the knot is fixed across scale and is the same as the location of the knot. Therefore, the state equation for this state variable is

$$x_1(j-1) = x_1(j) \tag{4.1}$$

This equation slightly differs from (3.1). First of all, since we are tracking the extrema across a sequence of dyadic scales rather than time, our independent variable is j. Second, we would like to start the tracker at the coarsest scales where the signals are smooth and therefore have few local maxima, which leads to a small number of initial hypotheses. We therefore move from coarse to fine scales, which corresponds to moving in the direction of decreasing j.

We obtain the second state equation by using the results presented in Section 2.2. The Lipschitz regularity of a linear spline at a knot is 1, and therefore the

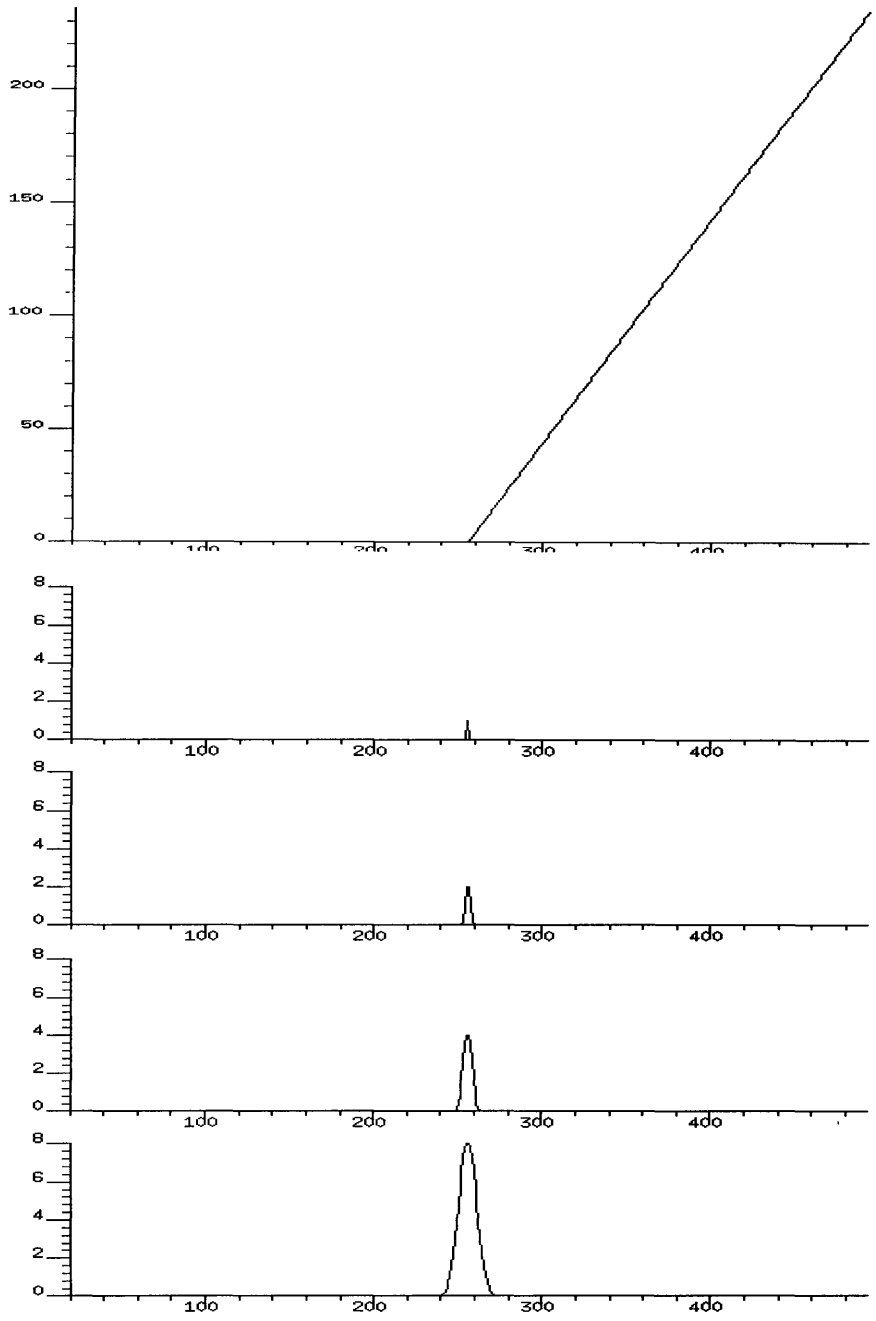

Figure 4.3: A linear spline (top) and its wavelet transform for scales 2^1,2^2,2^3, and 2^4, computed with the wavelet of Fig. 4.2

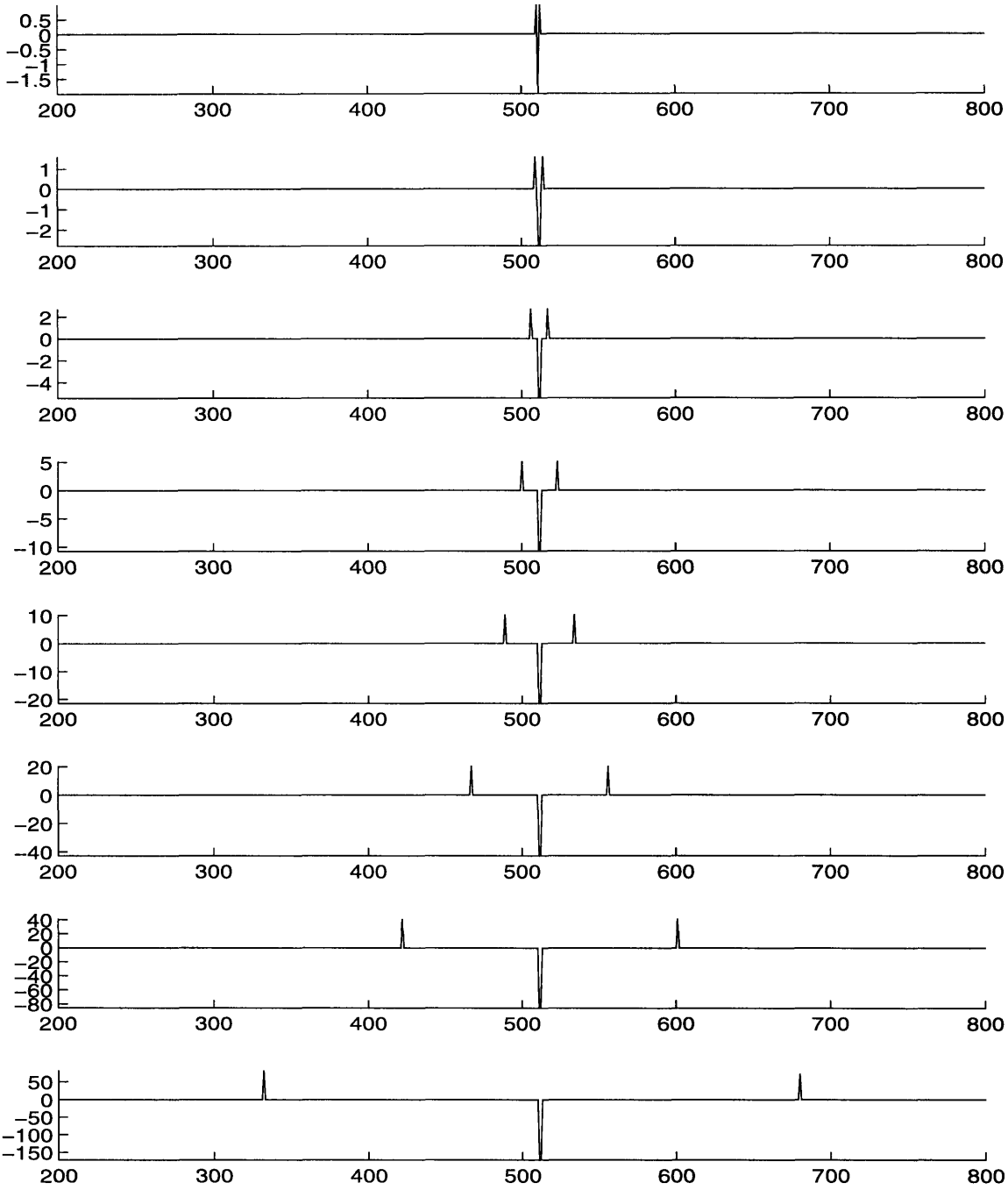

Figure 4.4: The wavelet transform extrema for a linear spline with one knot and a wavelet with four vanishing moments.

maximum at the scale 2^j has to be twice as large as the maximum at the scale 2^{j-1}:

$$x_2(j-1) = \frac{1}{2}x_2(j) \tag{4.2}$$

The slope difference at the knot is then given by $x_2(1)$. Combining (4.1) and (4.2) together, we get:

$$\mathbf{x}(j-1) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \mathbf{x}(j) \tag{4.3}$$

Since we "measure" our state variables directly, our measurement matrix C is an identity matrix, and so the measurement equation is

$$\mathbf{y}(j) = \mathbf{x}(j) + \mathbf{n}(j) \tag{4.4}$$

We model the noise as a zero-mean Gaussian random vector with uncorrelated components. The standard deviation of the second component at the finest scale, $\sigma_2(1)$, is estimated to be the standard deviation of the CWT maxima at that scale. Using the fact that both the number and the amplitudes of the CWT extrema decrease by a factor of two from every dyadic scale to the next coarser scale, we put $\sigma_2(j-1) = 4\sigma_2(j)$. The standard deviation of the first noise component, σ_1, was, by trial and error, determined to be 10 at the coarsest scale. We also put $\sigma_1(j-1) = 4\sigma_1(j)$. No significant changes in the results of the Monte-Carlo simulations discussed below occur if σ_1 is initialized at 1 or 100 instead of 10.

Handling association hypotheses is particularly simple in this case. The number of targets $N_{TT}(j) = 1$ is known and constant, and the target is always detected: $N_{TT} = N_{DT} = 1$. This means that there are no new targets, and the number of false alarms is always one less than the number of measurements. So, all association hypotheses are equiprobable given $\Omega_g(j+1)$, and we can rewrite (3.7) using (3.8):

$$Pr(\Omega_l(j)|Y_{cum}(j)) = \frac{1}{c_1}\mathcal{N}(\mathbf{y}(j) - \hat{\mathbf{x}}(j|j+1); 0, P(j|j+1)+R) Pr(\Omega_{l_1}(j+1)|Y_{cum}(j+1)) \tag{4.5}$$

We are now ready to run the filter and present the results.

4.3 Experimental Results.

First, we present the results of Monte-Carlo simulations conducted for a linear spline with additive Cauchy-contaminated Gaussian noise process. In other words, the probability density of every noise sample x is:

$$p(x) = (1-\varepsilon)\mathcal{N}(0,v) + \varepsilon\mathcal{C}(0,b),$$

where $\mathcal{N}(0,v)$ is the Gaussian probability density with mean 0 and variance v, and $\mathcal{C}(0,b)$ is the Cauchy probability density with parameters 0 and b:

$$\mathcal{C}(0,b) = \frac{1}{\pi}\frac{b}{b^2+x^2}$$

In order to get the kind of spiky noise that we claim our algorithm is robust to, we make the parameter ε small and b large, thus getting a Gaussian noise with very rare but large spikes. In particular, we took $v = 25$ and $b = 500$ in our simulations. A typical sample path (i.e., a linear spline with the knot at 128 + a noise realization) is shown in the Figure 4.5, for $\varepsilon = 0.02$.

If ε is close to zero, the noise is essentially Gaussian, and therefore a GLR-type Gaussian [14] estimator should perform better than or the same as our algorithm. However, as ε increases, the decline in the performance of the Gaussian estimator is more dramatic than that of the estimator described in this thesis. This is confirmed by the Figure 4.6 which depicts the mean-squared errors in estimating the knot location of a linear spline with one knot, for different values of ε. The error bars correspond to 95% confidence regions (i.e., two standard deviations.)

Next, we demonstrate the robustness of our algorithm to the detailed structure of the discontinuity by applying it to a signal which is not a purely linear spline. Particularly, we use functions

$$f(t) = \frac{(t+a)^2 - a^2}{128}u(t),$$

where $u(t)$ is the unit step. In other words, the function is a constant zero to the left of the knot and a parabola to the right (but the derivative of the parabola at the knot is nonzero.) One such function, for $a = 80$, is shown in the Figure 4.7.

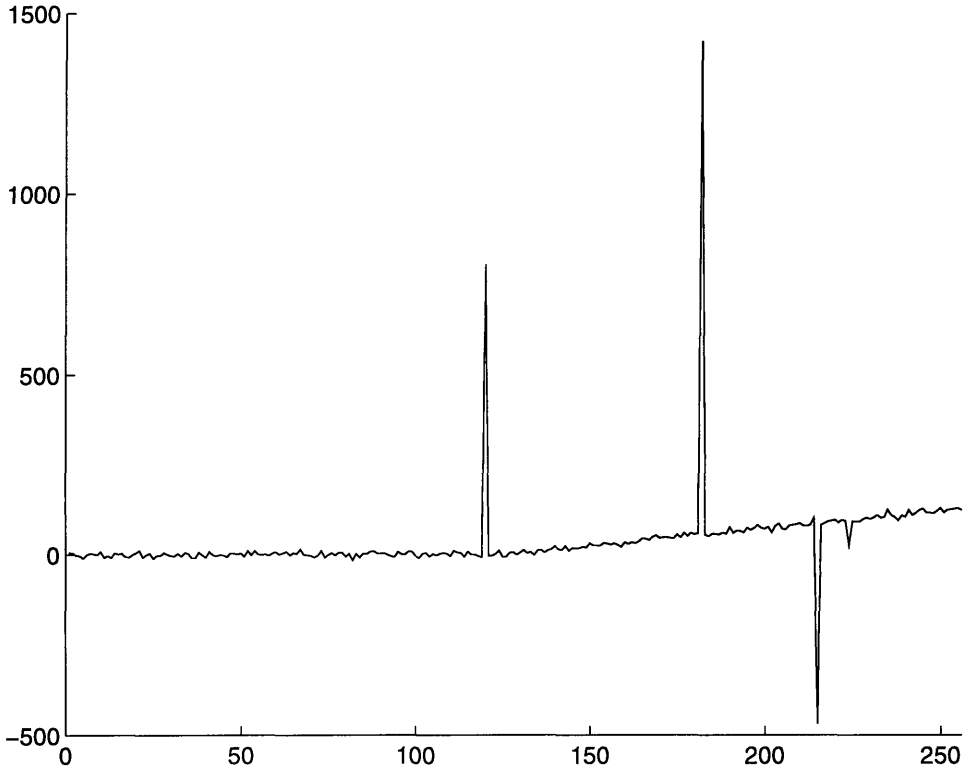

Figure 4.5: A typical sample path of the noise whose density is $p(x) = 0.98\mathcal{N}(0, 25) + 0.02\mathcal{C}(0, 500)$, added to a linear spline.

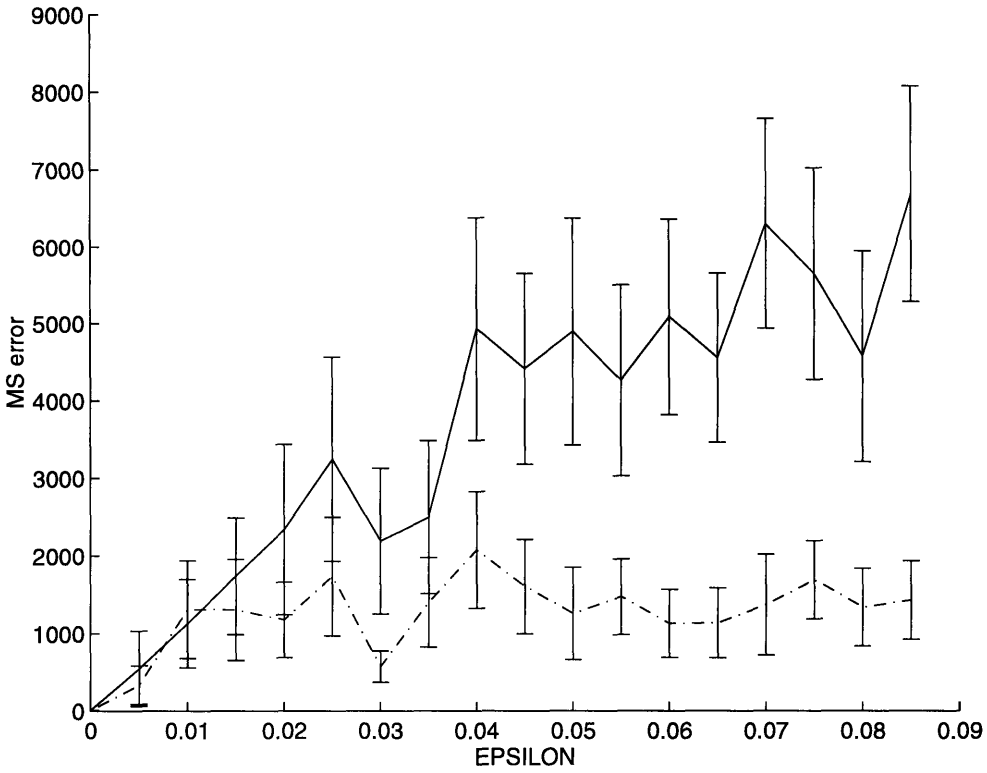

Figure 4.6: The MS error for the estimates of the knot location of a linear spline, versus ε. (The noise density is $p(x) = (1-\varepsilon)\mathcal{N}(0,25) + \varepsilon\mathcal{C}(0,500)$.) The dotted line corresponds to our algorithm; the solid line is GLR.

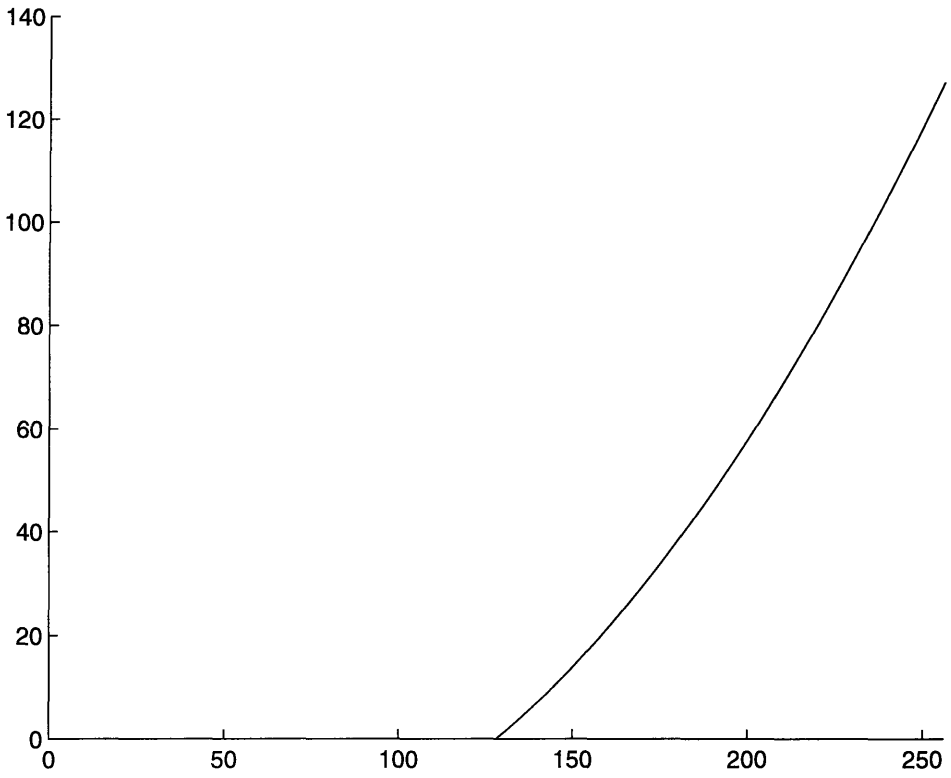

Figure 4.7: $f(t-128)$, where $f(t) = ((t+80)^2 - 80^2)u(t)/128$.

We again compare the performance of our algorithm with that of a GLR estimator tuned to linear splines, for different values of a. This time the noise is purely Gaussian. Again, our algorithm clearly outperforms the Gaussian estimator (Figure 4.8). As the curvature of the parabola decreases (and thus the parabola looks more and more like a straight line) – which corresponds to increasing a – the performance of the estimators becomes more and more similar. It is to be expected that, as the function approaches the linear spline, the performance of GLR becomes close to that of our estimator.

Thus, the algorithm described in this chapter is more robust than GLR when applied to linear splines with one knot.

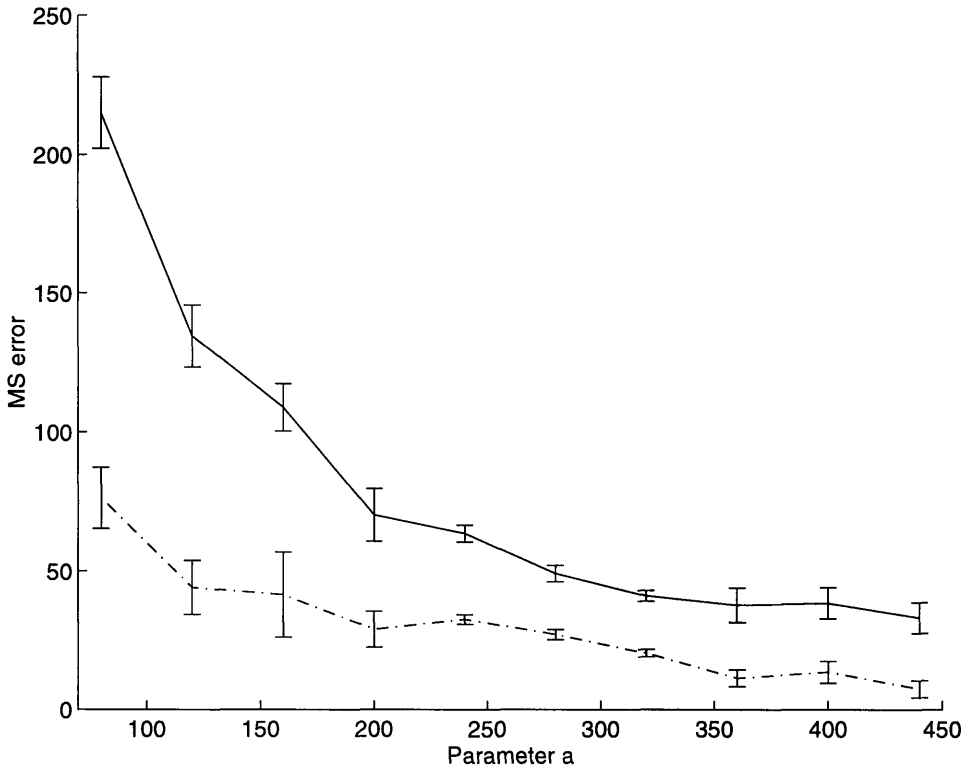

Figure 4.8: The MS error for the estimates of the knot location, versus a. The dotted line corresponds to our algorithm; the solid line is GLR.

4.4 Estimating the Spline Order.

In this section, we present a simple extension of the results described in the previous section. Given a function which is a spline of order less than or equal to M with one knot, we can set up an M-ary hypothesis testing problem to find out the spline order. In particular, the hypothesis H_i is that the spline order is i ($i = 1, ..., M$). We can compute the most likely fit under each hypothesis using the algorithm described in the preceding sections of this chapter. After this, we choose the most likely hypothesis. There are two easiest ways to handle every hypothesis. One can either use a wavelet with $i+1$ vanishing moments for H_i, or a wavelet with $M+1$ vanishing moments for all hypotheses. In the simulations below, we adopt the latter way.

This scheme does not necessarily work better than GLR-type detectors, even under the spiky-noise scenario. However, once the determination of the spline order has been made, our scheme will provide generally more accurate estimates of the knot than GLR.

We now illustrate the detection scheme with $M = 2$. Figures 4.9 and 4.10 show typical sample paths for first- and second-order splines with additive spiky noise described by

$$p(x) = (1 - \varepsilon)\mathcal{N}(0, 4) + \varepsilon\mathcal{C}(0, 50),$$

In both cases, the actual knot is at 128, and $\varepsilon = 0.01$. Figure 4.11 shows the probability of correct decision vs. ε. (Two hundred runs were made for each value of ε: 100 with the actual function being a linear spline, and 100 with the actual function being a quadratic spline.)

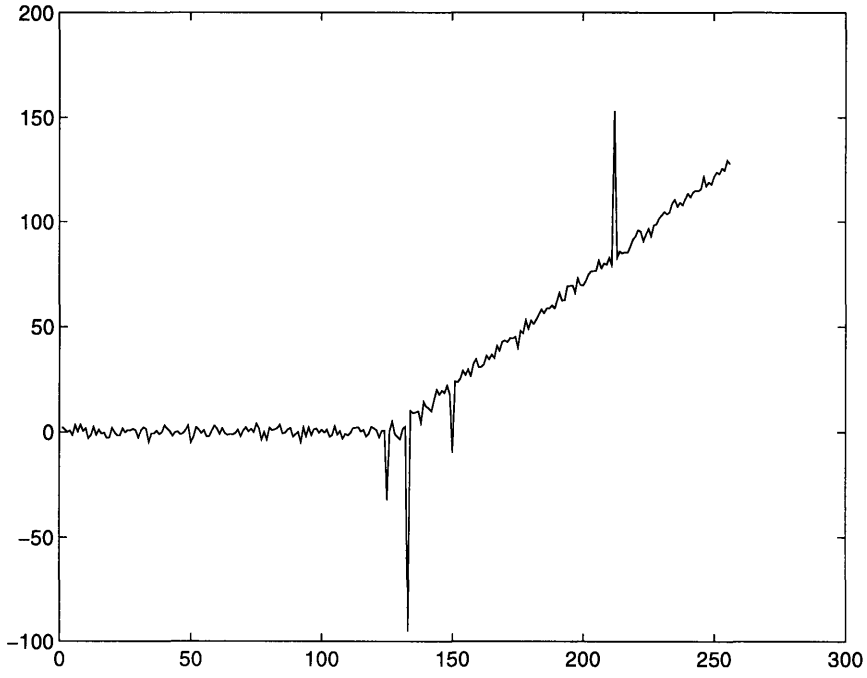

Figure 4.9: Linear spline with a spiky noise (the knot is at 128.)

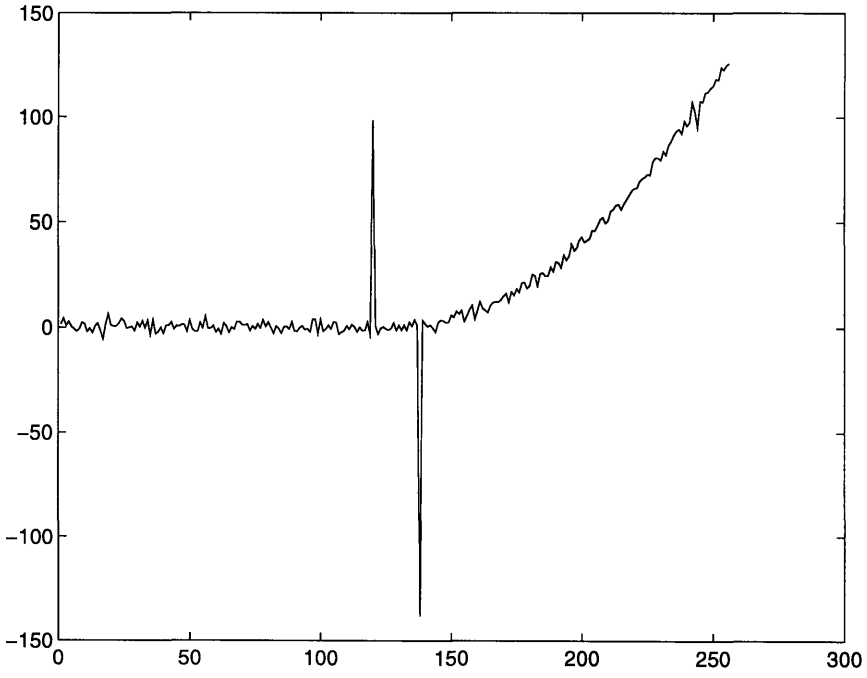

Figure 4.10: Quadratic spline with a spiky noise (the knot is at 128.)

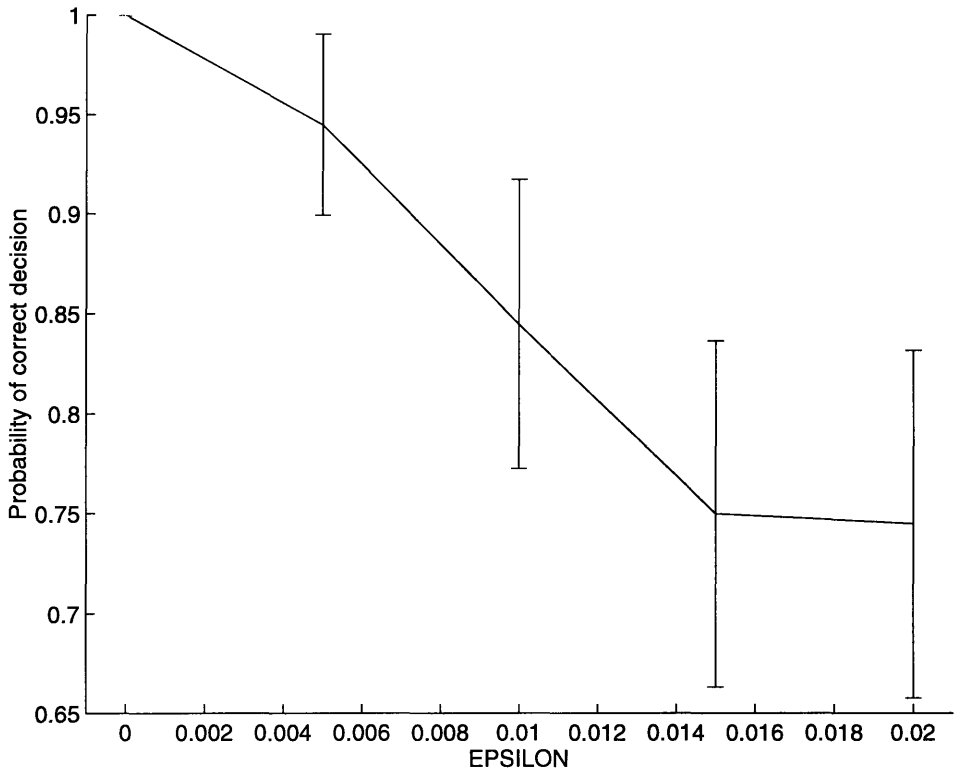

Figure 4.11: Probability of correct decision vs ε.

Chapter 5

CONCLUSION AND FUTURE WORK

In this thesis, a novel knot detection algorithm has been presented. The algorithm uses the extrema of the continuous wavelet transform to zoom in on the knots. It utilizes the theory on the relationship between the singularities of functions and corresponding CWT extrema tracks [3], [10], [11]. By viewing the scales of the CWT as successive "scans", or measurement sets, we cast the problem of tracking extrema across scale as a multi-target tracking problem. Each extrema track corresponds to a "target", i.e., a knot. By following these tracks from coarse to fine scales, we arrive at an estimate of the location of the knot and the slope change at the knot.

The resulting algorithm is robust both to the detailed statistical structure of the noise and to the precise nature of the abrupt change at the knot. It has been demonstrated that the algorithm is much more robust than the classical GLR-type algorithm.

There are several possibilities for future research. The most immediate concern is generalizing the algorithm to splines with several knots which are close to each other. If two knots are close to each other (i.e., closer than the support of the wavelet at some scale), then the CWT extrema tracks corresponding to these knots

are not going to be independent; they will interact. The next step in the research is to incorporate the knowledge of this interaction into the algorithm.

Another research direction is building more sophisticated models of the distributions of the CWT extrema due to noise. Even if the noise is Gaussian and white, finding the exact probability density of inter-arrival times of CWT extrema is very difficult, because the operation leading from the original signal to the CWT extrema is nonlinear.

Appendix A

A Fast Dyadic Wavelet Transform Algorithm.

In this appendix, we show how to discretize the wavelet transform for the class of wavelets described in Section 4.1. We use the notation of Chapter 2. Since we use 2^j to index the dyadic scales of wavelet transforms, we shall use i to denote $\sqrt{-1}$. We also use $\binom{k}{n}$ as a shorthand for $\frac{k!}{n!(k-n)!}$.

As in Section 2.3, we assume that the input is a discrete sequence $d[n] = S_\phi f(n, 1)$, and the goal is to be able to compute a sampling $S_\phi f(n + w, 2^j)$ of $S_\phi f(t, 2^j)$ and a sampling $W_\psi f(n+w, 2^j)$ of $W_\psi f(t, 2^j)$ for any dyadic scale 2^j, where $j \geq 1$. We show that this computation can be done recursively, by convolving $d[n]$ with dilated versions of certain discrete filters $H[n]$ and $G[n]$ an appropriate number of times. To do that, we just define $H[n]$ and $G[n]$ and describe the computations, without providing any intuition as to why the procedure works. A more extensive treatment of this algorithm can be found in [11].

To simplify notation, we define $S[n, j] = S_\phi f(n + w, 2^j)$ and $W[n, j] = W_\phi f(n + w, 2^j)$ for $j \geq 1$, where, as in Definition 2.3.2, w is a sampling shift which is determined by $\psi(t)$. For any discrete sequence $P[n]$ and any nonnegative integer j, we also define $P_j[m]$ by:

$$P_j[m] = \begin{cases} P[n] & \text{if } m = 2^j n \\ 0 & \text{if } \forall n \; m \neq 2^j n \end{cases}$$

Let us define a sequence $H[n]$ of length $2k+2$ as follows:

$$H[n] = \frac{\binom{2k+1}{k+1+n}}{2^{2k+1}}, \qquad n = -k-1, ..., k \tag{A.1}$$

Since $\binom{2k+1}{k+1+n} = \binom{2k+1}{(2k+1)-(k+1+n)} = \binom{2k+1}{(k+1)-(n+1)}$, we conclude that $H[n] = H[-n-1]$, i.e., the sequence is symmetric with respect to $n = -\frac{1}{2}$. It is also worth noting that

$$\sum_{n=-k-1}^{k} \binom{2k+1}{k+1+n} = 2^{2k+1},$$

which means that

$$\sum_{n=-\infty}^{\infty} H[n] = 1,$$

i.e., this filter is a $(2k+2)$-point weighted averager.

Its discrete-time Fourier Transform (DTFT) is

$$\hat{H}(\omega) = \frac{\sum_{n=-k-1}^{k} \binom{2k+1}{k+1+n}}{2^{2k+1}} e^{-in\omega},$$

which, by substituting $-n = k+1-m$ can be written as

$$\begin{aligned}
\hat{H}(\omega) &= \frac{1}{2^{2k+1}} \sum_{m=0}^{2k+1} \binom{2k+1}{m} e^{(k+1-m)i\omega} = \\
&= \frac{e^{\frac{i\omega}{2}}}{2^{2k+1}} \sum_{m=0}^{2k+1} \binom{2k+1}{m} (e^{\frac{i\omega}{2}})^{2k+1-m} (e^{-\frac{i\omega}{2}})^{m} = \\
&= e^{\frac{i\omega}{2}} \left(\frac{e^{\frac{i\omega}{2}} + e^{-\frac{i\omega}{2}}}{2} \right)^{2k+1} \\
&= e^{\frac{i\omega}{2}} \left(\cos \frac{w}{2} \right)^{2k+1},
\end{aligned}$$

where the second to last transition was obtained by using Newton's binomial theorem.

Define

$$\hat{\phi}(\omega) = \left(\frac{\sin \frac{w}{2}}{\frac{w}{2}} \right)^{2k+1} \tag{A.2}$$

It is easy to see that

$$\hat{\phi}(2\omega) = e^{-\frac{i\omega}{2}} \hat{H}(\omega)\, \hat{\phi}(\omega) \tag{A.3}$$

Now consider a second-difference sequence $G[n] = \delta[n+1] - 2\delta[n] + \delta[n-1]$, where $\delta[m]$ is the Kronecker delta sequence which is one when $m = 0$ and zero otherwise. The DTFT of $G[n]$ is given by

$$\hat{G}(\omega) = e^{i\omega} - 2 + e^{-i\omega} = 2(cos\,\omega - 1) \tag{A.4}$$

Let us define a wavelet $\psi(t)$ such that

$$\hat{\psi}(\omega) = e^{-\frac{i\omega}{2}} \hat{G}(\frac{\omega}{2})\, \hat{\phi}(\frac{\omega}{2}) \tag{A.5}$$

Substituting for $\hat{G}(\frac{\omega}{2})$ from A.4 and $\hat{\psi}(\frac{\omega}{2})$ from A.5 into A.5 and simplifying, we get:

$$\hat{\psi}(\omega) = e^{-\frac{i\omega}{2}} \frac{(i\omega)^2}{4} \left(\frac{\sin \frac{w}{4}}{\frac{w}{4}} \right)^{2k+3}, \tag{A.6}$$

We observe that the inverse Fourier transform of $\frac{\sin \frac{w}{4}}{\frac{w}{4}}$ is an even box function with amplitude 2 and unit area. We also know that multiplying by $i\omega$ in frequency domain is equivalent to differentiating the corresponding time function. Therefore, $\psi(t)$ is the second derivative of the convolution of $2k+3$ identical box functions – i.e., it is precisely the wavelet described in Section 4.1.

Using the definitions $S_\phi f(t, 2^j) = f * \phi_{2^j}(t)$ and $W_\psi f(t, 2^j) = f * \psi_{2^j}(t)$ of Sections 2.3 and 2.1, respectively, as well as the relations (A.3) and (A.5), we obtain the following recursive relations:

$$\hat{W}_\psi f(\omega, 2^{j+1}) = e^{-2^{j-1}i\omega} \hat{G}(2^j\omega)\, \hat{S}_\phi f(\omega, 2^j) \tag{A.7}$$

$$\hat{S}_\phi f(\omega, 2^{j+1}) = e^{-2^{j-1}i\omega} \hat{H}(2^j\omega)\, \hat{S}_\phi f(\omega, 2^j) \tag{A.8}$$

These equations hold for any nonnegative integer j. We can convert these equations to discrete-time domain by taking their inverse DTFTs. For any $j \geq 1$, we have:

$$S[n + 2^{j-1}, j+1] = S[n, j] * H_j[n]$$

$$W[n + 2^{j-1}, j+1] = S[n, j] * G_j[n]$$

In other words, we can get $S[n, j+1]$ and $W[n, j+1]$ by convolving $S[n, j]$ with an appropriately dilated version of $H[n]$ and $G[n]$, respectively, and shifting the result by 2^{j-1} samples to the right. However, when $j = 0$, these simple relationships do not hold, because 2^{0-1} is non-integer. In order to keep our computations simple, we shall disregard this half-integer shift and define

$$S[n, 1] = d[n] * H_j[n]$$
$$W[n, 1] = d[n] * G_j[n]$$

Thus, $S[n, j]$ and $W[n, j]$ are the samplings of $S_\phi(t, 2^j)$ and $W_\psi(t, 2^j)$ at half-integer points, for any $j \geq 1$. (I.e., the parameter w of the Definition 2.3.2 is $\frac{1}{2}$ in this case.) We arrive at the following algorithm (taken from Appendix B of [11]) to compute $\{W[n, j]\}_{j=1}^{J}$ and $S[n, j]$ from $d[n]$:

$W[n, 1] = \lambda_0 d[n] * G[n]$

$S[n, 1] = d[n] * H[n]$

for $(j = 1;\ j \leq J - 1;\ j++)$ {

$W[n, j+1] = \lambda_j S[n - 2^{j-1}, j] * G_j[n - 2^{j-1}]$

$S[n, j+1] = \lambda_j S[n - 2^{j-1}, j] * H_j[n - 2^{j-1}]$

},

where λ_j's are chosen so as to cancel discretization effects upon the CWT extrema. (In other words, the extremum of a continuous function which is sampled is not guaranteed to be precisely at a discretization point. Therefore, the extrema of the discretization of CWT slightly differ from the extrema of CWT. The numbers λ_j are chosen so as to eliminate this discrepancy.)

Appendix B

Proof of Theorem 2.2.7

Without loss of generality, assume that $t_0 = 0$. Let us define μ and I by: $\mu = \frac{t-\tau}{s}$, and $I = \int_{-\infty}^{\infty} f(0)\frac{1}{s}\psi(\frac{t-\tau}{s})\,d\tau$. Observe that $I = f(0)\int_{-\infty}^{\infty}\psi(\mu)\,d\mu \equiv 0$, and therefore

$$
\begin{aligned}
|W_\psi f(t,s)| = |W_\psi f(t,s) - I| \leq \\
&\leq \int_{-\infty}^{\infty} |f(\tau) - f(0)| \frac{1}{s}\left|\psi(\frac{t-\tau}{s})\right| d\tau \\
&\leq \int_{-\infty}^{\infty} C|\tau|^\alpha \frac{1}{s}\left|\psi(\frac{t-\tau}{s})\right| d\tau \\
&= C\int_{-\infty}^{\infty} |\psi(\mu)| \cdot |t-\mu s|^\alpha \, d\mu \\
&\leq C\int_{-\infty}^{\infty} |\psi(\mu)| \cdot |t|^\alpha \, d\mu + C\int_{-\infty}^{\infty} |\psi(\mu)| \cdot |\mu|^\alpha |s|^\alpha \, d\mu \\
&= C_1|t|^\alpha + C_2|s|^\alpha \\
&\leq A(|s|^\alpha + |t|^\alpha) \quad \square
\end{aligned}
$$

Bibliography

[1] Michele Basseville and Albert Benveniste, editors. *Detection of Abrupt Changes in Signals and Dynamical Systems.* Springer-Verlag, 1986.

[2] Lori-Ann Belcastro. Tomographic reconstruction of polygons from knot locations and chord length measurements. Master's thesis, Massachusetts Institute of Technology, 1993.

[3] Ingrid Daubechies. *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, 1992.

[4] Carl de Boor. *A Practical Guide to Splines.* Springer-Verlag, 1978.

[5] T.N.E. Greville, editor. *Theory and Applications of Spline Functions.* Academic Press, 1969.

[6] M. Holschneider and P. Tchamitchian. Regularite locale de fonction non-differentiable de Riemann. In P.G. Lemarie, editor, *Les ondelettes en 1989.* Springer-Verlag, 1990.

[7] Peter J. Huber. *Robust Statistical Procedures.* Society for Industrial and Applied Mathematics, 1977.

[8] R. E. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82-D, 1960.

[9] Thomas Kurien. Issues in the design of practical multi-target tracking algorithms. In Y. Bar-Shalom, editor, *Multitarget-Multisensor Tracking: Applications and Advances*, chapter 3. Artech House, 1992.

[10] Stephane Mallat and Wen Liang Hwang. Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38(2), 1992.

[11] Stephane Mallat and Sifen Zhong. Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7), 1992.

[12] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of Royal Society, London*, 207, 1980.

[13] Yves Meyer. Un contre-exemple a la conjecture de Marr et a celle de S. Mallat. Preprint, 1991.

[14] Alberto M. Mier Muth and Alan S. Willsky. A sequential method for spline approximation with variable knots. *International Journal of Systems Science*, 9(9), 1978.

[15] Igor V. Nikiforov and Ivan N. Tikhonov. Application of change detection theory to seismic signal processing. In Michele Basseville and Albert Benveniste, editors, *Detection of Abrupt Changes in Signals and Dynamical Systems*, chapter 12. Springer-Verlag, 1986.

[16] Donald B. Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, AC-24(6), 1979.

[17] Walter Rudin. *Functional Analysis.* McGraw-Hill, Inc, 1991.

[18] Larry L. Schumaker. *Spline Functions: Basic Theory.* John Wiley & Sons, 1981.

[19] Alan S. Willsky. *Recursive Estimation. 6.433 Course Notes.* MIT, 1992.

[20] Alan S. Willsky. *Stochastic Processes, Detection, and Estimation. 6.432 Course Notes.* MIT, 1992.

[21] Alan S. Willsky and H. L. Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control*, 21, 1976.

[22] D. Youla and H. Webb. Image restoration by the method of convex projections. *IEEE Transactions on Medical Imaging*, MI-1, 1982.