A scaling medium representation, a discussion on well-logs, fractals and waves

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus,
Prof. dr. ir. J. Blaauwendraad
in het openbaar te verdedigen
ten overstaan van een commissie,
door het College van Dekanen aangewezen,
op dinsdag 21 januari 1997 te 13:30 uur

door

Felix Johan HERRMANN

natuurkundig ingenieur

geboren te 's-Gravenhage

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. ir. A. J. Berkhout

Toegevoegd promotor:
Dr. ir. C. P. A. Wapenaar

Promotiecommissie:
Rector Magnificus, voorzitter
Prof. dr. ir. A. J. Berkhout (*Technische Universiteit Delft, Technische Natuurkunde*), promotor
Dr. ir. C. P. A. Wapenaar (*Technische Universiteit Delft, Technische Natuurkunde*), toegevoegd promotor
Prof. dr. ir. G. E. W. Bauer (*Technische Universiteit Delft, Technische Natuurkunde*)
Prof. dr. J. J. Duistermaat (*Rijks Universiteit Utrecht, Mathematisch Instituut*)
Prof. dr. ir. J. T. Fokkema (*Technische Universiteit Delft, Technische Aardwetenschappen*)
Dr. A. le Méhauté (*Institute Supérieur des Matériaux du Mans, ISMANS, Frankrijk*)
Dr. D. J. Smit (*Shell Research BV, R.T.S., Rijswijk*)

ISBN 90-9010132-2

SUPPORT
The research for this thesis has been financially supported by the DELPHI consortium.

Typesetting system: LaTeX 2ε
Printed in The Netherlands by: Beeld en Grafisch Centrum, Technische Universiteit Delft.

"Tooit men zich met een nieuwe manier van uitdrukken, dan werpt men met het ander gewaad de oude problemen van zich af."

L. Wittgenstein, 1946.

Contents

to convert the map of reflections to the location of the singularities is called *migration*

Chapter 1

Introduction

As a consequence of the ever increasing demand for hydrocarbons and the steady decline of newly explored resources an uprise has taken place towards a more production oriented seismic technology. Traditionally the field of seismic exploration committed itself primarily to *locating* resources in the earth's subsurface. Of late, attempts have been made to exploit seismic technology to deliver qualitative and quantitative information about the reservoir of interest. This new direction is far more ambitious since it strives for petrophysical and lithological information. This requires an integration of different types of data, acquired within different disciplines. It also calls for a good understanding on how to interrelate measurements taken at different scales. This has been the subject of this thesis, a study aimed to improve the understanding on the interaction between seismic waves and the medium, with variations at different scales. In figure 1.1, a schematic overview is depicted to illustrate how seismic waves are used to insonify the earth's subsurface in order to obtain information. The basic idea of this method is that waves, excited by a source, tend to *reflect* at regions where the earth's material properties show rapid variations, i.e. where singularities in the medium properties occur. In this way the seismic wavefield inherits information not only on the *locations* but also on the *nature* of the dominant singularities, possibly containing more, hopefully discriminating, information. This information can be picked up by detecting the wavefield as a function of the horizontal or vertical coordinate and as a function of time. Measurements taken along the surface are referred to as surface seismic data and are collected in so-called seismic shot records. An example of such a shot record is given in figure 1.2. The time traces, measured by the geophones or hydrophones, are plotted vertically as a function of the horizontal source-receiver offset. Despite the aforementioned trend, todays exploration seismic processing is still mostly involved in estimating the locations of the dominant singularities (Bleistein, 1984). These singularities are generally believed to be jump discontinuities, a special *class* of singularities. This is related to the idea that the earth is constructed of distinct layers, like an onion. The general procedure to convert the map of reflections to the location of the singularities is called *migration*

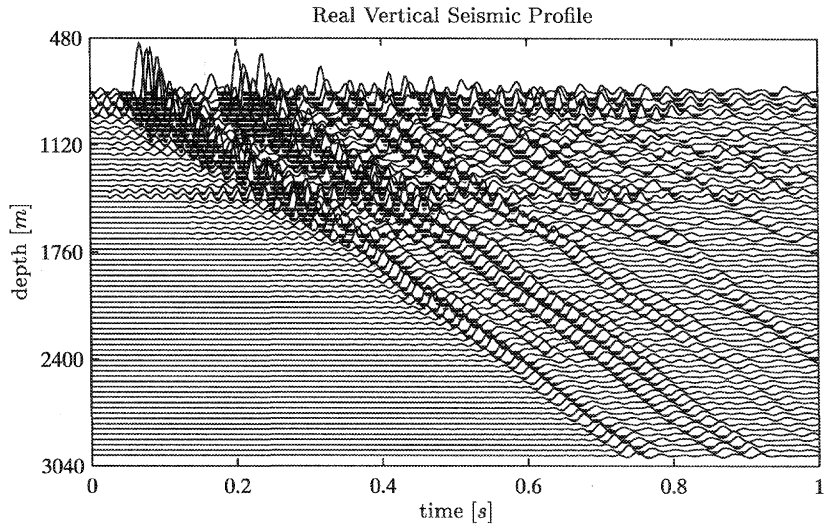

Figure 1.3 *A real Vertical Seismic Profile (VSP). The particle velocity is depicted as a function of the time, horizontally, and vertical offset, depth, vertically. Remark that as the wave travels down its "impulsive" nature gradually diminishes.*

dominant wavelength in the range of the medium fluctuations (with their observed scale divergence) one might be in a better position to close the gap between seismic and well-log based measurements. The latter is a problem of different scales and has to do with the question on how to coarse-grain the fine-grained well-data to the coarse-grained scale range inhabited by the probing seismic wavefield.

A first step to face the above challenge is to have a thorough look at the medium's complexity, followed by the introduction of a proper representation of the observed complexity making explicit reference to the scale. It has become clear that in this thesis decomposing the heterogeneity along the scale direction is a means to unravel the complexity. It has been observed that well-log data drastically change as a function of the scale due to the variability on every scale. This can be analyzed using the multiscale analysis tools described. To obtain an order of magnitude estimate for the local and global scaling, Hölder exponents are being calculated with the help of the continuous wavelet transform (Daubechies, 1988; Mallat and Hwang, 1992; Bacry et al., 1993). These Hölder exponents can be used to asses the degree of differentiability of the individual singularities, which is directly related to scale divergence. In that way the singularity spectrum provides information on the global degree of differentiability and integrability, on the generalized or Hausdorff dimensions and on the scaling behaviour of the statistical moments (Mandelbrot, 1974; Parisi and Frisch, 1985; Halsey et al., 1986).

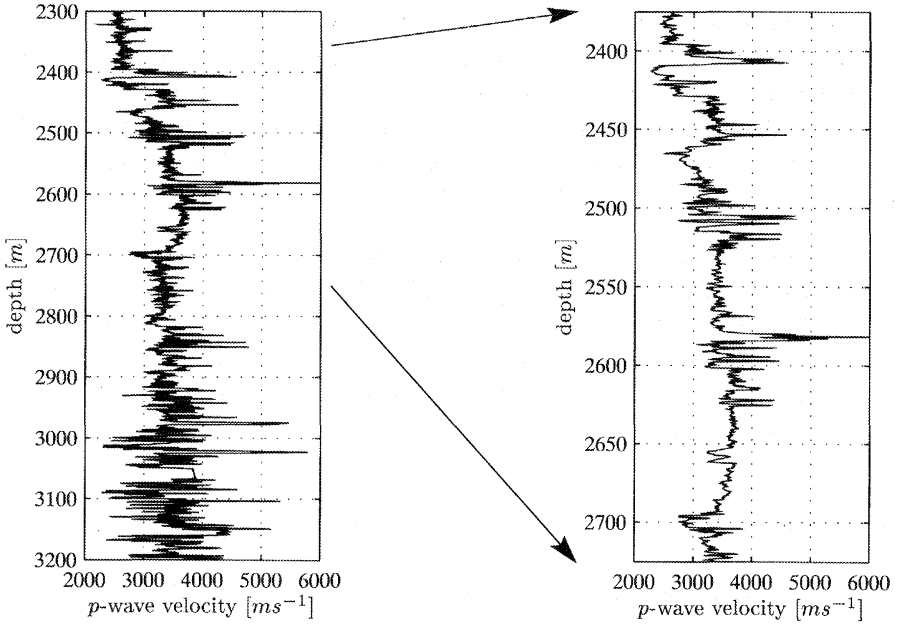

Figure 1.4 *An example of a well-log dataset. On the left the compressional wavespeed (P-velocity) and on the right a slightly zoomed in version. Remark the overwhelming complexity displayed by this well-log measurement.*

Fractals constitute a class of mathematical objects containing hierarchies of the singularities. It was Mandelbrot (1982) who first coined them as models capturing complexity across a wide scale range. They are in use to describe many different phenomena: share prices in the stock exchange, the irregularity of the coast of Britain, the intermittency and scale-invariance of the hydro-dynamic turbulence and the irregularity of geomagnetic noise. This wide range of phenomena is believed to include well-log measurements, as shown in the literature (Walden and Hosken, 1985; Todoeschuck and Jensen, 1989; Saucier and Muller, 1993) which will be confirmed by this thesis. The observation that well-log data display a predominant scaling behaviour substantiates the introduction of a representation with explicit reference to the notion of scale. Mathematically speaking this implies that functionals rather than ordinary functions will be used to represent the scaling well-log measurements.

The motivation to invoke the scaling medium representation was to accomplish a better understanding of the evidenced dynamics: *dispersion* and *specular reflection*. Dispersion (frequency dependent attenuation and time delay, see e.g. figure 1.3) is attributed to

an intricate interference mechanism induced by the medium's heterogeneity. Viscous relaxation effects have been assumed to be absent. From figure 1.3 it is clear that this is a cumulative effect leading to a loss of the initial pulse's sharpness on its way down. Specular reflections, on the other hand, are the result of an averaging process of a more local nature. This averaging process maps the singularities, occuring in the medium properties, to the space time characteristics of the wavefield, an observation substantiated by the irregular nature of the VSP's coda and the reflection seismic shot record.

The current model, utilized to describe the mapping of the singularities in the medium to the wavefield (Claerbout, 1971; Bleistein, 1984; Yilmaz, 1987; Wapenaar, 1996b; Berkhout, 1982), is based on the acoustic wave equation in which the medium properties occur as coefficients. In this formulation no explicit reference is made to the scale. Despite this possible deficiency, the model is being used there where the medium displays a strong dependence on the scale.

1.2 Initial approach

The initial goal of this PhD research project was to better understand the induced dispersion effects, attributed to the medium's irregular heterogeneity. To simplify the intricate interference mechanism perceived to be responsible for the characteristics of the probing pulse, an approximation was proposed. This approximation, known as the O'Doherty-Anstey formula (O'Doherty and Anstey, 1971), permits a formulation of the propagation operator, carrying the wavefield from one depth level to the other, in terms of the sample power spectrum of the depth traveltime converted reflection coefficients (Banik et al., 1985a,b; Resnick et al., 1986; Burridge et al., 1988; Burridge and Chang, 1989; Burridge et al., 1993; de Hoop et al., 1991b,a; Shapiro and Zien, 1993; Shapiro et al., 1994). The initial idea was to replace the sample power spectrum by its stochastic expectation, yielded by an appropriate random process. The random process initially used in this context resulted in an exponential decaying correlation function, a choice difficult to reconcile with the evidenced long tailed correlations (Walden and Hosken, 1985; Todoeschuck and Jensen, 1989) displayed by well data. For that reason a fractal representation was proposed, yielding an elegant parametric representation in terms of a single scaling exponent (Herrmann, 1991; Herrmann and Wapenaar, 1992, 1993, 1994; Burridge et al., 1993). Eventually this line of research amounted to a proposal for a migration scheme taking the effects of fine layering into account (Wapenaar and Herrmann, 1996). Later on, the author came to the insight that the underlying assumptions behind the derivation of the weak fluctuation O'Doherty-Anstey approximation were difficult to reconcile with the empirical multiscale findings. A similar observation applies to the recent elaborate results on media containing strong order one fluctuations (Asch et al., 1991, 1990; Papanicolaou et al., 1990; Burridge and Chang, 1989; Burridge et al., 1992; Lawecki et al., 1994;

Lawecki and Papanicolaou, 1994). In that case namely, a separation of scales is invoked where the correlation length of the medium is considered to be small compared to the spatial width of the pulse while the pulse is assumed to travel long distances. Again this presupposition does not combine well with the findings of the multiscale analysis where a separation of scales seems futile.

1.3 Current state of affairs

The presupposition of a separation of scales, invoking the necessary stationarity for the mean of the fluctuations, is also shared by localization theory (Anderson, 1958; Souillard, 1986; Pastur and Figotin, 1991). In this theory one is currently able to exclude the singular continuous part of the spectrum, under the above assumption, and to prove that the spectrum for the eigenvalues is pure point with the eigenvalues lying dense (Pastur and Figotin, 1991; Pastur, 1994). Given this observation one is able to come up with qualitative estimates for the exponential decay of the eigenfunctions in terms of Lyapunov exponents. Unfortunately the above presuppositions do not combine well with the multiscale behaviour empirically found which seems to preclude a separation of scales. Hence the question on how one-dimensional acoustic waves travel in one-dimensional media with powerlaw tailed correlations remains open. However, that does, on the one hand, not rule out the possibility to check for a power law type of behaviour displayed by the Lyapunov spectrum or to initiate a discussion on the potential role of scale – the scale derivative operator does not commute with the spatial derivative operator – within the context of describing wave interactions in scaling media.

1.4 An alternative approach

With the benefit of the hindsight, I have the impression that the problem at hand is perhaps too difficult. This notion became especially clear in relation to the very involved localization theory. I was unable to grasp its full details and the implications for wave propagation in media displaying long-tailed correlations. This may explain why partly this thesis is a conglomeration of yet unfinished ideas. I believe that the problem requires an explorative approach. In such a way the understanding of the neglected notion of scale might be improved. Speculations on a more general formulation of the problem, including the notion of scale, are presented in the *Epilogus*, which anticipates on future research and should not be regarded as an official part of this thesis.

1.5 Outline of this thesis

The contents of this thesis has been divided into three parts: The *Capita Prima* constituting the body of this thesis; the *Capita Selecta* consisting of selected topics being

of relevance for the main part and finally the *Epilogus* in which I roughly sketch how I intend to pursue research in the future. To be more specific the *Capita Prima* consist of three chapters. The first one addresses the issue of finding a proper representation for the well-log data. It also gives an overview of fractal models which can be used to categorize and simulate the well-log's complexity. Within the multiscale analysis the continuous wavelet transform plays a central role. This transform is based on translations and dilatations effectuated on a proper analyzing wavelet. In the second chapter the operators generating the translations and dilatations are being reviewed. In the final chapter attention is paid to the scattering and spectral problem associated with the acoustic wave equation. It appears that localization theory is possibly required to explain the dispersion observed.

In the second part of this thesis, the *Capita Selecta*, several concepts are being reviewed that are essential in the discussions in the main part and the *Epilogus*. They can be skipped by the reader familiar with the concepts of distribution theory, chapter 5; multi- and monofractals, chapters 6 and 7 and multiscale analysis by the continuous wavelet transform, chapter 8.

Part I

Capita Prima

Chapter 2

A scaling medium representation

In this chapter a multiscale analysis, characterization and modelling mechanism is laid down which is aimed at capturing the complexity as displayed by well-log measurements. The purpose of this chapter is fourfold. First it is intended to devise an analysis method that reveals the scaling behaviour displayed by many geophysical phenomena amongst which the well-log data. Secondly it intends to come up with a characterization facilitating the assessment of notions such as the degree of differentiability. Thirdly I will come up with a number of fractal models that yield a complexity that is comparable to the one evidenced from well data for example. Finally I propose a scaling medium representation in terms of the continuous wavelet transform.
key words: *scaling, continuous wavelet transform, singularity, Hölder exponent, (universal) multifractal, singularity spectrum, well-log data.*
prerequisites: *the chapters of the Capita Selecta.*

2.1 Introduction

One of the most striking aspects a (exploration) geophysicist[1] finds himself or herself faced with is the sheer complexity displayed by geophysical phenomena of various kind. It was Richardson (1922) who paraphrased the above observation so cunningly by parodying J. J. Swift's poem into

> Big whorls have little whorls that feed on their velocity
> and little whorls have smaller whorls and so on to viscosity
> – in the molecular sense (source (Schertzer and Lovejoy, 1993))

[1]The theory in this thesis can be of relevance also for the general physicist when one comes to think of the fields of turbulence or other fields where one tries to capture the spatial complexity and to gain a better comprehension of wave phenomena interacting with this complexity. Think for example of wave localization phenomena in aerogels, electron localization in solids or electromagnetic scintillations in the atmosphere.

In this way he hands over, in words, the basic ideas behind cascade models that have proven to be successful in tackling the complexity as being displayed by many geophysical processes. These cascade models share with these geophysical phenomena

- a certain *scale-invariance*: This scale-invariance is characteristic of many geophysical phenomena typically containing structures with sizes that span several orders of magnitude in length and/or time scale and that bear a certain similarity.

- an underlying *dynamical non-linear mechanism*: This type of mechanism is typically multiplicative yielding highly *intermittent* processes. This intermittency refers to another distinguishing characteristic of geophysical data in the sense that events tend to occur in a catastrophic fashion. That is to say that relatively quiet regions are superseded by wild and violent events occurring in a burstly fashion.

I think it is fair to say that the observation concerning the complexity may have profound implications on the way in which exploration geophysicists conduct their business. Indeed it is and was Mandelbrot (1974), followed by many others amongst whom le Méhauté (1991) and Nottale (1992), who advocated that the observed complexity does not strive very well with the assumptions, on the differentiability for instance, that are, sometimes lucidly, presupposed while setting up physical models. Within the context of this thesis this aspect is also important because the field of exploration geophysics finds itself confronted with a problem where the complexity seems to be omnipresent. The problem addressed in this thesis commits itself to improve the comprehension on how to interrelate measurements taken at different scales namely the fine-grained well-log data scale range and the coarser grained surface seismic or Vertical Seismic Profile data scale range. During the integration of these different data types one is generally interested in the dynamics of the wave interactions and this requires a true understanding on how waves interact in complex environments where the complexity persists over many scale ranges. Despite many efforts the gap between the seismic scale range and the well-log scale range has, to the author's knowledge, not yet been closed in a satisfactory way. It is this observation from which this thesis derives its inspiration.

At this point I ask myself the question whether one can use the observations on the complexity to one's advantage? In my case this comes down to initiating a discussion on the properties of wave phenomena taking place in media displaying a highly irregular behaviour. To be more specific wave interactions in media whose constitutive parameters[2] vary along the lines of the highly erratic, chaotic type of processes are of concern. Trying to come to terms with this challenge, I feel obliged to come up with a mathematical as well as physical proper and comprehensive representation for these constitutive coefficients

[2]These constitutive parameters refer in the case of acoustic wave motion to the density of mass and compressibility, both occurring as coefficients in the system of partial differential equations denoting the wave equation, see chapter 4.

that honors the notion of *scale* as founding principle of the evidenced complexity. This representation will serve as a starting point for a discussion on wave motion in relation to the notion of scale. The merit of the multiscale representation potentially lies in gaining a better understanding of the scaling structure in relation to wave interactions since it comes up with a tool:

- to analyze the *local* scaling being captured by *local order of magnitude estimates.* These estimates are related to the local degree of differentiability expressed in terms of Hölder exponents (Holschneider and Tchamitchian, 1990; Jaffard, 1991; Daubechies, 1992; Mallat and Hwang, 1992).
- to analyze the *global* scaling being captured by *global order of magnitude estimates* providing information on the *global regularity*, the hierarchy of the scaling, the generalized fractal dimensions and the scaling behaviour of the statistical moments (Mandelbrot, 1974; Hentschel and Procaccia, 1983; Collet, 1986; Schertzer and Lovejoy, 1987a; Siebesma, 1989; le Méhauté, 1991; Lichtenberg and Lieberman, 1992; Ott, 1993; Muzy et al., 1993; Bacry et al., 1993; Davis et al., 1994; Holschneider, 1995).

The above subdivision is inspired on the idea that the reflection of waves is governed by a local averaging mechanism, over the *gauge* of the probing pulse, while a global self-averaging mechanism, over the propagation distance, can conjectured to be held responsible for the behaviour of the pulse shape of the propagating wavefield.

I will commence the discussion by trying to convey the basic ideas that come with taking the notion of scale into consideration. This boils down to a rather intuitive introduction of the continuous wavelet transform. This wavelet transform forms not only the basis for the multiscale analysis methods that will be briefly reviewed in this chapter, the reader is referred to chapter 8 for a more detailed discussion, but also for the multiscale representation I intend to propose. The multiscale analysis reflects the above local/global subdivision and provides the necessary estimates for the characterization of the complexity in terms of mathematical quantities such as Hölder exponents or Hausdorff dimensions. This method is set to work on real well-log measurements. After that I will proceed the discussion by reviewing models that can be held responsible for generating the observed complexity and that will yield a way to simulate sequences with a type of scaling behaviour that matches certain aspects of the empirical observations such as well-logs. Then I will present my views on the mathematical and physical aspects that come with the presented multiscale analysis. The observations I make during that discussion eventually prompt me to propose a scaling medium representation that is consistent with the data and provides a robust framework in which I can give these irregular data sets a proper meaning. I will conclude this chapter with a short resume on the models currently used to characterize the complexity of the well data within the perspective of describing wave phenomena.

2.2 Having a look in the scale direction

The variability displayed by geophysical phenomena tends to persists over a large range of scales. This observation suggests to undertake an investigation of these data along different scales. Such an approach is quite natural given the scope and the objectives of this thesis where I pursue a better integration of seismic and borehole oriented data, a problem of different scales. Now how can one come up with an approach that is well equipped to handle/unravel the displayed scaling behaviour and that takes into consideration the notion that every measurement entails a coarse-graining/dressing of the actual/bare[3] physical quantity of interest? To answer this question let me first formally establish the notion that a measurement can be considered as a projection of the bare quantity onto a scale $\sigma = \sigma_0$, corresponding to the measurements device resolving capability. This map is effectuated via an inner[4] product of the actual bare quantity, f, and a real and symmetric smoothing kernel, $\phi_{\sigma_0,x}(x')$, located at $x = x'$,

$$f(\sigma_0, x) \triangleq \langle f, \phi_{\sigma_0,x} \rangle. \tag{2.1}$$

The width of the smoothing kernel is proportional to the scale index σ_0. The scaling, the dependence of the data on the scale indicator σ, can be examined by artificially coarse-graining the data even further via

$$f(\sigma\sigma_0, x) = \langle f, \phi_{\sigma\sigma_0,x} \rangle. \tag{2.2}$$

By setting $\sigma_0 = 1$ this expression simplifies to

$$f(\sigma, x) = \langle f, \phi_{\sigma,x} \rangle, \tag{2.3}$$

where σ is the scale indicator or index. The smoothing kernel in itself is obtained by the invocation of the combined action of the *shift* and *dilatation* operators on the initial smoothing kernel,

$$\phi_{\sigma,x}(x') = \mathcal{S}_{-x}\mathcal{D}_{\frac{1}{\sigma}}\phi(x') = \frac{1}{\sqrt{\sigma}}\phi(\frac{x'-x}{\sigma}), \tag{2.4}$$

while preserving the $L^2(\mathbb{R})$ norm, i.e. $\|\phi_{\sigma,x'}\|_2 = \|\phi\|_2$. The reader is referred to chapter 3 for a further substantiation of the shift operator $\mathcal{S}$ and the dilatation operator $\mathcal{D}$. Because the smoothing kernel is chosen to be real and symmetric equation (2.1) can be recasted in the form of a *spatial convolution*

$$f(\sigma, x) = (f * \phi_\sigma)(x) \quad \text{with } \phi_\sigma(x) = \frac{1}{\sqrt{\sigma}}\phi\left(\frac{x}{\sigma}\right). \tag{2.5}$$

[3]The terminology bare and dressed are borrowed from Renormalization Group theory where they refer to non-averaged and averaged quantities respectively (Wilson, 1983).

[4]Defined as $\langle f, g \rangle \triangleq \int f(x)g^*(x)\mathrm{d}x$, with * denoting the complex conjugate and f, g arbitrary functions.

Provided with the formalization of the coarse-graining in equation (2.3) one may wonder what happens when one considers the effect induced by an infinitesimal change in the scale indicator, $\sigma - \mathrm{d}\sigma \mapsto \sigma$. Following Holschneider (1995) one can derive for the difference $f(\sigma - \mathrm{d}\sigma, x) - f(\sigma, x)$, while $\mathrm{d}\sigma \to 0$, that

$$\mathcal{W}\{f, \psi\}(\sigma, x) \triangleq -\sigma \partial_\sigma f(\sigma, x) \tag{2.6}$$

in which $\mathcal{W}\{f, \psi\}$ represents the continuous wavelet transform defined by

$$\check{f}(\sigma, x) = \mathcal{W}\{f, \psi\}(\sigma, x) = \langle f, \psi_{\sigma,x} \rangle, \tag{2.7}$$

where $\psi_{\sigma,x}(x') \in L^2(\mathbb{R})$ denotes the analyzing wavelet given by

$$\psi_{\sigma,x}(x') = \frac{1}{\sqrt{\sigma}} \psi(\frac{x' - x}{\sigma}). \tag{2.8}$$

This mean zero wavelet, $\int \psi \mathrm{d}x' = 0$, is obtained by the action of the scale operator, $\mathcal{C}$, on the smoothing kernel

$$\psi(x) = j\mathcal{C}\phi(x), \tag{2.9}$$

with the scale operator defined by

$$\mathcal{C} = \frac{1}{2j}(x' \frac{\mathrm{d}}{\mathrm{d}x'} + \frac{\mathrm{d}}{\mathrm{d}x'} x'). \tag{2.10}$$

For further detail the reader is referred to section 3.6.3 of chapter 3.

At this point one might wonder what is gained by recasting equation (2.3) into equation (2.6)? First of all one has to take into consideration that $f(\sigma, x)$ contains all *details* in f ***up to*** the scale σ whereas $\check{f}(\sigma, x)$ delineates the details ***at*** scale σ. Moreover it also expresses the change in the scale direction in the sense that it measures the change effectuated by an infinitesimal change in the scale indicator σ. It is this latter property which is of primary importance. It allows one to examine the local scaling behaviour, the dependence on the scale indicator σ, by applying *zooms* with respect to the abscissa $x = x_0$,

$$r \mapsto \mathcal{W}\{f, \psi\}(r\sigma, rx_0) \tag{2.11}$$

for some fixed (σ, x_0). It also facilitates the examination of the behaviour for f at a certain scale, $\sigma = \sigma_0$, and as a function of x,

$$x \mapsto \mathcal{W}\{f, \psi\}(\sigma_0, x) \quad \text{for.} \tag{2.12}$$

This latter operation is called a *voice*[5].

[5] I borrowed the terminology *zooms* and *voices* from Holschneider (1995).

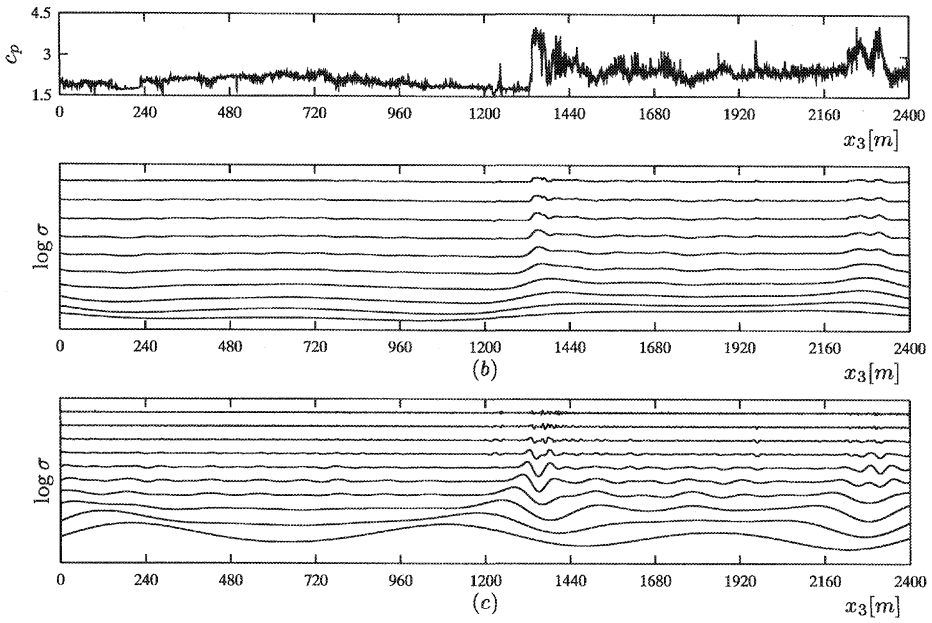

Figure 2.1 *A first example of the scaling displayed by a real well-log measurement. On the top the local compressional wave speeds are on display. The second row depicts the consecutive smoothings effectuated by a Gaussian bell-shaped smoothing kernel. The third row displays the continuous wavelet transform computed with the Mexican hat, the second order derivative of the Gaussian bell-shaped smoothing kernel.*

In figure 2.1 I depicted a first example of consecutive smoothings and details obtained by smoothing and continuous wavelet transforming a well-log profile for the local compressional wave speed. Clearly these type of profiles, see the top row of figure 2.1, demonstrate a significant scaling behaviour judged by the large activity of the wavelet coefficients, on display in lower row, and the substantial changes in consecutive smoothings that are shown on the middle row.

2.3 A multiscale approach

Given the preliminary introduction to scaling and the continuous wavelet transform I now would like to go into more detail. One of the points that come to mind first is the observation that the essential information in a signal[6] is generally carried by its regions of rapid variation, by its boundaries or edges (Mallat and Hwang, 1992; Mallat and

[6]In this case the signal refers to a well-log measurement.

Zhong, 1992). Mathematically these regions correspond to the so-called essential points (Gel'fand and Shilov, 1964; Zemanian, 1965) or singularities delineating the irregularities, transitions or edges in the signal. In words a singularity[7] is "*a point at which the derivative of a given function of a complex variable does not exist but every neighbourhood of which contains points for which the derivative exists*" (Webster, 1988). A basic property of a singularity is the notion that it manifests itself distinctly by displaying a specific type of scaling behaviour throughout a wide scale range. That is the main reason why it can be detected.

The primary aim of this section is to render the apparatus required to analyze and subsequently measure/quantify, see section 2.4, the different types of singularities. The proposed multiscale analysis presented here commits itself to unravel the signal's complexity either *locally*, via *zooms*, or *globally*, via a *partition functional* defined in terms of the *voices*. Both methods have in common that they allow for a quantification by means of exponents that express *order of magnitude estimates* for the powerlaw behaviour displayed by the wavelet coefficients in the zooms and by the partition function. From the introductory sections it became clear that the wavelet transform is well capable to live up to the task of performing the above type of analysis because it decomposes a signal into its multiscale constituents (Holschneider and Tchamitchian, 1990; Jaffard, 1991; Daubechies, 1992; Mallat and Hwang, 1992; Mallat and Zhong, 1992; Chui, 1992; Farge et al., 1993; Bacry et al., 1993).

By way of the choice for the continuous wavelets – allowing for superior multiscale analysis properties as compared to the discrete wavelet transform – the transformed domain is inherently redundant because of the wavelet's non-orthogonality. In order to circumvent this deviancy – the inspection of the whole redundant space-scale space – an efficient partitioning of this space-scale plane is proposed. This partitioning is effectuated by reviewing the scale-space behaviour of the wavelet transform modulus maxima lines which interconnect the extrema of the wavelet transform, the regions of rapid variation, across the different scales (Mallat and Hwang, 1992; Mallat and Zhong, 1992).

2.3.1 The continuous wavelet transform

It was briefly demonstrated that the continuous wavelet transform serves well as the main vehicle to conduct a multiscale analysis. The wavelet transform maps the original signal, the functional[8] f, depending on say the spatial coordinate x, to a double-indexed function $\check{f}(\sigma, x)$ depending on the scale or dilatation parameter σ and the spatial coordinate x. This mapping is defined by

$$\check{f}(\sigma, x) = \mathcal{W}\{f, \psi\}(\sigma, x) = \int_{-\infty}^{\infty} f(x') \frac{1}{\sqrt{\sigma}} \psi^* \left(\frac{x' - x}{\sigma} \right) \mathrm{d}x' \tag{2.13}$$

[7] Remark, however, that for fractal constructs the singularities lie *dense*.

[8] I use the term functional rather than function because the wavelet transform is able to host a larger class of mathematical objects then functions, see chapters 5 and 8.

where $\psi(x) \in L^2(\mathbb{R})$ is a proper[9] analyzing wavelet. When submitting this definition of the continuous wavelet transform to a close examination one can see that it represents a Cauchy type of sequence approximating the derivative, see chapter 5,

$$\langle f, \psi_{\sigma,x} \rangle = -\sigma \langle \frac{\mathrm{d}}{\mathrm{d}x'} f, \phi_{\sigma,x} \rangle, \tag{2.14}$$

where

$$\psi_{\sigma,x}(x') = \sigma \frac{\mathrm{d}}{\mathrm{d}x'} \phi_{\sigma,x}(x'), \tag{2.15}$$

represents the analyzing wavelet being given by the derivative of the smoothing function[10]. The way in which the derivative in equation (2.14) is taken is known as taking the derivative in the *distributional* or *weak sense* (Zemanian, 1965; Reed and Simon, 1980). By way of this definition one can interpret the wavelet transform as taking the derivative of the smoothed signal (Mallat and Hwang, 1992).

Since the dilatation parameter rules the effective width of the weak differentiating[11] kernel it is possible to analyze the functional f at any given but finite scale. So when applying a multiscale analysis one can resolve the local features of the functional f by zooming in on it, by increasing the resolution. This zooming is effectuated by choosing a sequence of analyzing wavelets with decreasing effective supports.

Similar to the selection of the optimal optics for a microscope, it is also necessary to impose certain conditions on the analyzing wavelet. By definition a wavelet has to have a mean zero, $\int \psi \, dx = 0$ and besides this constraint additional conditions can be imposed and this will allow for an extension of the class of signals which can be analyzed meaningfully (Bacry et al., 1993; Muzy et al., 1993). This latter notion can easily be understood given the above argument where the wavelet is seen as a test function to which one can assign certain properties in order to make equation (2.13) valid. For this reason I intendedly opted to refer to a signal as a functional, because a functional constitutes a larger class of objects, namely including the tempered distributions (Schwartz, 1957; Gel'fand and Shilov, 1964; Zemanian, 1965; Duistermaat, 1993; Mallat and Hwang, 1992) which do 'not fit' into the rather limited class of conventional functions. The reader is referred to chapter 5 for a summary of the important concepts that come with distribution theory.

[9] A wavelet that is smooth enough and having an adequate number of vanishing moments. If these conditions are not adhered to then the analysis will be dominated by the properties of the analyzing wavelet itself rather than by the function to be analyzed, and the results will be erroneous.

[10] I opted here to use a slightly different definition for the wavelet which is more in line with chapter 8.

[11] The order of the derivative depends on the number of vanishing moments, see chapter 8.

2.3.2 The wavelet transform modulus maxima

As I already mentioned most of the important information in a signal is carried by its singularities, by its regions of relatively rapid variations. Taking the derivative of such a signal pronounces this effect and can even lead to a divergence, an unbounded growth for the derivative at the singular points. With other words the derivative tends to diverge in the limiting procedure defining the derivative. Now the interesting point is that this notion directly translates to the behaviour displayed by the extrema in the modulus of the continuous wavelet transform. That probably inspired Mallat and Hwang (1992) and his co-workers to introduce a multiscale representation for the signal f in terms of a partitioning by the wavelet transform modulus maxima, the extrema for the modulus of the continuous wavelet transform.

To understand why the wavelet transform triggers on singularities one has to take into consideration that a wavelet with M vanishing moments, see definition 8.1 of chapter 8, acts as a M^{th} order distributional derivative and hence localizing singular behaviour in the M^{th} order derivative. To say it in another way the vanishing moments cause the wavelet transform to be orthogonal, read insensitive, with respect to polynomials of $\mathcal{O}(x^{M-1})$. It is this notion that makes the wavelet transform a powerful tool to analyze the singular behaviour of functionals f and it explains why the regions of fast variations are reflected into local multiscale extrema by the scale transformation.

Following Mallat and Hwang (1992), a modulus maximum (σ_0, x_0) of the continuous wavelet transform constitutes a strict local maximum for the modulus of the wavelet transform, $\max|\mathcal{W}\{f,\psi\}(\sigma,x)|$, for one specific scale and within the cone of influence given by $|x - x_0| < C\sigma$ with C being a constant depending on the wavelet (Mallat and Zhong, 1992). For a more formal definition given by Mallat and Hwang (1992) the reader is referred to chapter 8. To illustrate the role played by the wavelet transform modulus maxima I included figure 2.2 where the singularities and the corresponding modulus maxima, computed with the help of definition 8.2, are depicted.

By repeating the localization of the extrema across the different scales it is possible to define a wavelet transform modulus maxima line, a WTMML. This line commences at the smallest scale, at the resolution of the measuring device, and progresses, while connecting the extrema, to the coarser scales. Because the number of extrema within the cone of influence is restricted to one, a bifurcation occurs as soon as two cones, carrying the extrema, start to overlap. This notion is illustrated in figure 2.3 where the number of WTMML's declines as the resolution is decreased. Given this effective WTMML partitioning it is now possible to conduct a local analysis on the signal f. A first example of such local analysis is included in figure 2.3 (c) where the log amplitudes, yielded by the WTMML, the amplitudes along the WTMML's in space-scale space, are depicted versus the log scale. The different dashings are used to link the WTMML's in (b) to the amplitude behaviour in (c). Finally I would like to remark that it is possible to reconstruct the original signal

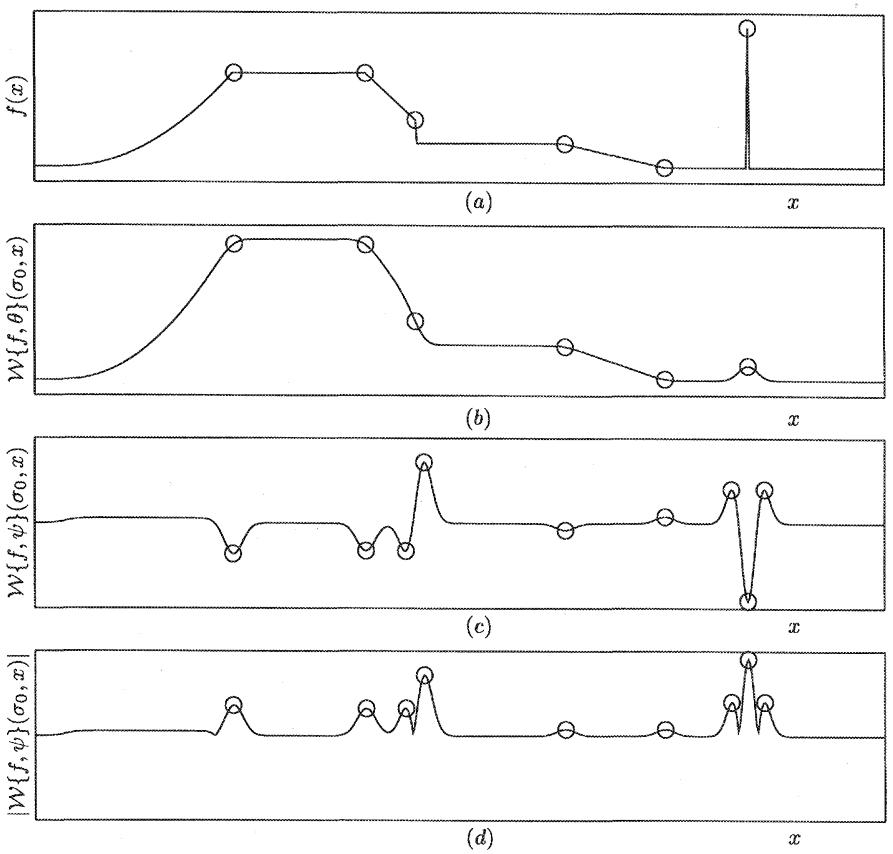

Figure 2.2 *This figure illustrates the* WTMM. *(a) The unsmoothed data set. (b) The data set smoothed with a Gaussian bell shaped smoothing function. Only a fixed scale is considered. (c) One scale of the wavelet transform of the data set. The wavelet used here is the first derivative of the Gaussian, so the wavelet has one vanishing moment. The scale of the wavelet is the same as that from the Gaussian of (b), hence this plot shows a smoothed version of the derivative of (a). The local maxima and minima indicate the points of sharp and slow variation. (d) The modulus of (c). The local maxima indicate the points of sharp variation and are denoted by the circles.*

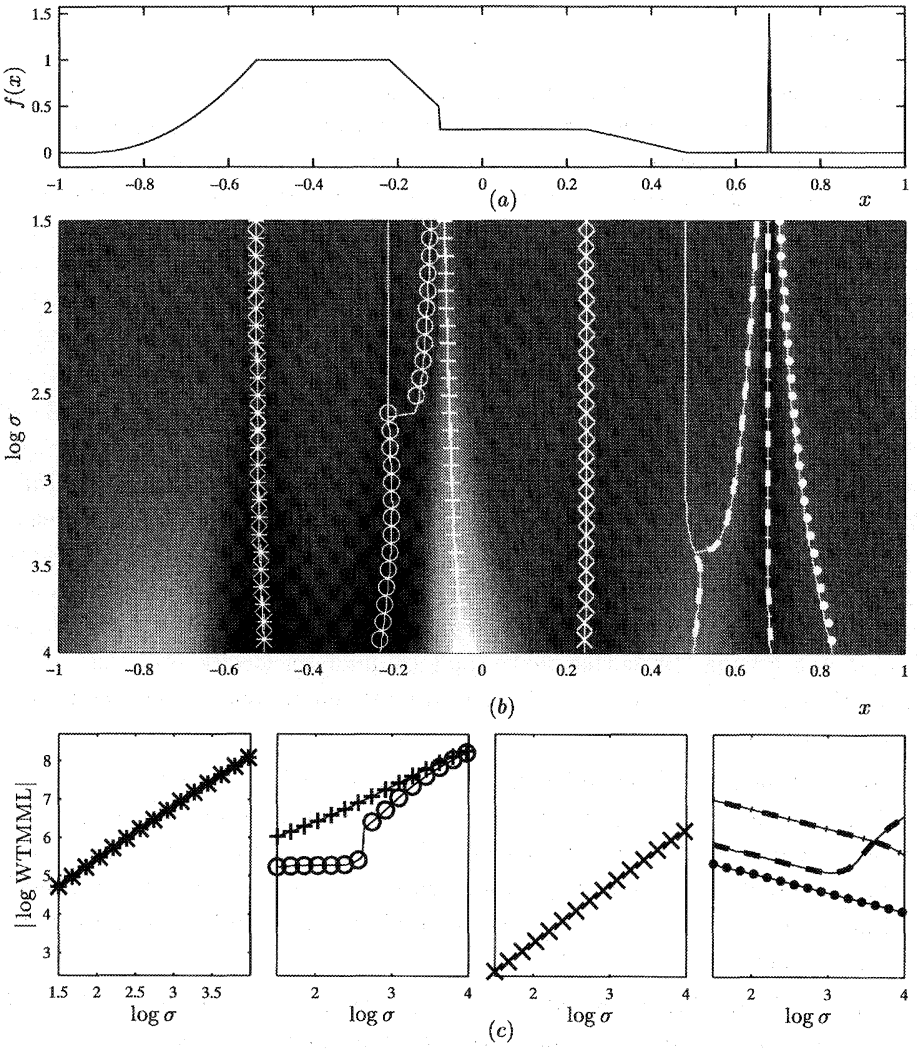

Figure 2.3 *This figure illustrates the* WTMML *partitioning. (a) The unsmoothed data set. (b) The continuous wavelet transform with the* WTMML*'s superimposed on it, notice the bifurcations. (c) The local multiscale analysis reviewing the amplitudes at the* WTMML*'s in* log-log *scale-amplitude space. The different dashings are intended to illustrate the correspondence of the* WTMML *in the space-scale space and the evidenced amplitudes along these lines. Notice the changes in amplitude behaviour when the bifurcations occur.*

f from the sparse representation in terms of the WTMML's, i.e. from a vector containing the abscissa and the corresponding moduli for every scale (Mallat and Hwang, 1992).

2.3.3 Local analysis

Given the focusing ability – within bounds imposed by the width of the cone of influence – of the continuous wavelet transform together with the effective partitioning yielded by the WTMML's, it is now possible to conduct a *local* multiscale analysis. Within this analyzing procedure the WTMML's are used as navigators directing the analysis in the redundant space-scale plane. The analysis itself simply consists of studying the behaviour of the modulus of the wavelet transform at the maxima as the resolution is being increased to the resolution at which the signal has been measured. While approaching the finest resolution the WTMML's start to point to the abscissa where the singular "points" in the signal are located. Note, however, that in an experimental setting one is always faced with an inner bound of the obtainable resolution. So the points refer to a small region of a size proportional to the accuracy of the measuring device. Therefore it is in this case more appropriate to refer to these singularities as *regions the behaviour of which can, within the resolution of the measured signal, not be discerned from being singular*[12], see the discussion in section 2.5. So the concept of a singularity can only be of relevance in the physical context, when taking the limit is interpreted as studying the behaviour of the signal as the small scale limit is approached. In chapter 8 I provide a series of theorems by Mallat and Hwang (1992) that make the local analysis precise. As to summarize the local multiscale analysis boils down to the examination of zooms along the partitioning provided by the WTMML.

In figure 2.3 I included an example of conducting a local analysis on the signal of figure 2.2. Indeed one can see that the WTMML delineate the regions where the functional is singular. One can also clearly recognize the bifurcations occurring when the cones carrying a WTMML start to overlap, when the WTMML start to merge. Judged by the amplitudes, depicted in the bottom row, these bifurcations do have a substantial effect on the amplitudes versus scale behaviour.

In this section I only briefly introduced the concept of a local scaling analysis. In section 2.4.1 I will introduce order of magnitude estimates for the behaviour of the modulus maxima as a function of the scale index σ. These estimates will not only provide an efficient quantification/parameterization but also offer a mathematical characterization of the signal's regularity within the available resolution range.

2.3.4 Global analysis

Within the realm of multiscale analysis one is generally not only interested in the *local* features of a signal. The *global* multiscale characteristics are of interest as well. For

[12] That is to say not to be discerned from being non-differentiable or singular in some order of their derivative.

example think of statistical physics where attempts are made to model and characterize the complex behaviour displayed by systems with a large number of degrees of freedom (Lichtenberg and Lieberman, 1992), such as hydrodynamic turbulence (Mandelbrot, 1974; Hentschel and Procaccia, 1983; Schertzer and Lovejoy, 1987b) or sedimentation processes (Walden and Hosken, 1985; Leary, 1991; Muller et al., 1992; Saucier and Muller, 1993).

The common denominator in all these systems appears to be a seaming insensitivity of the system to certain scale operations. In other words the system's behaviour on a larger scale can be seen as a scaled up, averaged, version of the system's action on a smaller scale, i.e. the system adheres to certain renormalization group properties (Wilson, 1983). Clearly this observation of a certain scale-invariance is closely related to the fractal concept where the objects are being constructed via an iterative scheme, say multiplicative cascades, where the system adds scaled down versions of itself to increasingly smaller scales. By means of this construction the fractal will display a certain scale-invariance as does the renormalization group.

Another important property shared by complex systems is their apparent degree of *non-stationarity* and *intermittency*. The non-stationarity refers to the notion that, for example, the two-point correlation function looses its meaning because of a divergence of the variance, i.e. the emergence of a non-integrable DC component for the power spectrum. The intermittency on the other hand rules the spikiness, the bursty fashion in which certain (geophysical) events occur. In such a situation the underlying system behaves highly non-linear resulting in an output where the relative quiescent regions are superseded by bursty and violent active regions.

As will become clear later the *non-stationarity* demands certain *integrability* conditions, number of vanishing moments, to be imposed on the analyzing wavelet whereas the *intermittency* requires the examination of a large range of statistical moments. In cases where the singularities become negative an additional regularity condition must be imposed on the analyzing wavelet.

As in statistical physics a multiscale partition function is introduced to explore the global characteristics of the singularity structure pertaining to the data. Following Bacry et al. (1993) and Muzy et al. (1993) one can define this partition function in terms of a "stacking" procedure along the voices yielded by the multiscale partitioning supplied by the WTMML's. This partition function is defined as follows,

$$Z\{f,\psi\}(\sigma,q) = \frac{1}{\sqrt{\sigma}} \sum_{m\in M} [\sup\{\text{WTMML}_\sigma(m)\}]^q, \tag{2.16}$$

where $\text{WTMML}_\sigma(m)$ refers to a maximum, a WTMM, at scale σ yielded by the m^{th} modulus maxima line and the supremum, sup, denotes the supremum of the WTMML. This supremum is necessary to circumvent a possible divergence for negative powers q in case the values of the modulus maxima become too small (Bacry et al., 1993; Muzy et al.,

1993). The reader is kindly referred to chapter 6 where I introduce the basic concepts surrounding (multi)fractals and to chapter 8 where I review the technical details, by a series of theorems taken from the work of Bacry et al. (1993), that lie behind conducting the global multiscale analysis with the help of the WTMML's.

Inspection of equation (2.16) shows that by varying the *positive* order for the power q the *active* regions, the strong singularities, are being emphasized for increasing positive q. An opposite effect occurs for the *passive* regions that are emphasized for increasing *negative* values for q. In this way the partition function $Z(\sigma, q)$ acts as a sieve where the q hunts for the different type of singularities, the active and passive regions.

For those not familiar with this type of partition function it might look somewhat artificial. I can assure that that is not the case, especially because it can easily be shown that $Z\{f, \psi\}(\sigma, q)$ is directly related to the sample form of the conventional structure function or variogram, $D(\sigma)$, by setting the wavelet ψ to the Poor Man's wavelet[13], $\psi_\sigma(x) = \delta(x + \frac{1}{2}\sigma) - \delta(x - \frac{1}{2}\sigma)$, and by setting $q = 2$. This variogram is well known in geostatistics and is used for studying processes with stationary increments (Mandelbrot and Wallis, 1969; Tartarskii, 1971; Yaglom, 1987; Bacry et al., 1993; Schmitt, 1993). Finally, as in case of the local multiscale analysis, one is interested in the behaviour for the partition function as the resolution is increased to the resolution at which the signal has been measured.

2.4 Multiscale quantification and modelling

Given the brief introduction on the local and global multiscale analysis by means of the WTMML framework one is now set to come up with order of magnitude estimates. These estimates not only locally characterize the decay/growth rate for the modulus of the wavelet coefficients along zooms guided by the WTMML but also refer to a characterization for the joint scaling behaviour displayed by the WTMML's and captured by the scaling behaviour of the partition function. As in many fields in physics it seems to be justifiable to limit oneself to order of magnitude estimates that aim at revealing the scaling behaviour in terms of powerlaws characterized by scaling exponents. This choice corresponds for the local analysis to come up with estimates for the *Hölder exponents*, α, in

$$|\mathcal{W}\{f, \psi\}(\sigma, x)| \propto \sigma^{\alpha + \frac{1}{2}}, \tag{2.17}$$

while for the global analysis one tries to reveal the *mass exponent function*, $\tau(q)$, from the scaling of the partition function

$$Z\{f, \psi\}(\sigma, q) \propto \sigma^{\tau(q)}. \tag{2.18}$$

[13]The Poor Man's wavelet defines the increment.

The reader is referred to chapter 8 for detailed definition and theorems substantiating equations (2.17) and (2.18). Mathematically the above exponents refer to

- a proper characterization, by means of the local Hölder/Lipschitz exponents (Holschneider and Tchamitchian, 1990; Jaffard, 1991; Daubechies, 1992; Mallat and Hwang, 1992) expressing the *local degree of regularity*, i.e. the *local degree of differentiability.*
- a proper characterization of the *global regularity*, by means of the singularity spectrum (Mandelbrot, 1974; Hentschel and Procaccia, 1983; Parisi and Frisch, 1985; Collet, 1986; Schertzer and Lovejoy, 1987a; Siebesma, 1989; le Méhauté, 1991; Lichtenberg and Lieberman, 1992; Ott, 1993; Muzy et al., 1993; Bacry et al., 1993; Davis et al., 1994; Holschneider, 1995), expressing the hierarchy of the different singularities. This singularity spectrum is related, via the Legendre transform, to the mass exponent function which, on its turn, is related to the generalized Renyi dimensions.

2.4.1 Local quantification

In this section the work of Mallat and Hwang (1992) will be followed closely. He provides – by means of a number of concise mathematical theorems, see chapter 8 – an apparatus, based on a WTMML partitioning, that is able to measure the local degree of differentiability of a functional containing an isolated singularity or singularities the cone of influence of which overlap for a certain scale range. The degree of differentiability or regularity is expressed by a Hölder exponent α, i.e. a functional being Hölder α at a point is α times differentiable at that point[14]. This exponent can be estimated by analyzing the amplitudes of the WTMM's along the WTMML's in the log-scale log-amplitude space. In this way the powerlaw relationship,

$$|\mathcal{W}\{f,\psi\}(\sigma,x)| \leq C\sigma^{\alpha+\frac{1}{2}} \qquad \forall(\sigma(s),x(s))_{m,\, m\in M} \tag{2.19}$$

is fully explored and becomes

$$\log|\mathcal{W}\{f,\psi\}(\sigma,x)| \leq \log C + (\alpha+\frac{1}{2})\log\sigma, \tag{2.20}$$

the asymptotic behaviour of which simplifies to

$$\limsup_{\sigma\downarrow 0} \frac{\log|\mathcal{W}\{f,\psi\}(\sigma,x)|}{\log\sigma} = \alpha+\frac{1}{2}. \tag{2.21}$$

Equation (2.21) constitutes the main vehicle to compute the Hölder exponent to be associated with the singularity.

While applying the local analysis method attention should be paid to the following:

[14]Only in case $\alpha > 0$ otherwise the functional is a tempered distribution.

- that the emergence of a WTMML may have been caused by a fast oscillation rather than a singularity. This oscillatory region may not be singular. However, the reverse argument holds: a signal containing a singularity emanates a WTMML when the proper wavelet is being used.
- that when the Hölder exponent equals the number of vanishing moments then one has to increase the number of vanishing moments. If this, after repetitive tries, does not help[15] one has most probably to do with a function that is C^{∞}, i.e. infinitely many times differentiable.
- that, for computational purposes, one has to restrict the number of vanishing moments.
- that there exist non-isolated singularities, singularities the WTMML's of which start to interfere at a certain resolution.

In chapter 8 I included a number of examples dealing with the Hölder exponent estimation of isolated algebraic singularities (Staal, 1995). To illustrate what happens in case the singularities can not longer be considered as isolated I included the subsequent subsection.

Measuring the Hölder exponent for non-isolated singularities

Generally one is more interested in finding local estimates for the Hölder exponents yielded by signals containing more than one singularity. In that case there will always be a scale range within which the WTMML's start to interfere with each other. With other words the cones of influence will start to overlap and this is indicated by a *bifurcation* of the WTMML. Despite the apparent mutual influence by the singularities it is still possible to come up with Hölder exponent estimates. This requires a rather technical theorem, see chapter 8 and for this reason I will limit myself to reviewing the simple example of a box-car[16] only. In figure 2.4 I have included this box-car which can, within the seismic context, be associated to a "thin" layer sandwiched between two homogeneous halfspaces. Inspection of the WTMML yielded by this box-car clearly shows that two scale ranges, separated by a bifurcation, can be distinguished. These two scale ranges refer, on the one hand, to a scale range acting on scales smaller than the spatial extent of the box-car, whereas on the other hand to a scale range exceeding the support of the box-car. Inspection of the amplitude behaviour of the WTMML's over the first scale range clearly shows that the estimates for the Hölder exponent correspond to a Hölder exponent[17] for an isolated discontinuity, $\alpha = 0$. But as soon as the WTMML's start to overlap, as soon

[15] In the sense that the α still equals the now increased number of vanishing moments.

[16] A box-car function is a characteristic function or indicator function, a function being non-zero over a specific range.

[17] Notice that due to the $L^2(\mathbb{R})$-normalization the slopes for the magnitudes of the WTMML's are off by a factor 1/2.

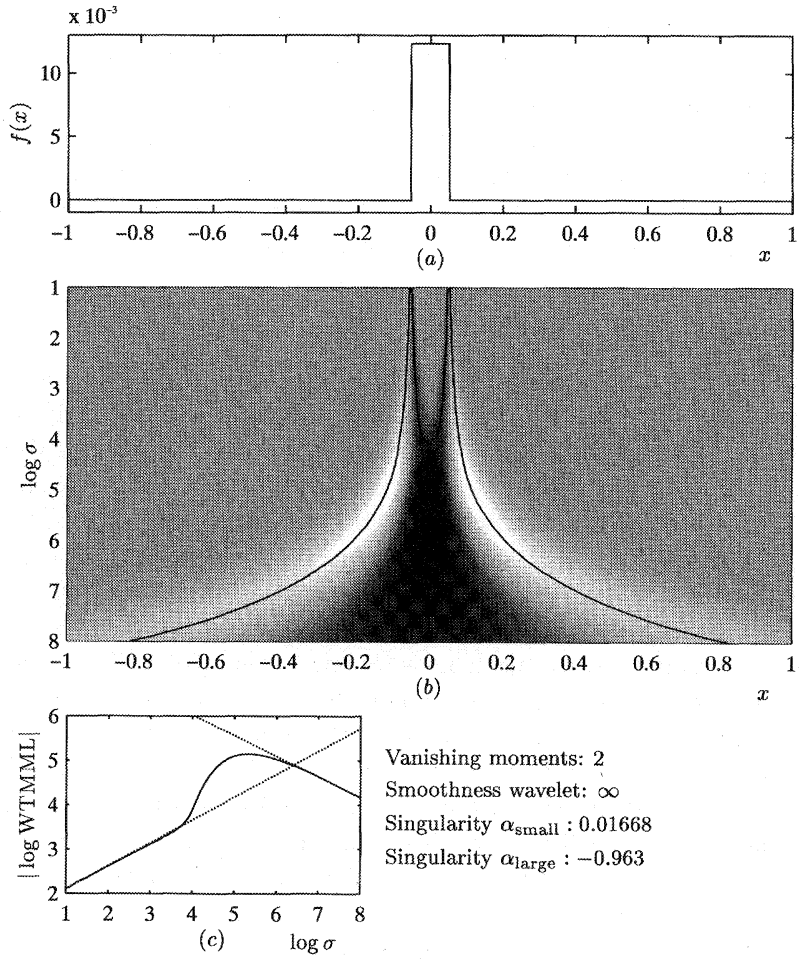

Figure 2.4 *This example illustrates the local scaling characteristics of two non-isolated singularities, discontinuities in this case. Clearly for the small scale range Hölder exponents are found consistent with the discontinuities. But as soon as the cones of influence start to overlap a bifurcation occurs having a drastic impact on the amplitude behaviour along the middle and to a lesser extent for the outer* WTMML*'s. The asymptotic behaviour for $\sigma \to \infty$ yields a scaling behaviour corresponding to that of a delta distribution.*

as the bifurcation occurs, a complete different behaviour becomes evident. The Hölder exponent to be associated to the decay rate of the WTMML equals $\alpha = -1$ for $\sigma \to \infty$. This means that for this scale range the box-car, layer, acts as a functional that can not be discerned form a δ-distribution. That is to say that the box-car function can be seen as an element of the delta convergent sequence (Gel'fand and Shilov, 1964).

The lesson to be learned from this example is that the signal, the functional, can display a different scaling behaviour for different scale ranges. However, this does not mean that the concept of parameterization by the Hölder exponents looses its meaning. It just tells one that care should be taken with respect to the scale range one is interested in. In section 2.5.3 I will pay more attention to this delicate issue.

2.4.2 Global quantification

Performing a global multiscale analysis obviously serves different merits than conducting a local analysis. Despite the fact that local information on the signal's irregularity is lost, it serves as an excellent tool to obtain information on the generalized Hausdorff dimensions, the mass exponent function $\tau(q)$ and the singularity spectrum $f(\alpha)$. In fact it gives – by means of the singularity spectrum – a comprehensive picture on the hierarchy of singularities present in highly intermittent, chaotic signals such as well-log measurements. For many purposes this type of global information suffices, because it governs the scaling behaviour of the statistical moments and provides information on the signal's intermittency, stochastic non-stationarity and global differentiability and integrability. It also circumvents problems related to the mutual interference between the different singularities. Before going into detail on the actual measuring of the singularity spectrum let me briefly define (Renyi, 1970; Mandelbrot, 1974; Hentschel and Procaccia, 1983; Parisi and Frisch, 1985; Schertzer and Lovejoy, 1987a; Siebesma, 1989; le Méhauté, 1991) the generalized or Renyi dimensions D_q and the mass exponent function $\tau(q)$ in terms of a WTMML partitioning (Bacry et al., 1993; Muzy et al., 1993).

Definition 2.1: Generalized or Renyi dimensions D_q
In order to define the generalized dimensions D_q use the partition function defined in equation (2.16),

$$Z\{f,\psi\}(\sigma,q) = \frac{1}{\sqrt{\sigma}} \sum_{m \in M} [\sup\{\text{WTMML}_\sigma(m)\}]^q, \tag{2.22}$$

where f is the singular function, ψ a proper analyzing wavelet and σ the scale. The set of generalized dimensions D_q is then defined by,

$$(q-1)D_q \triangleq \tau(q) \triangleq \lim_{\sigma \downarrow 0} \frac{\log Z\{f,\psi\}(\sigma,q)}{\log \sigma}, \tag{2.23}$$

implying a scaling behaviour for the partition function $Z(\sigma, q)\{f, \psi\}$, *for small* σ,

$$Z(\sigma, q) \propto \sigma^{\tau(q)} \tag{2.24}$$

and where $\tau(q)$ *is the mass exponent.*

The singularity spectrum itself is formally defined as (Bacry et al., 1993)

Definition 2.2: Singularity spectrum of a function
A singularity spectrum of a function $f(x)$ *is the function* $\mathrm{f}(\alpha)$, $\alpha \in H_f$ *(the set of finite Hölder exponents of* f*) such that*

$$\mathrm{f}(\alpha) = \dim_H\{x_0 \in \mathbb{R} | \alpha(x_0) = \alpha\} \tag{2.25}$$

where $\dim_H$ *denotes the Hausdorff dimension*

and expresses the Hausdorff dimension to be associated with the set of points the Hölder exponent of which lie between α and $\alpha + \mathrm{d}\alpha$. Quantitatively the singularity spectrum delineates the rate of occurrence of a certain singularity and therefore unravels the hierarchy of scaling exponents. Again the reader is referred to chapters 6 and 8 for a more thorough introduction of these multifractal concepts and for more technical details.

For example when the functional $f(x)$ contains only one isolated singularity then the singularity spectrum will be zero everywhere except at $\alpha = \alpha_0$, where it is infinitesimally small, since the dimension of a point is zero. On the other hand a singularity spectrum yielding $\mathrm{f}(\alpha) = 1$ at one particular abscissa $\alpha = \alpha_0$ refers to a functional being singular everywhere, where the singularities are all of the same strength, having the same Hölder exponent α_0. In chapter 8 a number of illustrating examples have been included.

2.4.3 Categorization of singular multiscale models

When studying the current literature dealing with the characterization and modelling of irregularity one is struck by the sheer variety of the different approaches. This calls for some sort of categorization although one always has to keep in mind that data must make the final "decision". The categorization I like to propose is very much in line with the ideas coined by Davis et al. (1994) and involves invoking a characterization of the singular processes that are being used as (toy-)models to generate, capture and parameterize the irregularity one is often confronted with when tackling problems in geophysics (Schertzer and Lovejoy, 1987a, 1993; Davis et al., 1994). The characterization is based on external characteristics that come down to

- **the signal's non-stationarity**: This non-stationarity[18] is linked to the observation that irregular processes tend to display long ranged correlations withstanding

[18] Remark that this non-stationarity refers only to a specific type of non-stationarity namely that of a powerlaw type of divergence for the covariance function.

the identification of some sort of correlation length marking a break in the scaling, i.e. a break in the scale range where the data shows some sort of scale-invariance. In cases of non-stationarity one observes that certain statistical moments diverge. The degree of this non-stationarity is expressed in terms of the exponent H which I introduced in chapter 7. For $H > 0$ this implies, for example, that the data can not longer be considered as densities since the mean is not longer a conserved quantity. This manifests itself, for instance, in the notion that the covariance function of Brownian motion, $H = 1/2$, does not longer exist because the variance diverges. Fortunately Brownian motion and its generalization fractional Brownian motion, $0 < H < 1$, are processes with stationary increments so one can work with structure functions instead, see chapter 7 (Mandelbrot and Wallis, 1969; Tartarskii, 1971; Yaglom, 1987).

- **the signal's intermittency**: This intermittency is linked to the spikiness, the erratic occurrence of major catastrophic events, the outliers. Clearly this property withstands a characterization by the Gaussian distribution because the measure of the process is not longer concentrated around the mean.

These two identifying characteristics are directly connected to the generating mechanism that is envisaged to capture the main characteristics, the overall texture, of the data. These mechanisms comprise

- **an additive mechanism**: This mechanism is being held responsible for the emergence of the long range correlations as evidenced in Brownian motion, fractional Brownian motion and Levy flights (Mandelbrot and Wallis, 1969; Mandelbrot, 1982; Montroll and Schlessinger, unknown; Schertzer and Lovejoy, 1993; Samarodnitsky and Taqqu, 1994; Klafter et al., 1996). All these processes, see chapter 7 for details, have in common that they are generated via a (fractional) integration acting on a random measure. Due to the integration they are heavily correlated, a notion becoming manifest in the divergence of certain statistical moments, such as the variance, and this gives rise to a stochastic non-stationarity. The cross-over from stationary to non-stationary behaviour can be defined in terms of the divergence for the DC component, $(k = 0)$, of the power spectrum, $S(k)$, with k being the spatial frequency. This cross-over occurs at $\beta = 1$ in

$$S(k) \propto \frac{1}{k^{\beta}}, \tag{2.26}$$

yielding stationarity for $\beta < 1$ and non-stationarity for $\beta > 1$. The process is said to be stationary in its first increments[19] when $1 < \beta < 3$. Visually speaking the degree of non-stationarity, being related to the degree of fractional integration, H, according to $\beta = 2H + 1$, rules the roughness of the process. This latter aspect

[19] First increment is given by $\Delta f(\sigma, x) \triangleq f(x + \sigma) - f(x)$.

is easily comprehended because for large H the decay rate of the power spectrum increases diminishing the high frequency contributions. Generally speaking, however, these type of processes fail to grasp the truly large outliers and they display a homogeneous type of scaling behaviour, they are monofractal, and this corresponds to the emergence of only one ruling scaling exponent, the degree of fractional integration. Specific care has to be taken concerning the measure which is going to be fractionally integrated. If this measure is not positive, like for a density[20], then one can not guarantee the fractionally integrated result to be positive and that can have serious consequences for its applicability. In figure 2.8 of section 2.4.4 I depicted an example that refers to fractionally integrating a realization of a singular white noise measure and of a universal multifractal density.

- **a multiplicative mechanism**: This type of mechanism is found in multiplicative cascades (Mandelbrot, 1974; Hentschel and Procaccia, 1983; Collet, 1986; Schertzer and Lovejoy, 1987a) where it is held responsible for the generation of highly intermittent densities where the quiet regions are superseded by large violent bursts that quickly return to where the measure is concentrated. In chapter 6 I demonstrate that this type of behaviour is displayed by the binomial multifractal. That type of multifractal is a simple example of a multifractal being generated by a multiplicative cascade yielding a heterogeneous scaling which is reflected in the fact that these constructs contain a whole hierarchy of scaling exponents. Within this continuum of scaling exponents and their associated Hausdorff dimensions one can recognize the co-dimension determining the *sparseness of the mean*, $C_1 = D - D_1$ as one of the ruling characteristic exponents. Here D denotes the dimension of the embedding space and, unless stated otherwise, equals 1. It is this exponent which predominantly captures the apparent intermittency and for the extreme sparse fractal sets, given by only one isolated algebraic singularity, it equals unity, $C_1 = 1$ which is consistent with the fact that the dimension of the singular support, being limited to a single point, is zero.

- **a mixture of the two**: In many situations the combination of two mechanisms gives superior results. In this case this corresponds to conducting a fractional integration on the densities which came out of the multiplicative cascades (Schertzer and Lovejoy, 1987a; Davis et al., 1994). This results in a shift to the right of the singularity spectrum, adds correlation and increases the smoothness.

To illustrate the above characterization I included figure 2.5. In this figure the stationarity and intermittency parameters are plotted versus each other. As annotation I included the main characteristics for the singularity spectrum, f(α), corresponding to the H and C_1 abscissa in the $H - C_1$-plane where $(H, C_1) \in [-\frac{1}{2}, 1] \otimes [0, 1]$ covers the necessary

[20] With a density I mean a positive measure with respect to which one can define a conserved quantity such as the mean.

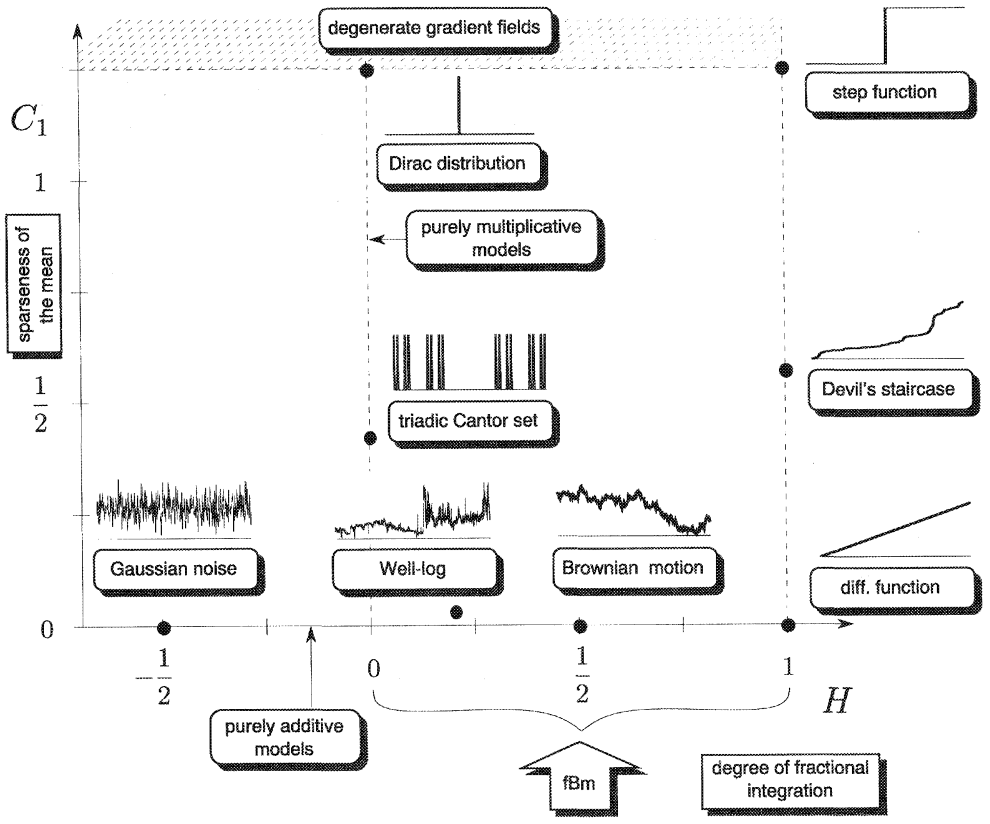

Figure 2.5 *This figure serves as an illustration of the proposed categorization of singular scaling models. The horizontal axis delineates the degree of fractional integration, H, shifting the singularity spectrum. On the vertical axis the amount of sparseness of the mean, C_1, is depicted. In the H-C_1-plane I included a number of examples of realizations and remarks on the properties of the singularity spectrum. I included the corresponding* $\mathrm{f}(\alpha)$ *spectra in figure 2.6*

ground. In the left-hand side lower corner one finds white noise a random process being singular everywhere with a scaling exponent of $\alpha = -\frac{1}{2}$, i.e. it is fractionally "integrated" to the degree $H = -\frac{1}{2}$, yielding a slope for the power spectrum, $S(k) \propto \frac{1}{k^\beta}$, of $\beta = 0$. The singularity spectrum in this case equals $\mathrm{f}(\alpha) = 1$ for $\alpha = H$ an notion easily reconciled with the fact that the dimension of the singular support of white noise, $D_0 = \mathrm{f}(\alpha_0)$, equals one. When moving to the right one fractionally integrates yielding an increase of the scaling exponents α. For $H = 0$, i.e. $\beta = 1$ the cross-over from stationary to non-stationary processes occurs. Then at $H = \frac{1}{2}$ one arrives at Brownian motion marking the

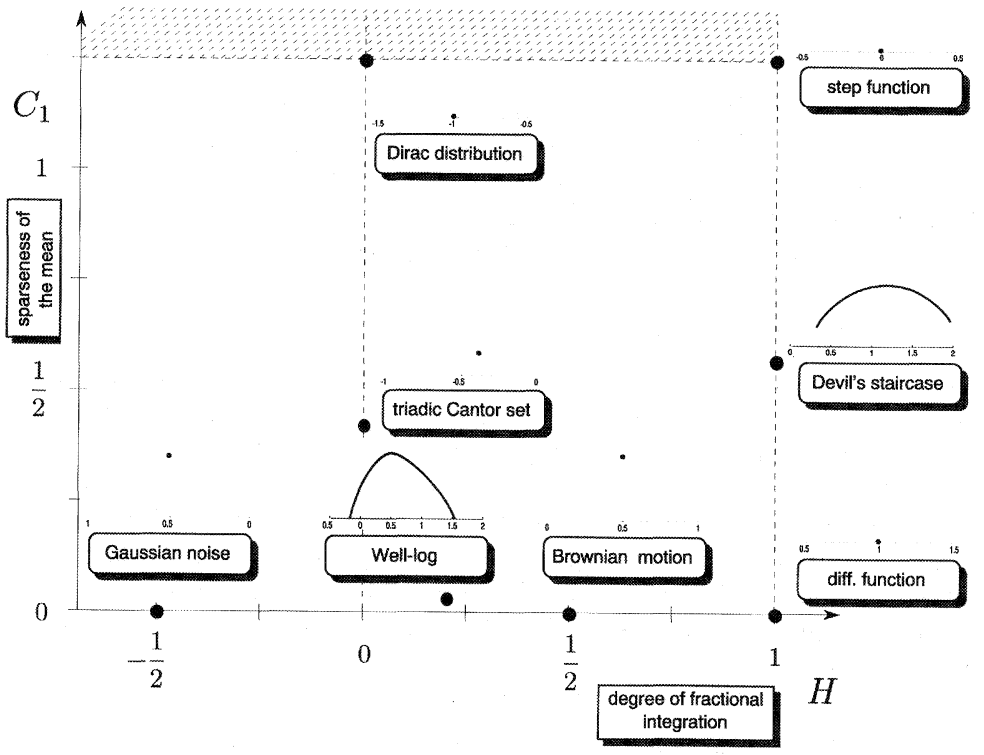

Figure 2.6 *This figure serves as an illustration of the proposed categorization of singular scaling models. The horizontal axis delineates the degree of fractional integration, H, shifting the singularity spectrum. On the vertical axis the amount of sparseness of the mean, C_1, is depicted. In the H-C_1-plane I included a number of examples of the corresponding $f(\alpha)$ spectra.*

transition from anti-persistent, $0 < H < \frac{1}{2}$ to persistent $\frac{1}{2} < H < 1$ fractional Brownian motion. Finally at $H = 1$ the process becomes smooth, i.e. C^1. The reader is referred to chapter 7 for a more general discussion on stochastic monofractals. In figure 2.8 of section 2.4.4 I included a series of examples elucidating the effect of smoothing by fractional integration.

When going up from bottom to top the field undergoes a different experience in the sense that it transforms from a pure additive process, displaying a homogeneous monofractal scaling behaviour, to a multiplicative process. This is accompanied by the emergence of a whole suite of singularities with different scaling exponents, an observation substantiated by the non-linearity of the mass exponent function, $\tau(q)$. What remains by pushing the

sparseness of the mean parameter, C_1, to the limit, are the isolated singularities. In figure 2.9 of section 2.4.4 I displayed a number of universal multifractal simulations that differ in their sparseness of the mean, their intermittency.

There exist several alternatives to generate intermittent densities and the binomial multifractal has been one of them. Judged by the relative success of the universal multifractal cascade model in characterizing geophysical phenomena (Schertzer and Lovejoy, 1987a, 1993; Schmitt, 1993; Davis et al., 1994) I will briefly introduce this model followed by a section where I bring fractals in relation to the renormalization group equations.

2.4.4 The universal multifractal cascade model

Global multiscale analysis conducted on a wide variety of geophysical data sets has shown evidence that there might exist a universality[21] class for multifractal cascade models (Schertzer and Lovejoy, 1987a, 1993; Schmitt, 1993). This universality class is composed of a continuous[22] multiplicative cascade where the properly normalized α_l *stable* Levy process[23] is used as the scale invariant *generator*. In this way Schertzer and Lovejoy (1987a) have been able to come up with an effective model parameterized by only three parameters that captures the continuum of scaling exponents as they occur in the singularity spectrum f(α). Since this model is based on a stable Levy process their work is set along the lines of probability theory. However that does not withstand a connection with the global multiscale framework reviewed so far.

First of all Schertzer and Lovejoy (1987a) introduce a partition function in terms of the stochastic expectation of a coarse-grained *density* living at a scale σ,

$$\epsilon(\sigma, x) = (\epsilon * \phi_\sigma)(x), \tag{2.27}$$

where the convolution runs over all realizations and where ϵ refers to the bare density, the density yielded by the fully developed cascade. The partition function associated with this quantity reads

$$\langle \epsilon(\sigma, \cdot) \rangle \propto \sigma^{-K(q)} \quad q \geq 0 \tag{2.28}$$

with the angular brackets denoting the stochastic expectation operation[24]. The $K(q)$ is related to the second Laplace characteristic function. Remark that this partition function resembles the definition for the partition functions introduced earlier. One only has to

[21] The author is aware of and agrees with the statement: "For quite a while I have set for myself the rule if a theoretician says "universal" it just means pure nonsense", by Wolfgang Pauli. However, in this context it can be argued that this class seems to give rise to a behaviour that is shared by surprisingly many different geophysical phenomena.

[22] Continuous means in this case multifractals which consist of a continuous densification of the resolution σ rather than resolution increments by a fixed factor, say $\sigma \rightarrow \frac{1}{2}\sigma$.

[23] Note that the Levy index α_l has nothing to do with the Hölder exponent α. See chapter 7.

[24] Not to confuse with an inner product denoted by $\langle \cdot, \cdot \rangle$.

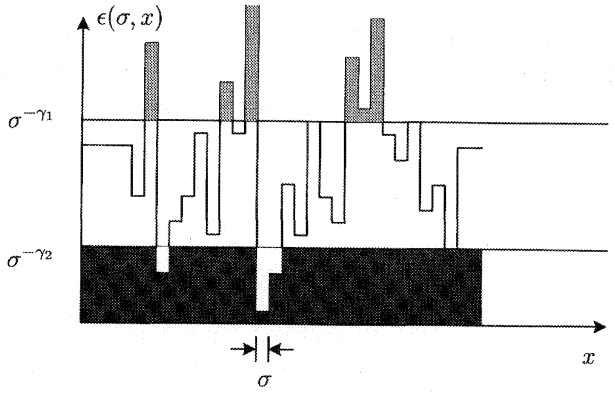

Figure 2.7 *In this figures it is schematically demonstrated what the co-dimension exactly expresses. The coarse grained field at the scale σ is depicted together with two thresholds $\sigma^{-\gamma_1}$ and $\sigma^{-\gamma_2}$, corresponding to two orders of singularity: $\gamma_2 < \gamma_1$.*

replace the *ensemble* averaging by a *spatial* averaging[25]. For that reason the $K(q)$ and $\tau(q)$ functions can shown to be related according to $\tau(q) = (q-1)D - K(q)$, with D being the dimension of the embedding space. Along similar lines a mutual relationship can be discovered between the probability distribution delineating the space occupied by singularities exceeding order γ,

$$\Pr(\epsilon(\sigma, \cdot) > \sigma^{-\gamma}) \propto \sigma^{c(\gamma)}, \tag{2.29}$$

and the singularity spectrum f(α). Here $c(\gamma)$ refers to the co-dimension function of the singularities being related to f(α) according to f(α) $= D - c(\gamma)$ with $\gamma = D - \alpha$.

To illustrate the notion of the co-dimension function I included figure 2.7 (Schertzer and Lovejoy, 1993; Pecknold et al., 1993). It illustrates schematically the notion that the co-dimension expresses the probability of finding a value for the coarse-grained field $\epsilon(\sigma, x)$ with a value exceeding the scale dependent threshold $\sigma^{-\gamma}$, $\Pr(\epsilon(\sigma, \cdot) > \sigma^{-\gamma}) \propto \sigma^{c(\gamma)}$. Remark that the behaviour of the partition function depicted in equation (2.28) or equivalently the partition function $Z\{f, \psi\}(\sigma, q)$ are not only related to the statistical moments but also, via the Mellin transformation, to the probability function (Schertzer and Lovejoy, 1993). A similar relationship exists between the $K(q)$ function or equivalently the mass exponent function[26] $\tau(q)$ and the co-dimension function $c(\gamma)$ or the singularity spectrum f(α). This relation is expressed by the Legendre transform, which is a saddle point approximation to the Mellin transform (Schertzer and Lovejoy, 1993).

[25] The author is aware of the fact that one requires ergodicity to do this. Is is beyond the scope of this thesis go into more detail on this.

[26] Which on its turn is related to generalized Hausdorff dimensions or Renyi dimensions, see chapter 6.

One of the key points of the work by Schertzer and Lovejoy (1987a) lies in finding closed form expressions for the second Laplace characteristic function, $K(q)$, and its Legendre transform, the co-dimension, $c(\gamma)$. These expressions read

$$K(q) = \begin{cases} \frac{C_1}{\alpha_l - 1}(q^{\alpha_l} - q) & 0 < \alpha_l < 2 \wedge \alpha_l \neq 1 \\ C_1 q \ln q & \alpha_l = 1 \end{cases} \tag{2.30}$$

and

$$c(\gamma) = \begin{cases} C_1(\frac{\gamma}{C_1 \alpha_l'} + \frac{1}{\alpha_l})^{\alpha_l'} & 0 < \alpha_l < 2 \wedge \alpha_l \neq 1 \\ C_1 e^{\frac{1}{\gamma}-1} & \alpha_l = 1 \end{cases} \tag{2.31}$$

with $\frac{1}{\alpha_l} + \frac{1}{\alpha_l'} = 1$ and provide an *effective parameterization* for $q > 0$. The parameters ruling the behaviour of the continuous cascade model, with Levy noise as its generator, comprise

- H describing the degree of non-conservation of the mean, see figure 2.8.
- C_1 describing the sparseness of the mean, see the left-hand side of figure 2.9, and is defined as

$$C_1 \triangleq \{\partial_q K(q)\}_{q=1} \tag{2.32}$$

  also known as the co-dimension where the measure is concentrated (Schertzer and Lovejoy, 1987a; le Méhauté, 1991).
- α_l describing the degree of multifractality, $\alpha_l = 0$ monofractal, $\alpha_l = 2$ log-normal, see the right-hand side of figure 2.9.

The additional parameter H I introduced expresses the degree of fractional integration implying a shift from a conservation of the mean[27], $\langle \epsilon(\sigma, \cdot) \rangle = 1, \quad \forall \sigma$, necessary for the initial definition of the cascade, to a non-conservation $\langle \epsilon(\sigma, \cdot) \rangle \propto \sigma^H$. In chapter 5 I introduced the operation of fractional integration/differentiation and I like to remark that extreme care should be taken by invoking this operation because in many cases it can only be given a proper meaning in the distributional sense. The impact of applying the fractional integration on the quantities above comes down to the mapping,

$$\begin{aligned} K(q) &\longrightarrow K(q) - qH \\ c(\gamma) &\longrightarrow c(\gamma - H). \end{aligned} \tag{2.33}$$

[27] Here in an ensemble sense, the multiplicative cascades of Schertzer and Lovejoy (1987a) are canonical while the binomial multifractals are microcanonical.

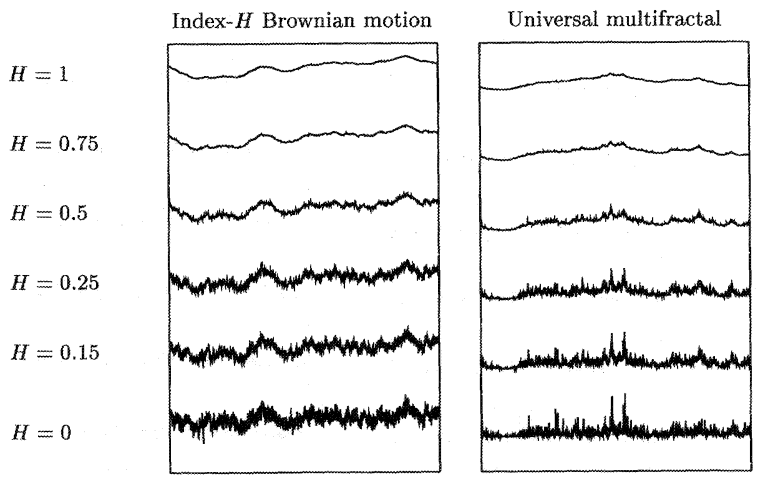

Figure 2.8 *Comparison is made between fractional integrations with values for H ranging from bottom to top, $H = 0, 0.15, 0.25, 0.5, 0.75, 1$, of white Gaussian noise (left-hand side), yielding index-H Brownian motion, and the same fractional integrations for a universal multifractal density (right-hand side) with $C_1 = 0.25$ and $\alpha_l = 1.5$. It is clear that the value of H rules the degree of "smoothness".*

This, on its turn, implies for the slope of the power density spectrum, using equation (2.33),

$$S(k) \propto \frac{1}{k^{1-K(2)}} \longrightarrow S(k) \propto \frac{1}{k^{1-K(2)+2H}}, \tag{2.34}$$

where $K(2)$ can be interpreted as a correction term on the effective slope of the power density spectrum given by β.

Estimates for the H can obtained by evaluating the Laplace characteristic function $K(q)$ at $q = 1$ and using the property of conservation of the mean, i.e $K(1) = 0$,

$$H = -\{K(q)\}_{q=1}. \tag{2.35}$$

At this point I will refrain from going into more details on the deeper theory behind the continuous cascade concept. To elucidate how these constructs look like I included some simulation examples in the following section.

Universal multifractal simulations

Generating continuous stochastic multifractals is not straightforward. Pecknold et al. (1993) propose a useful but unfortunately numerical unstable – for certain ranges of the

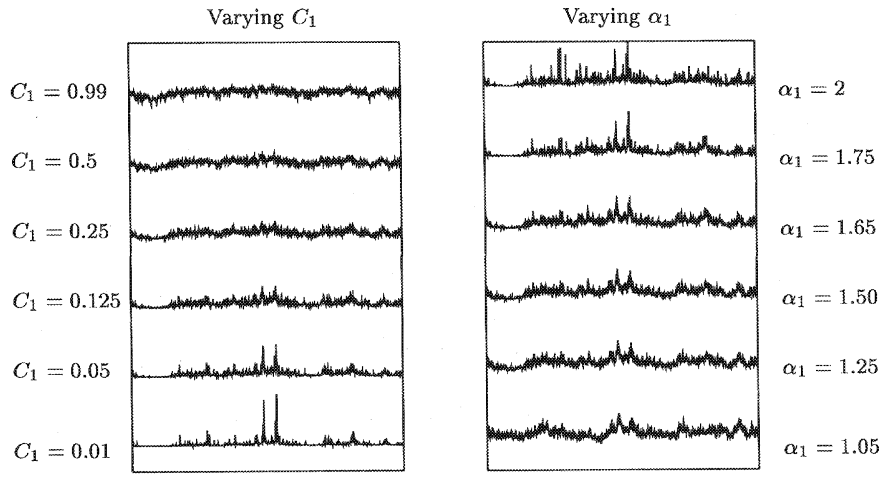

Figure 2.9 *Illustration of the dependence of universal multifractals on the parameters C_1 (left-hand side) and α_l (right-hand side). The values for C_1 are (from bottom to top) $C_1 = 0.99, 0.5, 0.25, 0.125, 0.05, 0.01$ with α_l fixed ($\alpha_l = 1.5$). It is clear that the C_1 determines the sparseness of the field, as C_1 becomes larger the simulations become more and more dominated by a small number of large outliers (singularities). The simulations for varying α_l $(2, 1.75, 1.65, 1.50, 1.25, 1.05)$ with C_1 fixed ($C_1 = 0.25$) display more deviations from the mean for decreasing α_l.*

multifractal parameters – technique to simulate continuous multifractal cascades with preset values for the α_l, C_1 and H.

On the first place the effect of fractional integration on a sequence of random numbers is reviewed in figure 2.8. The plot on the left depicts 6 different fractional integrations, for H increasing, from bottom to top, according to $H = 0$, 0.15, 0.25, 0.5, 0.75, 1, of a white Gaussian noise simulation and the same fractional integrations are displayed on the right-hand side for a universal multifractal with the parameters $C_1 = 0.25$ and $\alpha_l = 1.5$. All simulations consist of 16384 samples and are normalized to yield the same mean and variance. Inspection of these simulations shows that the parameter H rules the apparent "smoothness" of the simulations whereas the extension of the Gaussian process to a multifractal process clearly introduces intermittency in the realizations. The sensitivity of the simulations to changes in the other two ruling parameters are displayed in figure 2.9. On the left-hand side the different traces correspond to simulations with α_l fixed, $\alpha_l = 1.5$, and increasing C_1's, from bottom to top. As the values for the C_1 increase the simulations become more and more dominated by a few large outliers, the strong

singularities. On the right hand side the impact of variations in the other parameter α_l are reviewed. Again the values decrease from bottom to top and the simulations show a increasing number of deviations from the mean.

In summary it can be concluded that the generalization of the monofractal concept to processes which display a multifractal behaviour drastically influences the appearance of the simulations. In the presented simulations I was only able to show a limited number of realizations belonging to the universal multifractal family. However, the examples clearly indicate that the universal multifractals capture the problem of intermittency. Notice that this concept becomes particularly important when the distinctive features of reflectivity are concerned.

2.4.5 Scaling in relation to renormalization group equations

Renormalization group equations are those equations in physics that deal with scale dependence of certain phenomena (Wilson, 1983; Nottale, 1995, 1996). They accomplish this by prescribing how the physics at a large scale depends on what happens at a smaller scale. The way they appreciate this interrelation is by renormalizing the coarse-grained quantity. Within the cascade models a similar type of relationship is utilized. Only in that case it refers to the apparent interrelationship developing while going from the large scales to the smaller scales via some sort of fractalization process. In this view the Renormalization Group theory can be seen as a theory for an inverse fractal process, an observation made by Nottale (1992) and already remarked upon by Wilson (1983). Let me now explore these ideas a little bit further.

Following Nottale (1995) the renormalization group equations can, in their simplified form, be written as

$$\sigma \partial_\sigma f(\sigma, x) = g(f(\sigma, x)), \tag{2.36}$$

where the function $g(\cdot)$ performs the renormalization. Interpretation of the equation shows that the equation maps an infinitesimal dilated function onto itself. Via linearization of this equation, by expanding it into powers of f and neglecting all powers exceeding one, one obtains the simple relationship

$$\sigma \partial_\sigma f(\sigma, x) = C + \alpha f \tag{2.37}$$

with C being a constant, $C \neq 0$. Then it can be shown, following Nottale (1995) that there are different regimes pertaining to equation (2.37) ($\alpha > 0$) namely:

- $\sigma \gg \sigma_0$ where the scaling dominates with respect to constant C term and consequently this results in homogeneous scale-invariance for the solution[28]

$$f(\sigma, x) = (\frac{\sigma}{\sigma_0})^\alpha f(\sigma_0, x).$$

[28]The $=$ sign can be seen either in the sense of distributions or in the sense that it expresses equality in probability distributions, also denoted by $\stackrel{\mathrm{d}}{=}$.

- $\sigma \ll \sigma_0$ then $f(\sigma, x)$ becomes independent of σ.

In this the σ_0 is a reference scale depending on C and delineating the transition from scaling to non-scaling behaviour. For $\alpha < 0$ the opposite of the scaling behaviour is found.

This procedure demonstrates that one can obtain a scale invariant behaviour by solving a renormalization group equation of the type depicted in equation (2.37). This scale behaviour is, however, restricted to a scale range exceeding the transition scale σ_0. For $\sigma \ll \sigma_0$ the solutions becomes independent of the scale. This transition marks a break in the scaling and facilitates a separation of scales. By setting the constant C to zero, $C = 0$, the break is removed and one obtains a truly scale invariant solution.

As a matter of fact by setting $C = 0$ the problem becomes an eigenvalue problem pertaining to the scale derivative,

$$\sigma \partial_\sigma f = \alpha f \tag{2.38}$$

or alternatively to (see equation (2.10) too)

$$j\mathcal{C} f = (\alpha + \frac{1}{2}) f, \tag{2.39}$$

where the eigenfunctions f are of the general form, to be interpreted in the sense of distributions,

$$f(x) \propto x^\alpha \tag{2.40}$$

and where the scaling exponents α play the role of eigenvalues. It is this property that substantiates the initial *ansatz* to look for powerlaws when unraveling scaling complexity.

These deliberations clearly demonstrate that the solutions of equation (2.37) in the scaling regime or with C set to zero adhere to the scale-invariance – $f(\sigma x) = \sigma^\alpha f(x)$ with σ^α being the renormalization factor – satisfied by homogeneous distributions, see chapter 5. This type of behaviour can be extended to all abscissa via translations (Holschneider, 1995), eventually yielding a monofractal type of behaviour. In case of stochastic monofractals the scale-invariance becomes manifest in

$$f(\sigma x) \stackrel{\mathrm{d}}{=} \sigma^\alpha f(x) \tag{2.41}$$

which has to be interpreted as an equality for the probability distribution function pertaining to f and implying the following behaviour for the statistical moments

$$\langle |f(\sigma, \cdot)|^q \rangle = \sigma^{\alpha q} \langle |f(1, \cdot)|^q \rangle. \tag{2.42}$$

Up to this point I limited myself to describe constructs that display a homogeneous scaling behaviour, characterized by one single scaling exponent. The multifractal framework,

on the other hand, demonstrated that there exist alternative intermittent models that generate functionals containing whole hierarchies of different singularities. In that situation the scale-invariance becomes more intricate in the sense that it refers to a scaling exponent for each statistical moment implying that

$$\langle |f(\sigma,\cdot)|^q \rangle = \sigma^{\tau(q)} \langle |f(1,\cdot)|^q \rangle. \tag{2.43}$$

When dealing with a multifractal, the renormalization term, $\sigma^{\tau(q)}$, is not longer governed by a single scaling exponent α but by a series of exponents incorporated in the mass exponent function, $\tau(q)$, being non-linear in this situation. In terms of the partition function the above relationship comes down to

$$Z\{f,\psi\}(\sigma,q) = \sigma^{\tau(q)} Z\{f,\psi\}(1,q), \tag{2.44}$$

suggesting that the exponents $\tau(q)$ refer to eigenvalues pertaining to the following eigenvalue problem

$$-\sigma \partial_\sigma Z(\sigma,q) = \tau(q) Z(\sigma,q). \tag{2.45}$$

2.5 Mathematical and physical aspects of scaling

Now that the basic framework for conducting a multiscale analysis followed by a quantification has been laid down it is time to contemplate on the more mathematical aspects in a physical setting. On the first place the discussion will directed towards the selection of the mathematically proper analyzing wavelets. Then I will bring the notion of scaling in relation to the notion of differentiability[29]. I will pay specific attention on what the singularity spectrum and mass exponent function have to say on the global differentiability, integrability and rectifiability. The second main item of this section commits itself to say something on discretization and the mutual relation between the measuring device and analyzing wavelet. It is interesting to note that the octave pass-band filters utilized by the experimentalist – e.g. to anti-alias filter the data before discretization – find their exact counterpart in terms of wavelets. I will conclude this section by making an attempt to clarify the meaning of abstract mathematical concepts, such as differentiability, in the physical context. While doing so one immediately finds oneself confronted with the interference of the measuring device that effectuates an inevitable coarse-graining and in this way troubling one to issue rigorous mathematical statements. Despite this seeming difficulty I will express my personal opinion on why the mathematical concepts maintain their importance.

[29] This was already mentioned in the context of local scaling and Hölder exponent. The relevance of the singularity spectrum in this context will be treated here.

2.5.1 Pure mathematical considerations on multiscale analysis and quantification

Selection of the proper analyzing wavelet

From the mathematical point of view it is important that the analyzing wavelets being used adhere to the mathematical conditions that go with the proper mathematical definition of the continuous wavelet transform. Only then the multiscale analysis can be expected to work properly. The mathematical restriction to be imposed on the wavelet refer to a compliance with respect to the admissibility condition, see chapter 8, and to the invocation of a sufficient regularity. The first requirement is met by demanding a sufficient number of vanishing moments, M.

The two afore mentioned restrictions on the wavelet do not only serve an abstract mathematical purpose but, judged by the examples presented in chapter 8, may have a profound impact on the validity of the multiscale analysis and the quantification. Selecting improper wavelets in relation to the data, the object of study, yield erroneous results. This is because the wavelet's singularity detection range is restrained and depends on the regularity and the number of vanishing moments, M, pertaining to the wavelet. Unfortunately there does not exist a simple way out of this situation because one can jeopardize the effectiveness of the multiscale analysis by choosing a too regular wavelet and/or a wavelet with too many vanishing moments. Namely the smoother the wavelet the lesser its localization capabilities while the larger the number of vanishing moments the larger the number of extrema. Within the WTMML framework the localization aspect does not really give rise to a problem.

To be more specific, the number of vanishing moments constraints the maximum order of the detectable Hölder scaling exponent to $\alpha_{\max} < M$. The wavelet's regularity on the other hand restricts the accessible negative singularities to $\alpha_{\min} > -N$ where N denotes the wavelet's regularity, $\psi \in C_0^N$ (Bacry et al., 1993; Muzy et al., 1993).

In figure 2.10 it is shown what the relation is between the analyzing function and the observable range of singularities in the spectrum. From this schematic overview one can conclude that the Haar/box-car scaling function[30] is, as a consequence of its discontinuity and lacking of vanishing moments, only capable of detecting singularities in the range $-1 < \alpha < 0$. The variogram analyzing technique, the Poor Man's wavelet, being the "derivative" of the Haar scaling function, extends the observable detecting range to $0 < \alpha < 1$. But unfortunately this range is still limited due to the singular nature of the analyzing function itself. When within the analyzing technique, based on the WTMML's, a C_0^∞ wavelet with M vanishing moments is being used, then one is able to circumvent part of this problem because such a wavelet is able to detect singularities over the range[31] $-\infty < \alpha < M$. These observations are consistent with the examples discussed in chapter 8.

[30] Note that this is not a wavelet but a "smoothing" function.

[31] Note that for a discretized data set the infinite differentiability has to be replaced by a differentiability in the order of number of data points.

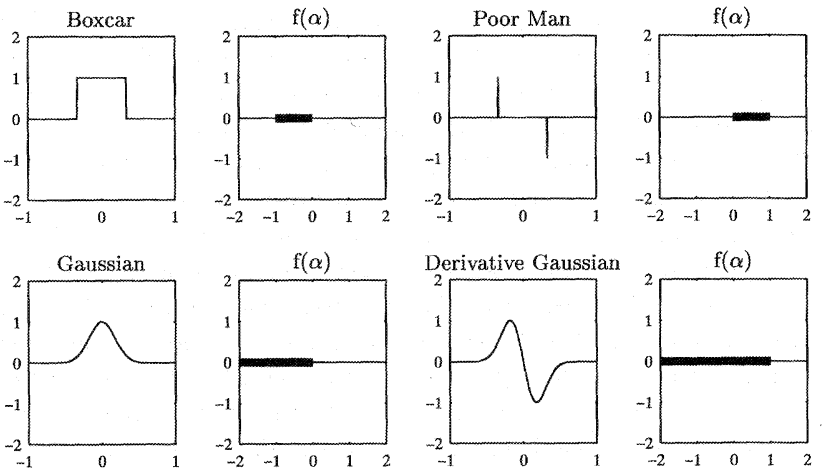

Figure 2.10 *The relation between the analyzing functions and the singularities that can be detected. The Haar scaling function – also known as the box-car or indicator function – has no vanishing moments, but a finite mean, so it is not able to detect singularity strengths larger than zero. Its smoothness is zero, hence it can not detect singularity strengths smaller than minus one. Its derivative – the Poor Man's wavelet – has one vanishing moment and a smoothness of minus one, enabling it to detect singularity strengths between zero and one. The Gaussian has no vanishing moments, but it is infinitely smooth, so it can detect singularities in the range* $(-\infty, 0)$. *Its first derivative is infinitely smooth too, but it has one vanishing moment, so it can detect singularities in the range* $(-\infty, 1)$.

Differentiability, integrability and rectifiability

It goes without saying that the intricate relationship between scaling and differentiability is very significant and deserves further exploration. From the local scaling analysis it is known that the Hölder exponents represent order of magnitude estimates for the local scaling and these exponents provide information on the local degree of differentiability. That is to say that a signal containing a singularity of strength $\alpha < 1$ is referred to as being non-differentiable at the location of the singularity while for $\alpha < 0$ the signal/function is referred to as singular. This notion can easily be extended to signals being non-differentiable/singular everywhere. For example Brownian motion being continuous everywhere but non-differentiable, $\alpha = \frac{1}{2}$, whereas white noise is singular and Hölder $\alpha = -\frac{1}{2}$ everywhere.

According to Nottale (1992) the apparent non-differentiable becomes manifest in a so-called *scale divergence*. This scale divergence refers to the following argument concerning

continuous but non-differentiable functions, i.e. Hölder regular functions with $0 < \alpha < 1$. For this purpose consider two points that lie on the curve delineated by the function $f(x)$. Call these points, with their respective coordinates, $A_0\{x_0, f(x_0)\}$ and $A_\Omega\{x_\Omega, f(x_\Omega)\}$. Since f is non-differentiable there exists a point $A_1\{x_1, f(x_1)\}$ with $x_0 < x_1 < x_\Omega$ that does not lie on the segment spanned by A_0A_Ω. This means that the total length becomes $l_1 = l(A_0A_1) + l(A_1A_\Omega)$ exceeding the distance between A_0 and A_Ω, i.e. $l_0 = l(A_0A_\Omega)$. Via iteration of this argument one can easily see that the total length diverges as the successive approximations of the non-differentiable function are being constructed. These approximations exist of line segments the length of which decreases for an increasing number of iterations. This length can be seen as the resolution or the gauge and the scale divergence refers to the divergence of the total length l as the resolution approaches zero, i.e. $l_n = l \to \infty$ as $n \to \infty$, a behaviour also known as non-rectifiability (le Méhauté, 1991). So whenever a function is continuous but non-differentiable it is non-rectifiable and the rate of the scale divergence is related to the degree of irregularity.

Let me now be more specific by posing the question on what can be learned mathematically from the quantities that characterized multifractals? First of all it can be stated that the mass exponent function $\tau(q)$ carries, for specific values of q, information on the integrability[32] or equivalently on the stochastic stationarity of the signal and is obtained via (Schertzer and Lovejoy, 1987a; Davis et al., 1994)

$$H \triangleq -\{\tau(q)\}_{q=1}. \tag{2.46}$$

The value of this exponent H refers to the degree of fractional integration which is related to the degree of non-conservation of the mean or to the degree of non-stationarity. This exponent has to be understood in the sense that when $H > 0$ the mean of a density depends on the scale,

$$\langle f(\sigma, \cdot) \rangle \propto \sigma^H,$$

hence the mean diverges as $\sigma \to \infty$.

Information on the global differentiability of f is carried by the Legendre transform of the $\tau(q)$ function, the singularity spectrum f(α). The abscissa for the endpoints of this spectrum provide the extrema for the estimates of the differentiability of the data being examined. The endpoint with the smallest value for α, $\alpha_{\min}$, refers to the strongest/wildest singularities whereas the abscissa for the other endpoint refers to the weakest/calmest singularities, i.e. the largest value for α, $\alpha_{\max}$. Needless to say the smaller the α the lesser f is differentiable, for $\alpha > 1$, the more non-differentiable for $0 < \alpha < 1$ and the more singular for $\alpha < 0$. The endpoints are defined (Parisi and Frisch, 1985; Collet, 1986;

[32] Of course it also contains information on the differentiability but that is easier revealed via the singularity spectrum that is found via the Legendre transform performed on the mass exponent function.

Siebesma, 1989; le Méhauté, 1991; Lichtenberg and Lieberman, 1992) as follows

$$\begin{aligned} \alpha_{\min} &\triangleq \{\partial_q \tau(q)\}_{q\to+\infty} \\ \alpha_{\max} &\triangleq \{\partial_q \tau(q)\}_{q\to-\infty}. \end{aligned} \tag{2.47}$$

These endpoints mark the interval of values taken up by the singularity exponents, i.e. $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, characterizing the singularities present in the data. The asymptotic definition of these endpoints can very well be reconciled with q being the selector for the singularities. The larger the q the smaller the scaling exponent which dominates the partition function.

Finally the singularity spectrum, $\mathrm{f}(\alpha)$, itself expresses the Hausdorff dimension to be associated with the singularities in the multifractal set that scale with a scaling exponent ranging from α to $\alpha + \mathrm{d}\alpha$. The singularity spectrum $\mathrm{f}(\alpha)$ expresses a probability, $\Pr(\sigma^\alpha < \check{f}(\sigma,\cdot) < \sigma^{\alpha+\mathrm{d}\alpha}) \propto \sigma^{-\mathrm{f}(\alpha)}$. Of the generalized fractal/Hausdorff dimensions the dimension in which the measure is concentrated is identified as an important one and reads

$$D_1 = \mathrm{f}(\alpha_1) = \alpha_1 \tag{2.48}$$

with

$$\alpha_1 = \{\partial_q \tau(q)\}_{q=1}. \tag{2.49}$$

This dimension is also known as the *information dimension* and is directly related (le Méhauté, 1991) to Shannon's entropy via

$$S_e(\sigma) = -\alpha_1 \ln \sigma. \tag{2.50}$$

Remark that this dimension is only properly defined for conservative fields, $H = 0$. If this is not the case one has to correct the α_1 for a shift by H. One can see that the information dimension is related to the C_1, the sparseness of the mean, which is just the co-dimension of the D_1, i.e.

$$D_1 = D - C_1, \tag{2.51}$$

with D being the dimension the embedding space. Finally one can designate a dimension to the set carrying the fractal, i.e the singular support of the fractal, and this one is defined by

$$D_0 = \mathrm{f}(\alpha_0) = \alpha_0 \tag{2.52}$$

with

$$\alpha_0 = \{\partial_q \tau(q)\}_{q=0}. \tag{2.53}$$

2.5.2 Physical aspects of measuring

Discretization

As already indicated in chapter 5, where I gave a preliminary discussion on distribution theory, a measurement is always the result of a non-linear averaging procedure, over a finite and non-vanishing interval, conducted on the physical quantity of interest. This observation also applies to the examples discussed so far. In fact all these examples concern discretized versions of the finite resolution measurement, i.e. a spatial bandpass filtering[33] and subsequent sampling[34] and optional digitization[35] yielding a discretized measurement of the spatial characteristics of the physical quantity. As a result of this procedure the measurement or functional is not known at every abscissa value for, say the spatial coordinate x, and its resolution will depend on the sample interval. This sample interval is inversely proportional to the Nyquist frequency delineating the high frequency cut-off of the bandpass filter characterizing the measurement device.

The spatial sampling procedure is mathematically defined by

$$\left(f_n\right)_{n\in\mathbb{Z}} = \int_{-\infty}^{+\infty} f(x)\phi(x - n\Delta x)\mathrm{d}x, \tag{2.54}$$

where ϕ is the appropriate kernel conducting the bandpass filtering, Δx the spatial sample interval and $\left(f_n\right)_{n\in\mathbb{Z}}$ the discretized approximation to $f(x)$. In the examples reviewed so far and in the remaining part of the thesis I took the liberty to refer to $\left(f_n\right)_{n\in\mathbb{Z}}$ as f or $f(x)$.

The inevitable constraint on the available resolution initiates the discussion whether the abstract concepts of mathematics, generally delimited by bounds at zero and/or infinity, make sense in the physical context where these limits are void of a meaning. In section 2.5.3 I will express my opinion on this important issue.

Measuring device versus the analyzing wavelet

When one comes to think of it there lies a deep analogy hidden between the selection of the proper analyzing wavelet and the choice of the appropriate measuring device. The underlying reason for this correspondence can be found in the shared property that taking measurements as well as conducting multiscale analysis involves the interaction of the known instrument/analyzing wavelet with the unknown signal. Hence one is during the selection process for the instruments/wavelet confronted with a lack of information on the true nature of the signal. From the examples I showed in chapter 8 it became very clear that an improper choice for the wavelet gives rise to erroneous results because the analysis became completely dominated by the analyzing function itself rather than

[33] Hopefully with a filter that is consistent with the regularity of the physical quantity of study. With other words the correct, "enough decades per octave", anti-aliasing has hopefully been used.

[34] This discretization has to be done in accordance to the so-called Nyquist sampling criterion.

[35] This digitization is required to represent the measured numbers digitally.

by the data. In practical circumstances the experimentalist runs the same risk since he finds himself confronted with a similar problem. His worries, for instance, may concern the selection of the proper anti-aliasing filter. These filters invoke a low-pass filtering, reducing the high frequencies in the signal, in preparation of a sampling. Perfect filters, filters with a an infinite sharp cut-off, can for practical reasons not be realized and one has to choose a filter with an adequate number of decibels per octave cut-off rate. The interesting point about this is that this decay rate is directly related to the regularity of the filter's impulse response. Hence the regularity conditions for wavelets find their direct counterpart in selecting the proper cut-off rate for the anti-aliasing filter.

The only way to check whether one made the correct choice in the above context is to fiddle around with the different types of filters/analyzing functions followed by a close inspection of the outcome. For the multiscale analysis this procedure boils down to an examination of the singularity spectrum for the possible occurrence of a phase transition while nearing the endpoints of the spectrum. A phase transition that emerges when approximating the $\alpha_{\max}$ endpoint can generally be attributed to an insufficient number of vanishing moments whether phase transition at the other side may point in the direction of insufficient regularity. In the cases where increasing the regularity and/or number of vanishing does remove the phase transitions the data itself may contain them.

The interesting thing about all this is that in practice people are in some cases taking more or less adequate measures. For instance in the field of geostatistics and atmospheric turbulence (Tartarskii, 1971; Yaglom, 1987; Schmitt et al., 1992; Schmitt, 1993; Davis et al., 1994) people use (generalized) structure functions also known as variograms instead of box-cars. This latter choice is equivalent to conducting an analysis with the poor man's wavelet having one vanishing moment and being singular by itself, see chapter 8. In this way the detection range shifted from $\alpha \in (-1,0)$ for the box-car to $\alpha \in (0,1)$ and thereby allowing for a sensible examination of stochastic processes with stationary increments. Notice, however, that this approach has the disadvantage that it is not equipped to detect singularities with negative scaling exponents. Obviously this due to the lack of regularity of the analyzing function. Summarizing care must always be taken while conduction measurements on processes containing singularity structures exceeding the detectable scale range.

2.5.3 Mathematical theory from a physical perspective

At this point the physicists under the readers may wonder what the physical significance is of the multiscale analysis in general and more specifically in the relation to differentiability. The reason for this question lies in the observation that without exception a physical measurement/observation represents a coarse-graining of the actual physical "reality". This inevitability seems to jeopardize – with the exception of the practical and fundamental criteria for the selection of the proper wavelets – all pure mathematical

arguments issued so far in relation to the multiscale analysis applied to actual physical measurements. The underlying reason for this dilemma lies in the fact that a mathematician delimits his statements with bounds at zero or infinity while these limits are physically speaking meaningless. For instance, the notion of differentiability, can within the mathematical context, straightforwardly be given a meaning for functions since they assign a value to a point. Physically speaking, however, the concept of a function is unmanageable because one is never able to resolve a quantity at a point! This is the prime reason why I am reluctant to consider, perhaps prematurely, a measured data set as a function. This is because the measurement process in itself can be seen a some sort of testing with a test function, an operation in line with a functional that assigns a value to a proper vector space of test functions. These functionals are able to host a much larger class of mathematical constructs amongst which the class of regular distributions that correspond to ordinary functions. Moreover they keep reference to the scale.

Now the point I want to make is that one has to tune the representation of choice for data depending on the data itself. In practical circumstances this comes down to choosing a representation that matches the data within the scale range of interest and that necessitates an interpretation of notions like differentiability in order to make them apply to physical data. For me such a physical interpretation is justifiable given the findings on what can go wrong in setting up a multiscale analysis as presented in chapter 8. Namely in those examples it became clear that when not adhering to strict mathematical conditions to be imposed on wavelets one gets an intermingling of the multiscale analysis with the data leading to erroneous results. Strictly speaking one might not have anticipated on such a notion because these examples inevitable referred to approximate versions of the actual mathematical objects.

For that reason I think it is legitimate to still talk about singular behaviour and non-differentiability in the context of finite resolution measurements. The explanation for this lies in the observation that the singular behaviour at the coarse scales may be interpreted as if it is preserved for the scale range untouched by the coarse-graining/smoothing operation. That is to say that an observer living in his characteristic scale still experiences the data as being non-differentiable or even singular even though his observation is inevitable coarse-grained. To be more precise the observer can not discern the data from being non-differentiable or singular within his scale realm. For example if one would ask an ant to draw a tangent line along the curbstone than the ant would be perfectly able to do so while for a human being – living at a larger scale – the curbstone acts as if it were discontinuous – one can stumble over it while the ant can not – withstanding one to draw the requested tangent line. However, when one sends the ant around the coast of England the ant would not only have also trouble to draw his tangent but would also have to cover a much longer distance than the human being would have to. Clearly the first example illustrates a separation of scales where the inner scale is smooth, when one considers the pave stone to be smoothly cut at some small length scale exceeding the char-

acteristic scale of the ant. The second example elucidated the seeming scale-invariance displayed by the coast of England, yielding an apparent non-rectifiability directly linked to its non-differentiability.

To summarize my statement is that the regularity assessment methods presented in this chapter preserve their meaning within the physical domain. I am aware that this line of reasoning can be full of pitfalls but I am afraid that physicists sometimes have to make choices. However, I think that the above argumentation fits well within the concepts of distribution theory where an instance of a Cauchy sequence can either be seen as an approximation to a generalized function or as a test function. For me the main consequence of this is that one requires a representation for the data that can in an appropriate manner be reconciled with the properties displayed by the data. With other words one must be very careful when selecting a representation for the data. One may even postulate that perhaps one is not "free" to choose a representation for the data at will!

For instance, in the continuum mechanics it is perfectly suitable to represent the constitutive parameters as smooth functions despite the fact that such a representation is certainly not defendable at scales pertaining to the Quantum Mechanics. Alternatively speaking one can make similar remarks with respect to the constitutive parameters that display a singular behaviour, in the preceding interpretation, within the scale range the dynamic interactions are believed to take place. This observation forms the key of the discussion in the epilogue. In the next section I try to elucidate the point made in this section via an example of scaling where the transition singular smooth occurs. But before doing that I like to endorse to the reader the opinion of 't Hooft (1994) who asks himself "*How do I interpret my mathematical equations?*". Furthermore I wholeheartedly subscribe the remark made by le Méhauté (1991) who refers to a scale divergence as: "*It is of course the way in which this upper bound*[36] *approaches infinity that determines the fractal dimension*, a property we shall meet when studying problems in physics."

An illustrative example expressing the relativity of regularity

In this section I will try to convey the inherent relativity linked to the concept of regularity or smoothness. For example, everybody would, on first sight, agree with the observation that the box-car in figure 2.11 is not smooth, non-differentiable and even discontinuous, whereas the Gaussian's appearance seems to be very smooth, C^{∞}. The careful observer, however, could have asked the critical question to which resolution the pronouncement of smoothness refers. To clarify this point let me submit the profiles of figure 2.11 to the local multiscale analysis presented in section 2.4.1. For the small scale range, being limited by the effective support of the two profiles, it is indeed clear that the box-car contains two discontinuities, judged by the estimates for the local Hölder exponents over this scale range. But when the scale range of observation is increased beyond this scale range a complete opposite behaviour becomes apparent for the box-car as well as for the

[36] Here le Méhauté (1991) makes reference to the divergence for the slope of the line joining $\{x_0, f(x_0)\}$ and $\{x_1, f(x_1)\}$ as $x_0 \to x_1$.

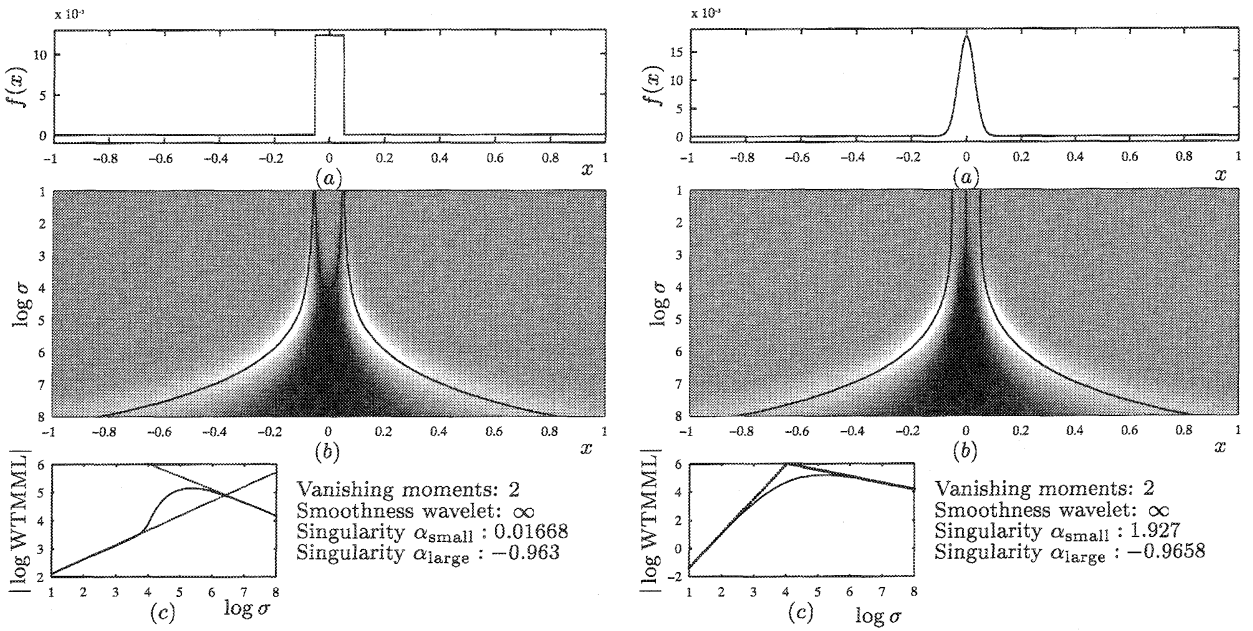

Figure 2.11 *In this figure a Gaussian with an effective scale* $\log \sigma = 4$ *is analyzed together with a box-car with the same support. Clearly the Gaussian and box-car display a deviating behaviour for the small scales but as soon as the scale of observation exceeds* $\log \sigma = 4$ *they display the same behaviour. This notion can be understood by taking into consideration that both functions are members of a delta convergent sequence. Finally notice that the Hölder exponent for the small scales yielded by the Gaussian bell-shape approximately equals the number of vanishing moments of the analyzing wavelet.*

Gaussian bell-shape when $\sigma \to \infty$. For this case namely, the Hölder exponent estimates correspond to the behaviour yielded by the δ-distribution. Especially for the Gaussian this may appear to be strange because one could have the impression that it is smooth; a property that certainly can not be associated with a δ-distribution. But on the other hand it is quite well understandable because both the box-car and Gaussian are functionals adhering to the conditions imposed on functionals generating a delta convergent sequence (Gel'fand and Shilov, 1964), see also chapter 5. This means that they both constitute an approximation to the δ-distribution, an observation in agreement with the presented regularity analysis. Summarizing: *when issuing statements about the smoothness of a measurement one has to include a scale of reference to which the pronouncement has to apply.*

At this point I may have left some of the readers puzzled. This is because the Gaussian is, of course, in a strict sense an infinitely differentiable function. But one might wonder whether the concept of a function, a rule locally assigning a value to a point, is a comprehensive concept given the inevitable limitation on the measurement's resolution? So, is it not better to use the concept of a functional? A functional is a rule assigning a value to each function in a set of so-called test functions of non-vanishing support (Zemanian, 1965). Here the ambiguity has been built in by associating the non-vanishing aspect of the test function's support to the boundedness of the measurement's device resolution. In this way a measurement, a functional, will only be an approximation to a possibly singular function, a function containing a singularity. That explains why the Gaussian is smooth for the small scale range but indiscernible from a singular non-differentiable function for the larger scales. This is because the Gaussian can be seen as an approximating measurement of the δ-distribution, with an instrument the response of which is given by the Gaussian.

The apparent ambiguity that goes with the interpretation of mathematical constructs in the physical framework could have been circumvented when representing the data in the space-scale plane. In that case, namely, one could easily have recognized the transition occurring in the scaling behaviour. For this reason Nottale (1992) proposed a representation that includes a reference to the scale by considering the functional/data itself together with all its coarse-grained/smoothed versions. My ideas will go along similar lines but instead I opt to use a wavelet representation for the functional/data and this will be the subject matter of the next section.

2.6 A scaling medium representation

The main purpose of this chapter is to define a mathematically consistent and physical feasible representation for the constitutive "parameters"[37] ruling the wave motion and

[37] I used the " " to anticipate to the observation that these parameters can not longer be considered as parameters when displaying a wild and irregular behaviour as compared to the dynamic scale range of the wave motion.

that display a highly irregular behaviour. The two constitutive parameters being of specific interest here are the density of mass, ρ, and the compressibility, κ. The derived quantity delineating the local compressional wave speed, c_p, is of interest as well. The profiles for these quantities are all obtained from real well-log measurements. These well-log measurements comprise of "*in situ*" measurements acquired along the vertical direction, the borehole, and extend over a distance of approximately 3 km. In the sequel of this section I will demonstrate that these well-log measurements are beautiful examples of highly irregular processes, the scaling behaviour of which can be very nicely captured by the quantification tools proposed earlier in this chapter.

In anticipation of the actual outcome of the multiscale analysis set to work on well-data I have to make precautions concerning the actual representation I am going to use to represent this type of data in a proper way. I am sure that the discussion on the implication of taking the wrong measures in the context of a multiscale analysis substantiated the importance of selecting the proper representation. To cut things short the continuous wavelet transform offers an excellent way of representing highly irregular data sets in a responsible way. Such an approach not only facilitates an effective partitioning by means of the wavelet transform modulus maxima, the WTMML, but also allows for a proper definition for the order of magnitude estimates assessing complexity both physically and mathematically.

The first feature, the effective partitioning, paves the way for the introduction of scaling versions of macro/meso models that share the "sparseness", think of blocky versions of the macro/meso models, that display the proper scaling behaviour over a wide scale range. The second feature is of equal importance because it refers to a mathematical classification being an indispensable prerequisite when one has certain mathematical manipulations on the data in mind. Within the context of the solution methods for the wave equation in its current formulation[38] these manipulations comprise differentiation, integration and multiplications[39]. In the sequel I will show that the operations of differentiation and integration can be given a meaning even though for cases where these operations are strictly speaking ***not*** not defined. This proposed procedure defines these operations in a *weak sense* and that goes at the expense of introducing a scale dependence. For the multiplication operation carrying, for example, the interaction between the medium and the wavefield, there possibly remains a problem when singularities emerge in the constitutive parameters. In the epilogue I will pay attention to the question of the added value of the multiscale analysis in relation to the current wave theory in complex media.

Bringing the above observations on the proper representation into the perspective on how

[38] Acoustic wave motion is ruled by a second order hyperbolic system of two coupled first order partial differential equations containing temporal, spatial derivatives and multiplications by the constitutive parameters.

[39] Or other type of analytical operations, e.g. divisions.

people treat erratic data sets such as well-data one may draw the conclusion that these notions are generally not contemplated in the sense that the discretized well measurements are treated as discretized functions and being manipulated as such. The reader may suspect that the representation I am about to introduce claims to provide the correct framework. The setup of this section is as follows. First I will introduce the representation. Then I discus its merits followed by applying it to real well data. The application in itself is intended to demonstrate that well-logs indeed display a behaviour that is within a certain scale range comparable with the highly erratic behaviour yielded by the singular models reviewed in section 2.4.3. I will conclude this section by proposing definitions for a generalized multiscale reflectors and a generalized multiscale version for the concept of the macro/meso model.

2.6.1 The representation

What do I intend to accomplish when proposing a scaling medium representation? The answer lies in the conjecture I presented in the introduction where I hold, on empirical phenomenological grounds, two types of averaging processes responsible for the behaviour of wave interactions in complex media such as the earth's subsurface. These averaging processes comprise of

- a *local* averaging process linked to a local scattering mechanism taking its effect over the gauge of the wavefield. It is expected that the reflection dynamics are related to this local type of averaging. Note, however, that the definition of the latter gauge forms one of the most intricate questions (le Méhauté, 1991; Nottale, 1992; le Méhauté, 1995; Nottale, 1995) in this area of research.

- a *global* spatial *self-averaging* over a gauge proportional to the propagation distance. It is expected that this averaging process is responsible for the apparent static *anisotropy* and dynamic *dispersion* induced by the irregular heterogeneity displayed by the medium being probed. The self-averaging refers to the notion that the coherent part of the propagating wavefield, not the coda, is relatively insensitive to the details of the local features it encountered on its propagation path. With other words some "law of large numbers" might apply.

Given these two physical constraints is it now possible to come up with a representation that

- only makes sense when a separation of scales seems futile.

- is mathematical consistent in the sense that it regularizes the *infrared* and *ultraviolet* catastrophes, that is to say it is able to deal with the apparent non-integrability and non-differentiability displayed by the data.

- provides a representation that captures the complexity and that allows for a reconstruction/inversion.
- features a sparse representation by means of an effective partitioning.
- features an effective quantification that on the one hand provides an effective parameterization while on the other hand it supplies a mathematical characterization.

Taking all these considerations in retrospective it is evident that the continuous wavelet transform constitutes the suitable candidate to do the job. Simply speaking the major argument favouring this choice can be reduced to the observation that a separation of scales is futile for well measurements within the scale range of interest. This observation is substantiated judged by the displayed behaviour in figure 2.1, where I included the consecutive smoothings and details yielded by a real well-log Moreover, the multiscale analysis results to be presented later in this chapter shows there is ***no*** evidence of a ***break/transition*** in the scaling. The well-data maintain their typical scaling behaviour across the spatial scale range believed to be inhabited by the seismic wave interactions. Remark, however, that in my opinion it is difficult, with the current wave theory, to come up with a specific estimate for this scale range in the sense that it is difficult to define the spatial gauge. The reason for this dilemma is, in my opinion, to be found in the observation that the main candidate, the apparent spatial wavelength, for the spatial gauge to be assigned to the wavefield in itself depends on the local velocity while this same velocity depends on the local scale/gauge as well! See the epilogue for a discussion on this interesting issue.

Separation of scales

Let me for the time being assume the seismic scale range to be proportional to the dominant spatial wavelength. Than it is possible to come up with a criterion that substantiates whether a separation of scale is feasible or not. Given the continuous wavelet transform it is easy to come up with such a criterion because this transform computes the scale derivative and therefore expresses the change in the signal, f, as the result of an infinitesimal scale change. Evidently, in those cases where the scale derivative is small, i.e. where

$$\mathcal{W}\{f,\psi\}(\sigma,x)|_{\sigma\in[\sigma_1,\sigma_2]} \leq C, \tag{2.55}$$

with C some sort of threshold being small or zero, one can neglect contributions from that particular scale range $\sigma \in [\sigma_1, \sigma_2]$. So, for those cases where the criterion of equation (2.55) holds one is allowed to invoke the separation of scales because, apparently, there are two distinct scale regimes namely, $\sigma \in [\sigma_{in}, \sigma_1]$ and $\sigma \in [\sigma_2, \sigma_{out}]$ with σ_{in} and σ_{out} referring to the inner and outer scale respectively, that may be treated separately. However, when one is unable to locate an interval where the scale derivative is negligible small it is pointless to advocate a separation. In that case the *trend* – the coarse-grained

part of f, containing information over the scale range $\sigma \in [\sigma_{sep}, \sigma_{out}]$ with $\sigma_{sep} \in [\sigma_1, \sigma_2]$ delineating the scale at which the separation is effectuated – and the remaining *detail* would drastically change as a function of the scale of separation, σ_{sep}. In that respect the trend is not being insensitive to scale changes, an observation in sharp contrast with the concept of a trend that is generally considered as being insensitive to scale changes. With other words the whole issue of defining a macro/meso model[40] is that it reflects an anticipated independency with respect to the scale, a notion that can not be reconciled with the apparent scaling. Finally notice that the criterion laid down in equation (2.55) bears a relationship with the criterion set by homogenization theory that determines whether the method of homogenization is expected to hold. Following Auriault (1991) this criterion reads

$$\frac{l}{L} = \varepsilon \ll 1, \tag{2.56}$$

with L denoting the macroscopic length scale of the excitation and l the characteristic length scale of the elementary representative volume. This ratio refers to the small parameter facilitating an asymptotic expansion. In the language of multiscale analysis this type of asymptotic approximation would be valid when the ratio of the scale indicator endpoints σ_1 and σ_2, $\frac{\sigma_1}{\sigma_2}$ becomes small. In the epilogue I will pay more attention to the important issue of a possible separation of scales.

Representation by the continuous wavelet transform

Judged by the behaviour displayed by the continuous wavelet transform conducted on a real well-log measurement as depicted in figures 2.1 and 2.15 it is clear that the criteria of equations (2.55) and (2.56) are almost certainly *not* met. The well data contains structures of all sizes and a separation of scales seems futile. Therefore one has to come up with a representation truly supporting the notion of scale.

In sections 2.2 and 2.3 of this chapter and in chapter 8 I showed that the continuous wavelet transform exactly offers such a representation for which also an *inverse* exists when the wavelets adhere to the *admissibility condition*. To be more specific the representation I opt for reads

$$\check{f}(\sigma, x) = \mathcal{W}\{f, \psi\}(\sigma, x) = \langle f, \psi_{\sigma,x} \rangle \tag{2.57}$$

and exactly corresponds to the continuous wavelet transform. Here the symbol f refers to the medium properties. In the sequel I will browse through a series of arguments that favour this choice of representation in terms of the wavelet coefficients.

A proper mathematical representation

Perhaps the most important feature favouring the above choice lies in the ability of the wavelet transform to "absorb" divergences occurring in f. This regularization is

[40] Here I mean the definition of a macro/meso model from well-log data, not the estimation of a macro/meso velocity model from the seismic data.

effectuated by either selecting a sufficient regular wavelet counter balancing the possible *ultraviolet divergence*, $|F(k)| \to \infty$ as $k \to \infty$, for the spectrum and caused by the occurrence of negative singularities or by setting the number of vanishing moments in such a way that the *infrared divergence*, $|F(k)| \to \infty$ as $k \downarrow 0$, is regularized in the sense of Hadamard's finite part, see chapter 5. This latter divergence is caused by an apparent non-integrability.

To put it differently the singularities in the data act as *fractional differentiators*, for $\alpha < -1$, or as *fractional integrators* for $\alpha > 0$, and the analyzing wavelet has to be set to undergo the action of these singularities. Now the interesting point is that the wavelet transform undergoes this action in exactly the same fashion as is being done in distribution theory. There the singular behavior of f is carried over to the appropriate test function via the method of testing. Therefore it is fair to say that within the wavelet theory $\check{f}$ can be considered to be a *functional*, a rule that assigns a number to a proper vector space of test functions, rather than a ordinary function.

Suppose the singularities in the signal, f, range over the interval $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ with $\alpha_{\min} < -1$ and $\alpha_{\max} > 0$. Then the wavelet will cast the singularities that lie outside the interval $-1 < \alpha < 0$ towards the inside of this interval while the wavelet undergoes the action that is attributed to invoking the shift, i.e.

$$\langle f, \psi_\sigma \rangle = \sigma^m \langle f_{reg}, \mathcal{I}^m \psi_\sigma \rangle. \tag{2.58}$$

The value for m are those values that implement the map, $\alpha \mapsto \alpha' = \alpha - m$ with $\alpha \in (-1, 0)$ and where m is found by taking the upper entier of α, $m = \lceil \alpha \rceil$ for all singularities α in f. The $\mathcal{I}^m$ denotes the m^{th}-order derivative operator for $m < 0$, the identity operator for $m = 0$ and the m^{th}-order primitive operator for $m > 0$. Obviously as a consequence of the above action the wavelet has to comply to sufficient regularity conditions as well as integrability conditions depending on the minimum and maximum values of m and the result will be that f_{reg} only contains singularities, ranging from -1 to 0, that can be accommodated by the inner product and by the Fourier transform. The reader interested in more technical detail is referred to chapters 5 and 8. To conclude I like to remark that, to the author's opinion, the just described procedure constitutes the underlying principle for detecting events in signals and coincides with the way in which, for instance, humans visually detect events (Marr, 1982).

Reconstruction

By virtue of selecting wavelets that adhere to the admissibility condition one can reconstruct – via the resolution of identity – the original functional via

$$f(x) = \int_{-\infty}^{+\infty} \int_0^{\infty} \check{f}(\sigma, x') \psi(\frac{x - x'}{\sigma}) \mathrm{d}x' \frac{\mathrm{d}\sigma}{\sigma}, \tag{2.59}$$

which has to be interpreted in the sense of distributions. Here the symbol f refers to an ordinary function as well as to a generalized function, a possibly singular functional.

Regularization by the continuous wavelet transform

In those cases where a certain smoothness is required or where f contains singularities which are Hölder α with $\alpha < -1$, one has to regularize the functional f with a test function. Such a procedure exactly corresponds to only partially reconstructing f via equation (2.59), i.e. via

$$f(\sigma, x) = \int_{-\infty}^{+\infty} \int_{\sigma}^{\sigma_{out}} \check{f}(\sigma', x')\psi(\frac{x - x'}{\sigma'})\mathrm{d}x' \frac{\mathrm{d}\sigma'}{\sigma'}, \tag{2.60}$$

with f representing the bare/undressed physical quantity with a finite outer scale, $\sigma_{out} < \infty$, to underline the physical situation. It can be demonstrated that equation (2.60) implies a *regularization*, see chapter 5, by a test function (Zemanian, 1965), an operation defined by

$$f(\sigma, x) = (f * \phi_\sigma)(x) \qquad \sigma > 0, \tag{2.61}$$

with ϕ_σ being a smoothing kernel leaving the DC component untouched. The fact that this equation can be regarded as a regularization is important because the process of regularization by a test function (Zemanian, 1965) affirms a certain smoothness/differentiability on the quantity $f(\sigma, x)$.

It is not hard to reconcile these two equations (2.60) and (2.61) since they can shown to be equivalent formulations. To understand this remember that the smoothed function, $f(\sigma, x)$ contains all details *up to* the scale σ while the reconstruction integral of equation (2.60) runs over the details containing the information *at* the scales ranging from σ to σ_{out}. It has to be remarked, however, that there is an important difference between these two types of regularizations. By way of its construction equation (2.60) is able to host a larger class of objects that are not necessarily integrable. As demonstrated this non-integrability is absorbed by the analyzing wavelet a notion that can not be attributed to equation (2.61) where specific integrability conditions have to be imposed on f, see theorem 5.1. A distributional interpretation of this equation, however, boils down to an equivalent formulation as laid down in equation (2.60).

The final point I like to make is that the index of the test functions used to define operations such as differentiation and integration in a weak sense, by means of Cauchy sequences, finds its direct equivalent in the scale index. So the way of thinking behind wavelet theory is very similar to the founding concepts of distribution theory where, for example, the Dirac delta distribution is given a meaning in the sense of a delta convergent sequence (Schwartz, 1957; Gel'fand and Shilov, 1964). This sequence is a Cauchy sequence of test functions, read wavelets, of decreasing support and this type of construction is very similar to the *resolution of identity*, see chapter 8, that comes down to the same thing.

Differentiation and integration

Let me in this subsection make some more specific remarks on the question how to differentiate or integrate functionals that are non-differentiable or non-integrable or alternatively that can not be discerned from being non-differentiable or non-integrable within a certain scale range. Provided with the representation in the wavelet domain the answer is simple: one just has to impose *one* additional vanishing moment in case one wants to integrate, to the first order, or to increase the regularity by *one* in case one wants to differentiate to the first order. In this way the *partially* reconstructed function, $f(\sigma, x)$, becomes integrable or differentiable but this goes at the expense of a insertion of a explicit scale dependence.

Finally remark that these measures, mathematically speaking implying *regularizations*, exactly correspond to the representation proposed by Nottale (1995). The prime reason why Nottale (1995) introduces such a formulation is that it affirms the invocation of a certain smoothness making $f(\sigma, x)$ differentiable for the scales smaller than σ under the condition that $\sigma > 0$ while $f(\sigma, x)$ maintains its irregularity for the scales exceeding σ. Then Nottale (1995) continues to consider the scale derivative of the smoothed quantity $f(\sigma, x)$ and that exactly corresponds to the continuous wavelet transform.

Sparse representation by the WTMML

The wavelet transform modulus maxima formalism offers the necessary efficient partitioning of the space-scale plane. This partitioning is based on the property that the singularities in f are directly reflected into the behaviour for the extrema of the wavelet coefficients. Moreover Mallat and Hwang (1992) showed that the representation in terms of the WTMML can be inverted although not uniquely and at the expense of some accuracy.

2.6.2 Application to well-log data

In the previous section I stepped through the main arguments and properties substantiating the choice of using the continuous wavelet transform domain as the multiscale representation. It is now time to reflect on this choice a little bit more in the context of applying these ideas to real well-log measurements. I will start by showing that the "conventional" quantities – the *variogram* and *power spectrum*, reflecting the second order two point statistics – display a power law type of behaviour. As input for the analysis I took the compressional wave speed profile I depicted already in figure 2.1. Despite the hiatuses inherent in these methods, see the preceding discussion, one can find estimates for the slopes of these quantities that provide information on the degree of fractional integration H. I continue the discussion by applying the proposed multiscale analysis and quantification to the same well-data. This concerns an evaluation of the local analysis of the scaling as well as an unraveling of the global scaling all conducted within the sparse

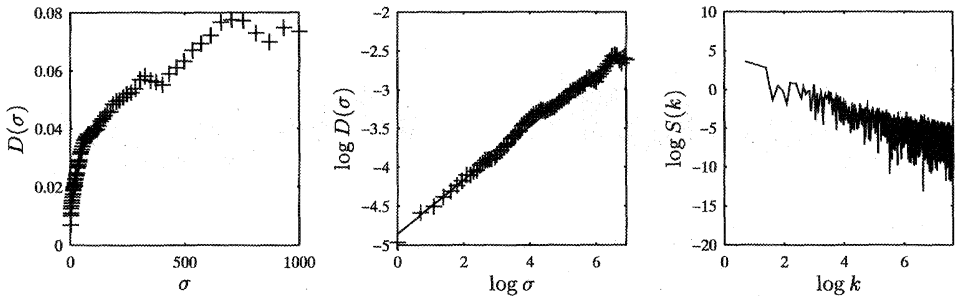

Figure 2.12 *Illustration of the power law behaviour displayed by the structure function (middle plot which is a* log-log *version of the left plot, the structure function) and the power spectrum (right), obtained from a real well-log measurement. The* log-log *plots are depicted together with linear fits the slope of which is related to the monofractal scaling exponent H.*

partitioning provided by the WTMML. Then I will show that the universal multifractal cascade model not only offers an effective parameterization for the singularity structure but also a way to generate synthetic well-data, the statistical properties of which display a singularity structure that bears a great similarity with the evidenced scaling complexity yielded by the well-data.

A conventional "scaling" look at a well-log

The most simplified models within the scaling concept are formed by monofractals, e.g. H-indexed fractional Brownian motions (Mandelbrot and Wallis, 1969; Tartarskii, 1971; Yaglom, 1987; Walden and Hosken, 1985; Todoeschuck and Jensen, 1989; Herrmann, 1991; Herrmann and Wapenaar, 1992, 1993, 1994) or non-conservative densities, the behaviour of which is quantified by the scaling exponent H expressing the degree of fractional integration. The scaling for these type of processes entails a powerlaw behaviour for the second order statistics as expressed in the following relations for the power density spectrum, see figure 2.12 on the right,

$$S(k) \sim \frac{1}{k^{\beta}}, \tag{2.62}$$

where β refers to the slope of the power spectrum, $S(k)$, with $\beta = 2H + 1$, and for the structure function or variogram, see figure 2.12, left and middle,

$$D(\sigma) = \langle |f(x+\sigma) - f(x)|^2 \rangle \sim \sigma^{2H}. \tag{2.63}$$

Despite the fact that this approach has been successful in the characterization for the *second* order statistics, $q = 2$, of the medium fluctuations, it fails to give an adequate

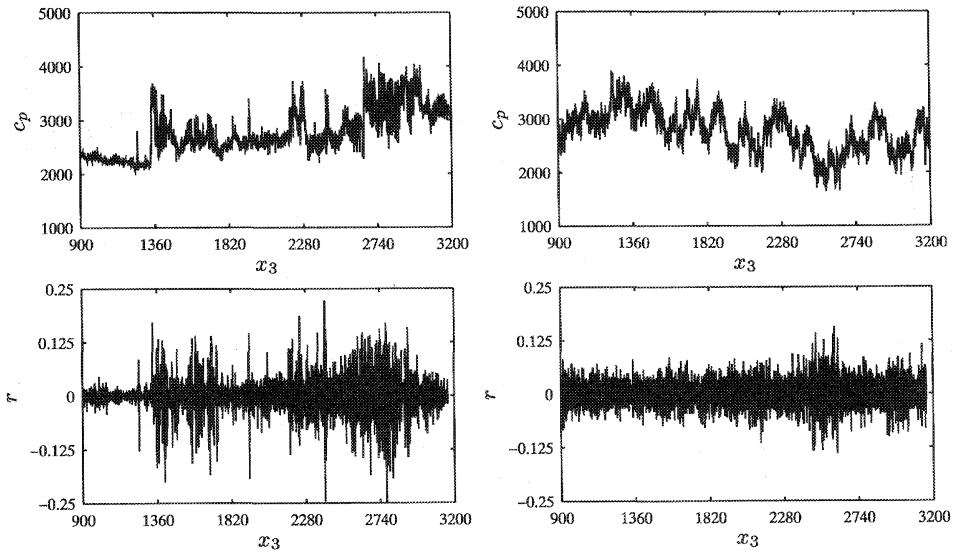

Figure 2.13 *Comparison between the compressional wave speeds and acoustic normal incident reflection coefficients yielded by a real well-log measurement (left column) and their monofractal counterparts (right column). The monofractal parameters were set according to estimates obtained from the well-log measurement. Notice the absence of outliers on the right.*

description for the properties of the higher order statistics. This shortcoming manifests itself particularly when highly intermittent signals[41], signals with distinct *active* bursts and *passive* regions and large *outliers*, are concerned. The degree of intermittency expresses the ratio between the relative *active* and *passive* regions in a signal and unfortunately monofractals, such as fBm, are not well equipped to effectively capture this property. This observation is illustrated in figure 2.13, where a comparison is made between plots for the compressional wave speeds and the corresponding reflection coefficients of a real well-log measurement, left column, and similar plots for a synthetic monofractal simulation, right column. These acoustic normal incidence reflection coefficients were computed using the definition for the reflection coefficients at a discontinuity obtained by supplementing boundary conditions on the pressure and normal component of the particle velocity. The reflection coefficient for the pressure and for a constant density medium reads $r = (c_p^l - c_p^u)/(c_p^l + c_p^u)$ where c_p^l and c_p^u denote the compressional wave speeds in the lower and upper layer. The simulation was obtained after running a monofractal random generator with its parameters set according to estimates for the

[41] Well-logs are amongst them.

monofractal parameters, the variance and slope of the power spectrum $\beta = 2H + 1$. It is clear, especially for the reflection coefficients, that the monofractal characterization misses the large outliers, the large "macro" reflectors. In order to overcome this apparent deviancy I need to come up with a more general approach, the one I proposed in the preceding part of this chapter.

2.6.3 A local representation and quantification by the WTMML

Provided with the WTMML partitioning I am able to locate and quantify incipient singularities via zooms guided along the WTMML. In figure 2.14 I included an example of the local multiscale analysis and quantification procedure set to work on the well-log measurements I already depicted in figure 2.1. As expected the WTMML's emanate at the abscissa where the well-log behaves singular or singular in its derivatives. The amplitudes along the WTMML's provide information on the local scaling of the singular regions and is used as input to the local Hölder exponent estimation scheme. This latter action leads to an effective linearized parameterization for the scaling behaviour to be associated with the abscissa where the WTMML's emerge. Moreover, it also leads to a mathematical characterization of the measurement's regularity, differentiability, within the prescribed scale range.

The well-log measurement under consideration is a measurement of the compressional wave speed. It was measured at a resolution of 15 cm. Afterwards it was smoothed to a resolution of 60 cm to remove the part of the spectrum contaminated by the well-logging tool. After that I took a subset of the measurement to obtain a convenient data set of 4096 samples. It occupies about 2.5 km of data. The analysis was done with the second derivative of a Gaussian, so with a wavelet which is C^∞ and which has two vanishing moments. In figure 2.14 I included an example of the proposed local multiscale representation and quantification/parameterization applied to the well-log. One can very nicely see that there are indeed "regions" the singularity strength of which varies significantly, an observation consistent with a multifractal behaviour. Finally, remark that the original signal can be reconstructed (Mallat and Hwang, 1992). Alternatively one can also try to reconstruct the signal when only information over a limited scale range is available, i.e. in the seismic scale range. This opens a way to define a multiscale representation, a generalized scaling macro model, characterizing the edges and allowing for a reconstruction. The reader who is interested in a explorative survey of using the WTMML method to conduct multiscale analysis on seismic data is referred to Staal (1995), Hoekstra (1996) and Dessing (1997).

A global multiscale representation and parameterization

As an initial *ansatz* I conjectured the effects of wave propagation, induced by heterogeneity, to be ruled by some sort of spatial self-averaging process over the propagation distance. This prompted me to come up with a global multiscale analysis scheme that

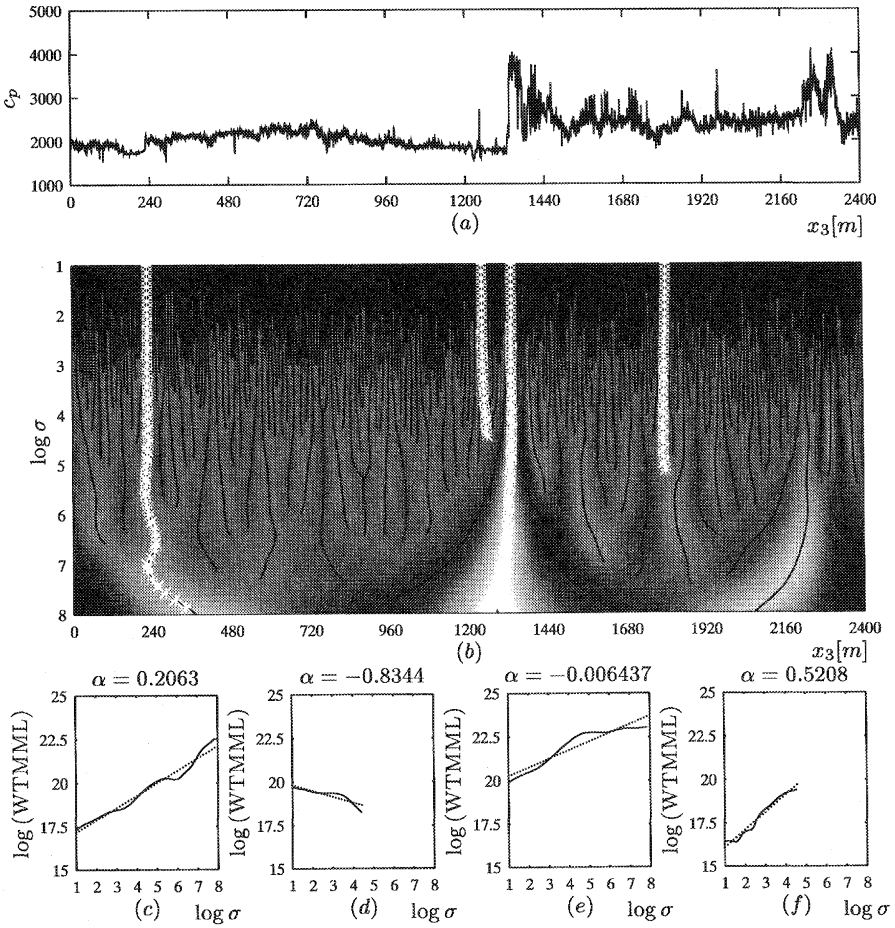

Figure 2.14 *Example of the local analysis and subsequent parameterization conducted on a real well-log measurement. The sample interval is* 60 cm *and the total length of the data set is* 2.4 km. *On the top is the well-log displayed, i.e. a compressional wave speed profile. In the middle the* WTMML*'s are displayed. On the bottom it is shown how to estimate the local Hölder exponents for a selected number of* WTMML*'s. Notice that positive, edge like, as well as negative, needle like, singularities occur.*

- captures the scaling of the statistical moments.
- unravels the singularity structure
- determines the generalized fractal dimensions.
- assesses the global differentiability and integrability.

Given the outcome of such a global multiscale analysis, conducted on real well-data, one is set to determine the degree of non-stationarity, H, the amount of intermittency, C_1 and the universal multifractal parameter α_l from the mass exponent function and singularity spectrum. These parameters not only provide a classification for the type of multifractal best suited to describe well-data but also offer a way to generate synthetic data sets with preset multifractal properties.

Computation of the global multiscale analysis

In figure 2.15 I included an example of the partitioning of a real well-log, figure 2.15 (a), by the continuous wavelet transform modulus maxima lines, figure 2.15 (b). I calculated the corresponding partition function $Z_\sigma\{\cdot,\psi\}(q)$, depicted in figure 2.15 (c), with the wavelet's scale ranging from 5 m to 160 m. The mass exponent function $\tau(q)$ is plotted in figure 2.15 (d) with the $q \in [-10, 12]$ while the f(α) spectrum is on display in figure 2.15 (e). Figure 2.15 clearly demonstrates that within the scale range 5 m to 160 m the well-log reveals a hierarchy of singularities with different scaling exponents.

The estimation of the weakest singularity, describing the maximum smoothness, shows that, cf. equation (2.47),

$$\alpha_{\max} \simeq 1.6.$$

The strongest singularity, describing the minimum smoothness is found to be, cf. equation (2.47),

$$\alpha_{\min} \simeq -0.18,$$

and the integrability is given by, cf. equation (2.46),

$$H \simeq 0.21.$$

These estimates show clearly that in the seismic scale range the well-log measurement can not be distinguished from a multifractal. This multifractal contains singularities with scaling exponents that correspond to singular "points", $\alpha < 0$, non-differentiable but Hölder continuous "points" $0 < \alpha < 1$ and "points that are singular in their higher derivatives. The value for the H refers to a non-conservation of the mean and that corresponds to a stochastic non-stationarity. This latter property entails a non-integrability that becomes evident in the notion that the mean and/or variance diverge(s).

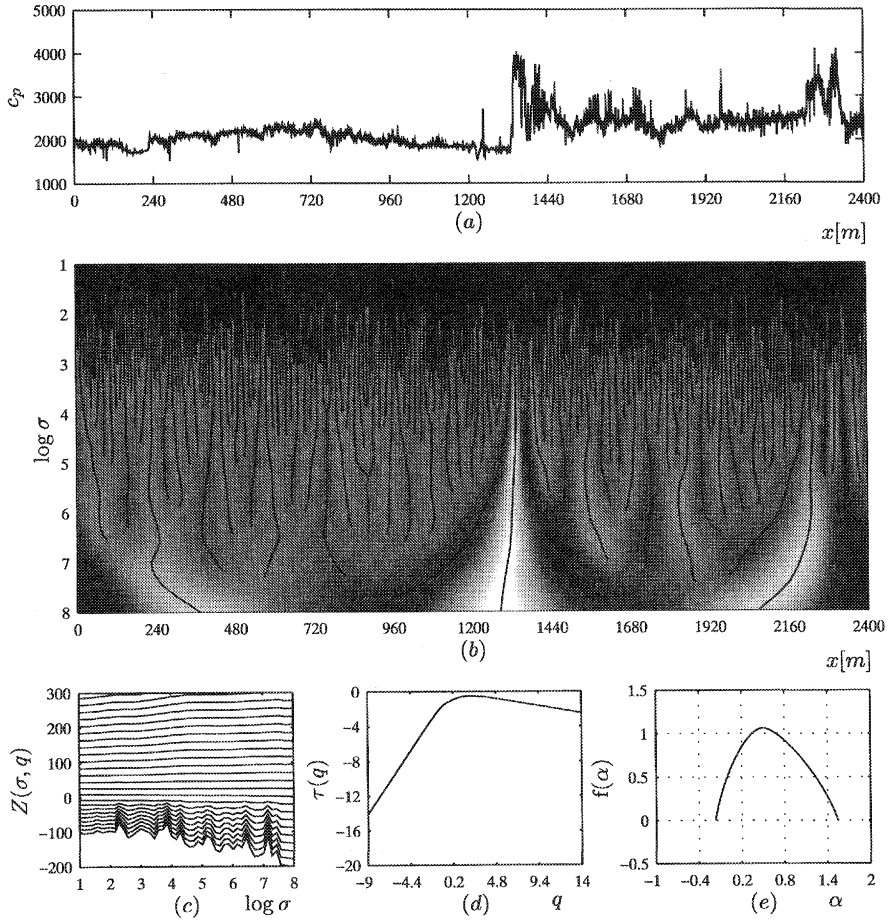

Figure 2.15 *(a) A well-log measurement of the compressional wave speed. The sample interval is 60 cm and the total length of the data set is 2.4 km. (b) The wavelet transform modulus maxima lines of (a). The wavelet used is the second derivative of a Gaussian, so it is infinitely smooth and it has two vanishing moments. The scale of the wavelet ranged from 5 m to 160 m. On the fine scales the modulus maxima lines start at every point, indicating that the second derivative of the well-log is singular everywhere. Via the partition function $Z(\sigma, q)$ of (c) it is possible to compute the mass exponent $\tau(q)$ of (d). This function in turn allows for the computation of the singularity spectrum $f(\alpha)$ which is depicted in (e). The weakest singularity is found to be 1.6, showing indeed that the signal is singular everywhere in its second derivative. The strongest singularity equals -0.18. Finally the degree of non-conservation equals $H = 0.21$.*

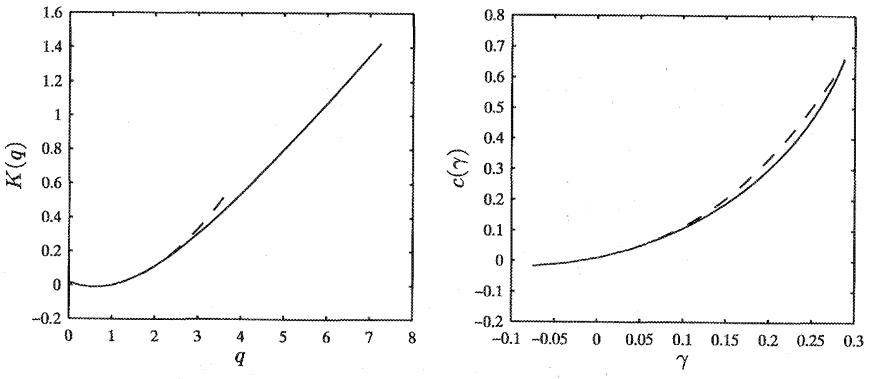

Figure 2.16 *Estimation procedure for the universal multifractal parameters applied to a real well-log measurement. On the left the second characteristic function $K(q)$ is depicted together with its fit given by equation (2.30) whereas on the right the co-dimension function for the singularities $c(\gamma)$ is displayed, also accompanied by its fit given by equation (2.31). The estimated parameters used for the fits equal, $C_1 \simeq 0.056$ and $\alpha_l \simeq 1.83$ and yield the estimation of the universal parameters. The q range used in the nonlinear fitting procedure, for the α_l, was taken to be $q \in [0, 3.5]$.*

Determination of the multifractal model parameters

What remains for me to do is to come up with a model that characterizes the singularity spectrum. Keep in mind that such an effort will not explain the actual local specifics of certain a data set; it merely characterizes the scaling complexity as a whole. The continuous universal multifractal cascade model constitutes a process with a singularity structure, within the α-range corresponding to positive q, that is comparable with the observed spectra. In this way the universal multifractals capture the intermittency and non-stationarity. They also adhere to the condition of strict positiveness, a notion not supported by additive Brownian motions and their generalizations, fractional Brownian motions.

In figure 2.16 I included an example where the universal multifractal parameters are estimated, via a non-linear curve fitting procedure, from the estimated[42] $K(q)$ and $c(\gamma)$ functions. The obtained parameters equal, see equations (2.30) and (2.31), $C_1 = 0.056$ and $\alpha_l = 1.83$. The q range in the non-linear estimation procedure, see below, for the α_l was taken to be $q \in [0, 3.5]$. Let me briefly explain the procedure I used to come up with estimates for the universal multifractal parameters α_l, C_1 and H from the sample $K(q)$ and/or $c(\gamma)$ functions. The procedure amounts to

[42] Via the estimated $\tau(q)$ and $f(\alpha)$ of figure 2.15.

	H	C_1	α_l
Well I (comp. velo.)	0.21	0.034	1.55
Well II (comp. velo.)	0.22	0.049	1.67
Well II (density)	0.23	0.048	1.38
Well II (kappa)	0.23	0.027	1.68
Well III (comp. velo)	0.11	0.044	1.64
Well III (shear velo)	0.14	0.043	1.61
Well III (density)	0.14	0.044	1.72
Well III (gamma ray)	0.21	0.031	1.66

Table 2.1 *This table displays the estimation results for the multifractal parameters C_1, α_l and H from real well-log measurements.*

- estimation of the $K(q)$ and $c(\gamma)$ functions from the $Z(\sigma, q)$ function.
- using the relation $H = \{K(q) - qH\}_{q=1}$ to estimate the H from the $K(q)$ function.
- using the relation $C_1 \triangleq \{\frac{d}{dq}(K(q) - qH)\}_{q=1}$ to estimate the C_1 from the derivative towards q of the $K(q)$ function.
- conducting a non-linear fit, using the closed form expressions for the stochastic expectation of the $K(q)$ and/or $c(\gamma)$ functions on the sample $K(q)$ or $c(\gamma)$ functions, i.e. finding an α_l which minimizes the error (in a least squares sense) between the *parametric* and *sample* curves.

Figure 2.16 displays the result of this estimation procedure set to work on the $c(\gamma)$ function obtained from the f(α) spectrum depicted in the figure 2.15. Comparing the sample curves, denoted by the solid line, with the parametric curves, denoted by the dashed line and defined in terms of the non-linear estimates, shows a good agreement up to values of $q = 3.5$. Notice that the limited q-range is not surprising since only one realization, the well-log, is considered.

In table 2.1 I included a list of estimates for the universal multifractal parameters obtained from a series of different well-logs acquired at different locations around the world. These estimates indicate that the three parameters do not vary substantially from one well-log to the other[43]. Even estimates obtained by analysing other physical quantities, such as the gamma ray, show a similar kind of behaviour. The same applies to the compressibility in well II, which was obtained by using the relation between the density, compressional wave speed and the compressibility, i.e.

$$\kappa(z) = \frac{1}{\rho(z)c_p^2(z)}. \tag{2.64}$$

[43]The values for the H of well II seem to be an exception to that.

The reason for this latter property to exist lies in the fact that the multiplicative multifractals are, beyond a renormalization constant, insensitive to fractional powers. The explanation for the apparent "universality" for the estimated values of the coefficients pertaining to well-logs is, however, less obvious. This insensitivity might be understood by taking into account that the universal multifractal parameterization merely captures the overall scaling, i.e. the "scaling structure" of the unknown underlying physical process[44] which is responsible for the sedimentation process. Evidently this process is, amongst others, driven by the dynamics of the hydrodynamic turbulence in the atmosphere.

Finally remark that the concept of multifractals has already received quite a substantial attention in the literature (Muller et al., 1992; Saucier and Muller, 1993). Recently, interesting results on the scaling of the Ice-Core of Greenland (Schmitt et al., 1994), have been added to this yielding estimates for the multifractal parameters close to the estimates obtained for the well data. The underlying reason for this coincidence may lie in the idea that well-logs as well as Ice-Pack measurements reflect the turbulent changes in the weather/climate on a geologic time scale.

A synthetic well-log simulation

With the estimated values for the multifractal parameters it is possible to generate a realistic synthetic multifractal with its parameters set according the estimated values summarized in table 2.1. In figure 2.17 such an example is given for the well-log I. Comparing both the compressional wave speeds and reflection coefficients of the original well and the corresponding random multifractal simulation shows that the generalized approach captures the intermittency as well as the texture better than the monofractal example depicted in figure 2.13.

2.7 Review of the conventional approaches

Let me now bring the multiscale approach as proposed in this chapter in the perspective of the conventional approaches being applied – predominantly in the field of waves in finely layered media – to represent the complexity of the constitutive parameters as they are being observed in well-log measurements. Probably the first thing that comes to mind is to represent the well-log measurements as a stack of thin[45] layers defined in terms of the consecutive discretization points offered by the well-log measurement. Clearly such an approach is a special case of a piecewise continuous representation that is perceived as being the proper choice for the representation. Provided with the estimated singularity spectrum, acquired from real well-log measurements, it is difficult to reconcile the piecewise continuous representation – where the spectrum would have consisted of

[44] An exception to this forms the non-linear Navier-Stokes equation which, on theoretical grounds, gives rise to a $H = \frac{1}{3}$ (Schertzer and Lovejoy, 1987a,b) for fully developed turbulence (Kolmogorov's Law, see Kolmogorov (1941)).

[45] Small compared to the dominant wavelength.

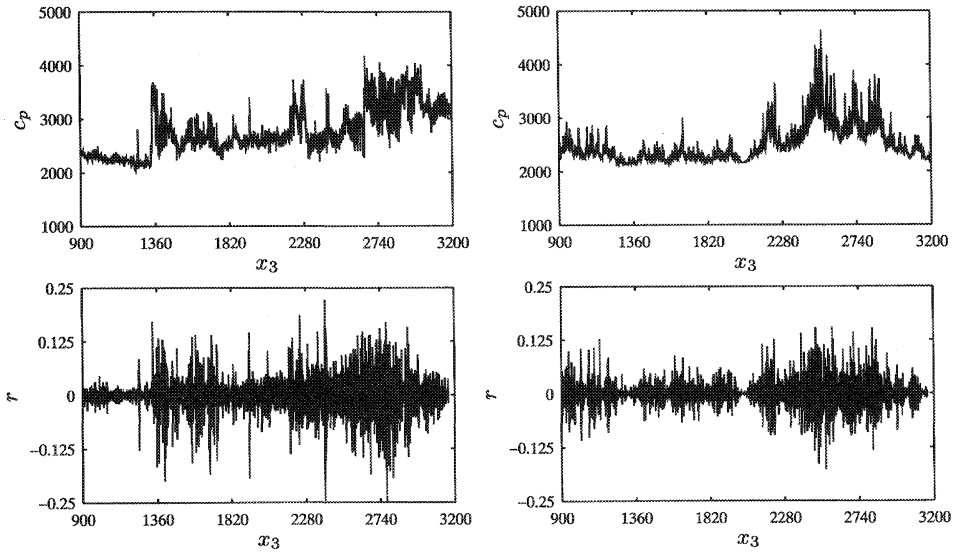

Figure 2.17 *Comparison between the compressional wave speeds and acoustic reflection coefficients yielded by a real well-log measurement (left column) and their multifractal counterparts (right column). The multifractal parameters were set according to estimates obtained from the well-log measurement. Remark that the simulation only captures the global characteristics not the specifics.*

two points namely $\alpha = 0$ for the discontinuous and $\alpha \in \mathrm{N} \wedge \alpha > 1$ for the arbitrarily smooth part – with the evidenced distribution of singularities in the singularity spectrum.

As a second choice – more with an effective representation for the wave dynamics on the seismic scale range in mind – people base their approaches on *a separation of scales.* This separation either refers to the invocation of some kind of smooth/trend background model with possibly discontinuous fluctuations superimposed on it or to a blocked version of the well-log. This blocking operation often occurs without any theoretical[46] justification although the work of Vermeer (1992) forms an exception to this. He already introduces the continuous wavelet transform to allocate the singularities at the different scales but then he continues to assign discontinuities to the abscissa where the well-log is singular. This is understandable with his geological application in mind but it misses the characterization of the singularity strength and may influence the wave interactions.

The main idea behind the invocation of a separation of scales is the conjecture that the

[46]Remark that the blocking can be brought into the perspective of multiscale analysis by using the box-car/indicator function as the smoothing kernel. However, in that case one is bound to an observable detection range for the singularities of $\alpha \in (-1, 0)$, a notion effecting the coarse-graining operation.

smooth trend model can be held responsible for the propagation effects experienced by the probing wavefield while the reflection is attributed to the remaining fine-scale detail. This conjecture is perceived quite differently in the field of finely layered media where the fine-scale fluctuations are seen as thin layers the scattering at which are designated as being responsible for an apparent dispersion mechanism. But still the main idea is to think of the complexity to consist of some well separable trend defining a macro/meso model with a possibly random perturbation superimposed on it.

Perhaps unnecessarily let me mention again that the multiscale analysis and quantification on real well-data have demonstrated that a separation of scales might be futile. The data displays a powerlaw type of behaviour over a wide scale range and in that case I think it is unlikely to expect satisfactory results in a quest, based on a separation of scales, for a better integration of well and surface seismic data, a problem of scales. During the integration a better treatment of the dynamics is essential and one can not longer be satisfied by treating the kinematical aspects solely. Here I mean with kinematics the predictions for the traveltimes given by ray-asymptotic methods.

Concerning the separation of scales I like to add that certainly the well-logging tool implements some kind of averaging (Hsu and Burridge, 1991) yielding a break in the scaling at about a scale range of 1.5 to 2.5 meter. The occurrence of this break, however, does, in my opinion, not justify the use of an exponential decaying covariance function, the correlation length of which exactly matches the smoothing interval of the tool. In that respect I, frankly speaking, differ from the opinion that the highly erratic complexity displayed by the well-data can appropriately be characterized by a random process parameterized by a single parameter, the correlation length. This correlation length delineates the transition from a white noise type of behaviour for the low spatial frequencies, $S(k) \propto C \quad k \ll \zeta_c^{-1}$ with C a constant and ζ_c, the correlation length, to a Brownian motion type of behaviour for the high frequencies, $S(k) \propto k^{-2}, \quad k \gg \zeta_c^{-1}$.

Inspection of the literature shows that the majority of papers dealing with waves in finely layered media utilize a random process for the medium fluctuation with an exponential decaying covariance function. There exist several alternatives to define such a process. Banik et al. (1985b) and Baluni (1985) started the discussion by introducing the random telegraph model comprising of alternating layers with two different acoustic impedances and Poisson distributed layer thicknesses. Burridge and Chang (1989) followed by Kerner (1992) and Shapiro and Zien (1993) use a first order Markov process to generate a random sequence for the fluctuations. This type of process also results in an exponentially decaying correlation function but having equidistant layers rather than Poisson distributed layer thicknesses.

Herrmann (1991) interrupted the series of papers based on the exponential type of correlation function by proposing the use of the fractional Brownian motion[47] as a model –

[47]It were Walden and Hosken (1985) followed by Todoeschuck and Jensen (1989) who first suggested

with infinite correlation length and powerlaw type of behaviour for the energy spectrum and structure function – for the medium fluctuations evidenced from well-log measurements (Herrmann and Wapenaar, 1992, 1993). This approach followed by Burridge et al. (1993) provided a characterization of the "blueness" for the power spectrum of the reflection coefficient sequence and found its way into the O'Doherty-Anstey formula (O'Doherty and Anstey, 1971). This formula approximately describes, in terms of the power spectrum of the reflectivity, the propagation of waves in media that vary weakly in one dimension. It appeared that this combination yields interesting results since the transmitted wavefield strongly depends on the blueness, the degree of anticorrelation displayed by the reflection coefficient sequence (Herrmann, 1991). Unfortunately the weakness condition, imposing a scaling behaviour for the reflectivity corresponding to that of Brownian motion, possibly withstands a proper application of this method given the observation $H \simeq 0.21$. In the epilogue I will discuss waves in relation to scaling. Besides this fundamental observation it also fails in the following two ways:

- it does not handle the intermittency, a striking characteristic of well data where erratically occurring catastrophic events seem to be all important.
- it is not strictly positive, it is not a density, so it can not be used to sensibly characterize the constitutive parameters which are void of a physical meaning when negative.

On the other hand remember that the intrinsic scale-invariance and the roughness characterization certainly had their merits.

This review has not been intended to be all inclusive but it shows that the multiscale approach as being presented in this chapter has not yet found its way in wave theory dealing with highly erratic, non-differentiable media. On the other hand there has been a recent upraise in the use of multiscale models to characterize the well-log's complexity in a more general context (Vermeer, 1992; Saucier and Muller, 1993; Collier, 1993; Li and Haury, 1995; Painter, 1995). Unfortunately, until now, a discussion on the implication of the finding on wave dynamics is lacking.

I like to conclude this section by stating that I did not pay attention to periodic and hence infinitely smooth perturbations. The reason for this lies in the outcome of the multiscale analysis and in the fact that the periodic structures give rise to a stop-band type of behaviour for the wavefield, an observation not empirically substantiated. Despite this argument against the use of periodic perturbation Jannane et al. (1989) and a very large group of co-authors published a short note where they claim – based on synthetic modelling – a separation of scales in relation to what kind of structures can be resolved

to use fractional Brownian motion to characterize the "red" power spectra for the fluctuations experienced from well-log measurements and fractional Gaussian noise to describe the "blue" spectrum for the reflectivity.

from seismic data synthetically. Given the multiscale analysis presented in this chapter I would be very reluctant to make such a statement from the medium's complexity point of view. Of course the statement of Jannane et al. (1989) refers to the seismic data where they used the wave equation to conduct the mapping.

2.8 Concluding remarks

In the introduction of this chapter I conjectured that that there are two fundamental averaging processes, a global self-averaging over the propagation distance and a local averaging process over the effective gauge. These averaging processes can be associated with wave interactions, with the transmissivity and reflectivity, and become important in media with a complex structure on the scale range inhabited by the transient wavefield. Despite the current lack of a true understanding and theoretical basis of this conjecture it still forms a worthwhile point of departure for the introduction of a multiscale medium representation in the context of wave dynamics. In this representation the propagation and reflection are associated to the afore mentioned averaging processes. It is this view that inspired the two ways of looking at the displayed scaling within the multiscale representation for the medium. The multiscale representation in itself can be thought of as an abandoning of the rather restrained and perhaps unphysical representation of the medium's complexity by functions in favour of a representation by means of the continuous wavelet transform. In this wavelet representation one defines the proper functional, assigning a value to a proper space of test functions, an action striding well with the way in which physical interactions are perceived to take place.

The scaling medium representation

To summarize the proposed scaling medium representation provides

- a true invocation of scale in the sense that the wavelet representation refers to the *details* in the function ***at*** a *specific* scale.
- an efficient partitioning in terms of the WTMML delineating the information carrying singularities that can be associated with the occurrence of reflections.
- a reconstruction mechanism yielding scale-indexed coarse-grained and hence regularized approximations of the bare, possibly singular, physical quantity. Notice that this smoothing operation has, because of the proper choice for the wavelet, been done in a mathematical proper way.
- a local and global multiscale analysis and quantification. The first is aimed at capturing localized estimates for the regularity to be associated to local reflection while the latter captures the more general scaling behaviour by unraveling the intertwined singularity structure.

- the possibility for a representation in terms of the multiresolution analysis defined in terms of the orthogonal discrete wavelet transform as proposed by Mallat (1989). Via selection of the proper family of discrete wavelets is possible to span a Sobolev space of choice.

To recapitulate the whole idea of the scaling medium representation and its partitioning I included figure 2.18. In this picture I show that the wavelet representation clearly reveals the scaling complexity across the different scales. Moreover the WTMML's delineate the singularities. In the conventional way of representing this type of data sets, namely as ordinary functions, the notion of scaling would have been obscured by the sole way in which the data is represented. The remaining question is whether this type of representation becomes manifest in the way in which physical models are being devised. I will address this question in the epilogue.

A multiscale reflector

Since the multiscale analysis corresponds to a local averaging procedure triggering on the regions of rapid variation, one can associate it with an anticipated scattering to occur at the singularities. The sheer elegance of the WTMML formalism happens to be the property that the WTMML's exactly point to the regions where the medium acts as if being singular. Generally speaking I expect the existence of an intrinsic relationship between occurrence of local scattering and the emergence of a WTMML. Therefore I like to propose with the following definition for a local specular multiscale reflector:

Definition 2.3: A multiscale reflector
A multiscale reflector constitutes the set of points delineated by a WTMML,

under the restriction that the decay/growth rate for magnitude of the WTMML's has an Hölder exponent associated to it that lies within the "observable[48]" range of the wave dynamics. For me personally it remains to be shown what that range really is. At least for $\alpha > 1$ I do not expect specular reflections anymore but instead turning ray type of reflections. The recent work by Wapenaar (1996a) has shown interesting results on solving the local scattering problem at singularities of the above algebraic type. In that approach boundary conditions are imposed while the accessible range for the Hölder exponents is restricted to $\alpha > -1$. However, these results do in my opinion not preclude a discussion on the role of the notion of scale in relation to wave dynamics, see the epilogue.

Generalized scaling macro/meso model

Given the definition for the multiscale reflector it is not hard to imagine a generalized type of macro/meso model that represents a sparse[49] representation of the scaling over

[48] In the sense that the wave theory may loose its meaning or may not specularly scatter at it.
[49] Here I mean with sparse the notion that for the coarse scales the number of WTMML's decrease.

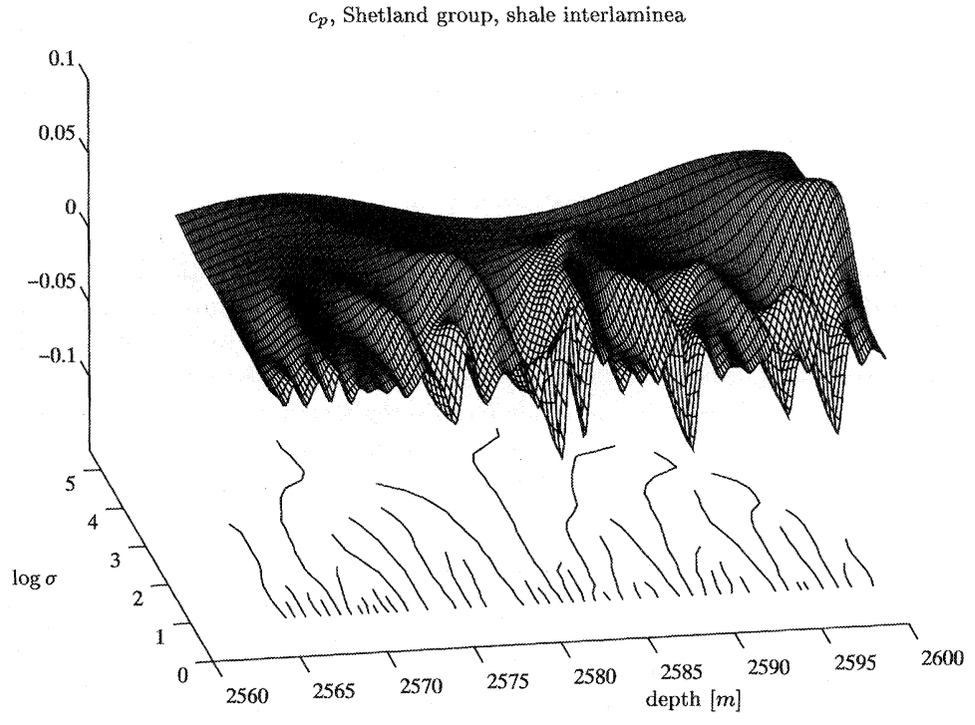

Figure 2.18 *Illustration on how to "look" at a scaling measurement. On the top the continuous wavelet transform, conducted on a selected well-log interval, is on display. The scale decreases while going out of the plane. At the bottom I included the* WTMML*'s that delineate the location of the singularities.*

a specified scale range. This generalization has been one of the goals of this chapter and indeed the WTMML formalism provides all necessary prerequisites. Clearly the approach extends the conventional way of representing a transition – by the invocation of a jump discontinuity that generates a specular reflection – to a more general type of singular behaviour captured by the wavelet coefficients along the WTMML. It was shown that the singular behaviour can be quantified by the Hölder scaling exponent. For a jump discontinuity this exponent equals $\alpha = 0$ and hence represents a special case. Despite the findings of Wapenaar (1996a) who solves the wave interaction at these algebraic singularities it remains, for me, an open question how to come up with a formulation, preferably from first principles and in terms of the wavelet coefficients or equivalently in terms of the WTMML's, for the specular wave dynamics induced by the singularities.

Only then one has an apparatus at hand that integrates the seismic and well-log data.

To summarize

In this chapter I committed myself to come up with a proper representation and characterization for the complexity as being displayed by the constitutive parameters evidenced from well-log measurements. The main point I wanted to make is that a representation for the discretized well-log measurements in terms of ordinary functions is full of pitfalls and better be circumvented. For that reason I proposed a representation that is mathematically sound and which does not mystify the all important notion of scale. In that respect it is worthwhile to mention that the introduction of a macro/meso model derived from a well-log is futile when the magnitude of the scale derivative or equivalently the wavelet coefficients can not be neglected. However, the WTMML partitioning offers a powerful alternative because it provides a sparse representation of the space-scale plane aiming at the singularities. By anticipating the notion that the scattering is expected to occur at these singularities the WTMML partitioning is seen as a generalized multiscale "macro/meso model".

Give the proposed multiscale representation I am also able to demonstrate that certain mathematical operations that are normally not allowed on the data can be given meaning in the new representation. This can only be done at the expense of adding one fundamental parameter namely the scale index. On the other hand I was able with the proposed representation to come up with a characterization and even a parameterization for the complexity by means of local regularity estimates in the form of Hölder exponents and global regularity estimates with the help of the singularity spectrum. These Hölder exponents may well prove to be indicators, delineating certain geological features and or transitions, while the singularity spectrum entails an unraveling of the hierarchical scaling structure. Moreover this spectrum provides information on the data's non-stationarity annex integrability, intermittency and generalized fractal dimensions.

But above all, the complexity displayed by the well-data withstand a separation of scales. This becomes, for instance, manifest in the notion that it is difficult to still speak of the earth's subsurface as being composed of well separated layers as an onion. These observations deserve in my opinion a careful rethinking of the underlying assumptions that yielded the wave equation, see the epilogue.

Chapter 3

Operators

In this chapter I lay down an operator formalism pertaining to the shift and dilatation operations required to set up the multiscale analysis and representation presented in chapter 2 and chapter 8. The purpose of this chapter is fourfold. First it is intended to substantiate the shift and dilatation operation by linking these operations to the spatial derivative and scale derivative operator. Secondly I will demonstrate that the Fourier and Mellin/scale transform represent unitary transformations being defined in terms of the generalized eigenfunctions associated with these two differential operators. Thirdly I will make exploratory remarks on the link between the spectral scale representation and (multi)fractals. Finally this chapter serves as a preparation for the discussion on wave dynamics to be presented in chapter 4 and the epilogue.
key words: *self-adjoint/Hermitian operators, eigenvalue problems, spectral representations, commutation relations, Fourier Transforms, Mellin/scale transforms, fractals, singularity spectra.*

3.1 Introduction

Finding a suitable representation for a detected signal plays an important role in inferring information. The reason for this lies in the fact that a representation constitutes an expansion of the signal onto a basis of expansion functions. These expansion functions are found by solving the eigenvalue problem connected to the operator that is held responsible for the detected signal. Obviously such an expansion can only be successful when expansion functions are used that behave natural towards the signal. Now how can one find these expansion functions? To answer this question one enters in the realm of physics where one commits oneself to find the proper physical model explaining for the observed physical phenomenon. In those cases where the predicted data by the physical model can be reconciled with the actual measured data one is hopefully[1] in the position

[1]It may either occur that the measurement process is interfered – e.g. by the inevitable smoothing/coarse-graining of the instrument or by the lack of data – or that the physical model was incorrect despite that it may have given acceptable results. Unfortunately none of these artifacts can be ruled out in the physical setting.

to infer information.

Since the advent of Quantum Mechanics physicists have found a very elegant way, by means of operator theory, to find the suitable representation for measured signals (Messiah, 1958; Reed and Simon, 1978, 1979; Dautray and Lions, 1990; Cohen, 1995). This representation occurs in terms of basis functions that solve the eigenvalue problem pertaining to the operator which is responsible for the emanated signal. For example, the physical variable frequency emerges when the operator concerns the spatial frequency operator – containing the spatial derivative – and the pertaining Fourier transform establishes the spectral representation. Here the frequencies correspond to the eigenvalues while the complex exponentials represent the eigenfunctions. This type of approach can be applied to arbitrary operators whose attributed solution to the eigenvalue problem yields a way to analyze, manipulate and construct signals that are believed to be connected to this operator. In this chapter I will pay primarily attention to the substantiation of the shift and dilatation operation that represent the two operations on which the multiscale analysis presented in chapters 2 and 8 is based. This corresponds to reviewing the *space* and *scale* representations pertaining to the frequency operator, that allows for the formal definition of the shift operation, and the scale operator establishing the dilatation operation. At first I will introduce the concept of operators, their Hermitism and their unitarian behaviour. Then I will solve the eigenvalue problem and observe that when the operator is Hermitian one automatically gets real eigenvalues while the eigenfunctions constitute an orthogonal and complete set of basis functions enabling a proper spectral representation. Given this spectral representation one can demonstrate that functions of operators can readily be given a proper meaning and this makes it possible to issue statements on the average values that are cognate to the operator. The joint behaviour of physical quantities is addressed next and leads to the introduction of commutator relations expressing the fact that when two operators do not commute one cannot resolve their associated physical quantities up to arbitrary accuracy in the $L^2(\mathbb{R})$ sense. That is to say that the "variance", the "bandwidth", of one variable cannot be improved independently of the other, a notion one is all to aware of given the Heisenberg uncertainty relation, expressing the uncertainty between the Fourier pair momentum and position in Quantum Mechanics while Gabor (1946) showed its counterpart for simultaneous time and frequency analysis. The discussion is continued by paying explicit attention to the space and scale representations followed by a section numerically elucidating the shift and dilatation operations. I will conclude this chapter by bringing the concepts of fractals and the scale representation in relation with each other.

3.2 Operators

Strictly speaking operators refer to constructs whose action alters functions or functionals. They comprise actions such as differentiation, integration or squaring. Within the physical context these constructs generally refer to operators of the Hermitian type. For

two functions, $f(x)$ and $g(x)$ one can write

$$\langle \mathcal{A}f, g\rangle = \langle f, \mathcal{A}^\dagger g\rangle, \tag{3.1}$$

where the angular brackets refer to the inner product defined as

$$\langle f, g\rangle \triangleq \int f(x)g^*(x)\mathrm{d}x \tag{3.2}$$

and where the calligraphic symbol is reserved to denote the operator and where the adjoint of the operator $\mathcal{A}$ is denoted by $\mathcal{A}^\dagger$. In those cases where the adjoint of the operator, $\mathcal{A}^\dagger$ equals itself, i.e. $\mathcal{A} = \mathcal{A}^\dagger$, one speaks of a self-adjoint or Hermitian operator and this corresponds to an operator whose action forces the identity

$$\langle \mathcal{A}f, g\rangle = \langle f, \mathcal{A}g\rangle \tag{3.3}$$

to hold. In that case equation (3.1) becomes the definition for a self-adjoint operator. An operator is unitary when its adjoint equals its inverse, i.e.

$$\mathcal{U}^\dagger = \mathcal{U}^{-1}, \tag{3.4}$$

where $\mathcal{U}$ is an unitary operator. Notice that the inverse of an operator is an operator that makes the following identity to hold

$$\mathcal{A}^{-1}\mathcal{A} = \mathcal{A}\mathcal{A}^{-1} = \mathcal{I}, \tag{3.5}$$

where $\mathcal{I}$ is the identity or unit operator.

The importance of the unitary type of operators lies in the fact that they do not alter the norm of the function they are acting upon, i.e. $\|f\|_2 = \langle f, f\rangle = \|\mathcal{U}f\|_2$. For operators to posses such a property they have to be of the form (Cohen, 1995)

$$\mathcal{U} = e^{j\mathcal{A}} \tag{3.6}$$

with $\mathcal{A}$ being Hermitian. To understand the unitary property one has to prove that

$$\mathcal{U}^\dagger = e^{-j\mathcal{A}^\dagger}. \tag{3.7}$$

To do this one can use a Taylor expansion for the exponent,

$$\begin{aligned}\mathcal{U}^\dagger = (e^{j\mathcal{A}})^\dagger &= (\sum_{n=0}^{\infty}\frac{(j)^n}{n!}\mathcal{A}^n)^\dagger \\ &= \sum_{n=0}^{\infty}\frac{(-j)^n}{n!}(\mathcal{A}^\dagger)^n = e^{-j\mathcal{A}^\dagger},\end{aligned}$$

and then it becomes straightforward to show that

$$\mathcal{U}\mathcal{U}^\dagger = e^{j\mathcal{A}}e^{-j\mathcal{A}^\dagger} = e^{+j\mathcal{A}}e^{-j\mathcal{A}} = \mathcal{I}, \tag{3.8}$$

implying $\mathcal{U}^\dagger = \mathcal{U}^{-1}$ and defining the unitary operator. Finally remark that unitary operators are not Hermitian but they obey

$$\langle \mathcal{U}f, g\rangle \triangleq \langle f, \mathcal{U}^\dagger g\rangle = \langle f, \mathcal{U}^{-1} g\rangle. \tag{3.9}$$

Now let me go back to how one can interpret equation (3.6). Indeed by expanding it into a Taylor expansion one can give the operator exponential $e^{j\mathcal{A}}$ a meaning. Strictly speaking such an approach is allowed but it is often not well tractable. In the sequel I will show that when solving the eigenvalue problem it becomes rather straightforward to give functions of operators, such as the operator exponential, a meaning. Besides this evident advantage one also obtains information on the nature of the spectrum of the eigenvalues and a way to expand arbitrary functions into a basis yielded by the eigenfunctions. In this manner one can readily represent the operator and the functions they are acting on in the spectral domain.

3.2.1 The eigenvalue problem

Suppose one has a Hermitian operator then the solution to the eigenvalue problem entails

$$\mathcal{A}v(a, x) = av(a, x), \tag{3.10}$$

where attaining the expression for the double-indexed functions, $v(a, x)$, that equate the above relation, is referred to as solving the eigenvalue problem. This family of functions is denoted as the eigenfunctions being indexed by the a's, the corresponding scalar eigenvalues. Because of the imposed Hermiticity the eigenfunctions are complete, orthogonal and the eigenvalues are real. The orthonormality of the eigenfunctions is expressed by

$$\langle v(a', \cdot), v(a, \cdot)\rangle = \delta(a' - a), \tag{3.11}$$

while their completeness becomes manifest by

$$\langle v(\cdot, x'), v(\cdot, x)\rangle = \delta(x' - x). \tag{3.12}$$

These two properties facilitate the expansion of an arbitrary f into

$$f(x) = \int F(a)v(a, x)\mathrm{d}a \tag{3.13}$$

with $v(a, x)$ playing the role of the transformation "matrix" also known as the basis kernel (Cohen, 1995) whereas the quantity $F(a)$ delineates the spectral representation for $f(x)$ in the a-domain[2]. The spectral representation is obtained via

$$F(a) = \langle f, v(a, \cdot)\rangle. \tag{3.14}$$

[2]Notice that I refrained in this discussion to make remarks on the specific nature of the spectral measure. Here I assume the spectral measure to be absolutely continuous and I also neglect the possibility that the eigenfunctions do not strictly lie in $L^2(\mathbb{R})$.

Again remark that in case of a Hermitian operator the eigenvalues are real, a property shared by all observable physical quantities, and they may attain an (in)finite range of values or they may even be discrete. The effective range for the eigenvalues becomes apparent when the eigenvalue problem is solved. Given this solution one is able to predict the attainable values for the physical quantity a. This range is known as the spectrum of a, $\sigma(\mathcal{A})$, and hopefully matches the experiment. Unfortunately this prediction can be wrong implying that the operator, representing the physical model, does not represent the physical process correctly; the eigenvalue problem might have been solved wrongly or the measurement was erroneous (Cohen, 1995). However, keep in mind that the solution of the eigenvalue problem only gives the attainable eigenvalues to be associated with the operator. When particular measurements are under consideration one has, of course, to do with the actual values that are incorporated in the function $F(a)$ whose range may or may not extend over all possible values of a. But again solving the eigenvalue problem has nothing to do with the particularities of the measurement. It is merely the expansion of the operator onto the basis of eigenfunctions that counts and that has to be constructed for the different measurements yielding different outcomes for the (density) "spectrum" $F(a)$. Unfortunately the term spectrum has an ambiguous meaning in the (physical) literature since it refers both to the attainable eigenvalues, the set $\sigma(\mathcal{A})$ being defined by the set of eigenvalues such that $(\mathcal{A} - a)$ is not invertible as a bounded operator, as well as to the expansion coefficients, $F(a)$, expressing the rate of occurrence of a certain eigenvalue, e.g. a frequency for the Fourier transform. Hopefully the meaning of the spectrum becomes clear from the pertaining context. The reader is referred to chapter 4 where I will pay attention to the nature of the spectrum and the accompanying physical states in the context of acoustic wave motion.

3.2.2 Spectral representation of operators

When the eigenvalue problem is solved one is able to expand an arbitrary function into a basis formed by the eigenfunctions,

$$f(x) = \int F(a)v(a,x)\mathrm{d}a \tag{3.15}$$

and this expansion is also known as the *spectral representation* of f in the domain defined by the eigenfunctions of the operator $\mathcal{A}$. In a similar fashion one can acquire a spectral representation for the operator or functions of the operator. By these constructs I mean operators that are defined by expanding the function $Q(y)$ into a Taylor series and subsequently replacing the ordinary dependent variable y by the operator $\mathcal{A}$,

$$Q(y) = \sum_n c_n y^n \quad \longrightarrow \quad \mathcal{Q}(\mathcal{A}) = \sum_n c_n \mathcal{A}^n. \tag{3.16}$$

When $\mathcal{A}$ is Hermitian it can be shown that $\mathcal{Q}(\mathcal{A})$ will share this Hermitian property in case the Taylor expansion coefficients, c_n, are real.

What is the action of $\mathcal{Q}(\mathcal{A})$ on the eigenfunctions of $\mathcal{A}$? To answer this regard

$$\mathcal{Q}(\mathcal{A})v(a,x) = \sum_n c_n \mathcal{A}^n v(a,x) = \sum_n c_n a^n v(a,x), \tag{3.17}$$

from which one can recognize that the following relation holds

$$\mathcal{Q}(\mathcal{A})v(a,x) = \mathcal{Q}(a)v(a,x). \tag{3.18}$$

Using this result it becomes rather straightforward to define the action of the operator $\mathcal{Q}(\mathcal{A})$ on an arbitrary function, $f(x)$, via the spectral representations of the function and the operator. So write,

$$\begin{aligned}\mathcal{Q}(\mathcal{A})f = \mathcal{Q}(\mathcal{A})\{f\}(x) &= \mathcal{Q}(\mathcal{A}) \int F(a)v(a,x)\mathrm{d}a \\ &= \int F(a)\mathcal{Q}(\mathcal{A})v(a,x)\mathrm{d}a\end{aligned}$$

which becomes

$$\mathcal{Q}(\mathcal{A})f = \int F(a)Q(a)v(a,x)\mathrm{d}a, \tag{3.19}$$

where the equality of equation (3.18) is applied. Remark that in this case one no longer requires to expand $\mathcal{Q}(\mathcal{A})$ into a Taylor series expansion because one has ordinary calculus at hand when the scalar eigenvalues are concerned.

3.2.3 Commutators

Up to this point the discussion primarily focussed on the examination of one single physical quantity and its associated operator. What happens when one is dealing with two operators $\mathcal{A}$ and $\mathcal{B}$. To follow Cohen (1995): *Generally the order of operations is not interchangeable, as, for example, putting on socks and shoes.* In that case the order clearly matters and the operators are said to be non-commutative. In cases where the order of the operations is irrelevant the operators are said to commute. To check whether operators commute forms an important aspect of physics and the method comes down to evaluating the possible difference between the operator combinations $\mathcal{AB}$ and $\mathcal{BA}$ both acting on a suitable function. The operator expressing this difference is called the *commutator* of $\mathcal{A}$ and $\mathcal{B}$ and is denoted by

$$[\mathcal{A},\mathcal{B}] = \mathcal{AB} - \mathcal{BA}. \tag{3.20}$$

This operator plays a central role for the uncertainty principle for arbitrary variables and the reader is referred to the pertaining discussion below.

3.2.4 An example

To illustrate some of the aspects of operators reviewed so far I would like to discuss one example. Consider the space operator in the space representation, $\mathcal{X}$, that equals a multiplication by x in the space representation. The eigenvalue problem associated with this operator reads

$$\mathcal{X}v(x',x) = x'v(x',x), \tag{3.21}$$

where the x''s refer to the eigenvalues whereas the solution for the eigenvalue problem are given by the Dirac distribution,

$$v(x',x) = \delta(x-x') \qquad \forall x' \in \mathbb{R}. \tag{3.22}$$

Since the operator is Hermitian it follows that these eigenfunctions are complete and orthonormal while the eigenvalues are continuous and real. However, care should be taken in the sense that the Dirac distribution does not constitute an ordinary function, it is a so-called generalized function that can only be given a meaning in the distributional sense. Please, when the reader is unfamiliar with the concepts of distribution theory, consult chapter 5.

3.2.5 Averages and the characteristic function for operators

One of the main advantages of the operator method lies in the fact that it is relatively straightforward to calculate averages and characteristic functions that are associated with the operators. Following Cohen (1995) I adapt to take $|F(a)|^2$ for the density, a natural choice when the energy in a signal is concerned. The reason why $F(a)$ does not lend itself to be a density lies in the fact that it is a complex quantity so it is impossible to connect a norm to it. Remark that as an alternative choice one can take $|F(a)|$. Finally notice that the densities here bear a fundamental similarity with the distribution functions. When these latter density functions are supplemented with an additional normalization yielding $\|F(a)\|_2 = 1$ one can rightfully interpret them as probability density functions.

The average of a function g living in the spectral domain a can be written as

$$\langle g \rangle = \int g(a)|F(a)|^2 \mathrm{d}a, \tag{3.23}$$

where $F(a)$ denotes the spectral representation of the function $f(x)$ in the domain pertaining to the operator $\mathcal{A}$. In the operator domain this definition is equivalent to

$$\langle g \rangle = \int f^*(x) g(\mathcal{A}) f(x) \mathrm{d}x. \tag{3.24}$$

The characteristic function

The characteristic function is defined in a similar way

$$M(\alpha) = \langle e^{j\alpha a} \rangle = \int e^{j\alpha a} P(a) \mathrm{d}a, \tag{3.25}$$

where $P(a) = |F(a)|^2$ is the distribution (Cohen, 1993, 1995). The distribution $P(a)$ can be shown to be related to the characteristic function according to

$$P(a) = \frac{1}{2\pi} \int M(\alpha) e^{-j\alpha a} \mathrm{d}\alpha. \tag{3.26}$$

Clearly the distribution $P(a)$ and the characteristic function $M(\alpha)$ form a Fourier pair. In the operator domain equation (3.25) reads

$$M(\alpha) = \int f^*(x) e^{j\alpha \mathcal{A}} f(x) \mathrm{d}x. \tag{3.27}$$

Mean and variance

For later purposes it is necessary to introduce quantities that express the mean and variance of a physical quantity. The mean of the physical quantity a pertaining to the operator $\mathcal{A}$ is given by

$$\langle \mathcal{A} \rangle = \int f(x)^* \mathcal{A} f(x) \mathrm{d}x = \int a |F(a)|^2 \mathrm{d}a \tag{3.28}$$

while its variance is given by

$$A = \sigma_a^2 = \langle \mathcal{A}^2 \rangle - \langle \mathcal{A} \rangle^2 = \int f(x)^* (\mathcal{A} - \langle \mathcal{A} \rangle)^2 f(x) \mathrm{d}x \tag{3.29}$$

$$= \int (a - \langle a \rangle)^2 |F(a)|^2 \mathrm{d}a. \tag{3.30}$$

Examples

Anticipating on the discussion on the space representation let me apply the above averages.

mean space and spatial extent:

The mean space is defined as

$$\langle \mathcal{X} \rangle = \int F(k)^* \mathcal{X} F(k) \mathrm{d}k = \int x |f(x)|^2 \mathrm{d}x, \tag{3.31}$$

with $F(k)$ representing the spectral representation of $f(x)$ where k refers to the spatial frequency.

The mean spatial extent is defined as

$$X = \sigma_x^2 = \langle \mathcal{X}^2 \rangle - \langle \mathcal{X} \rangle^2 = \int F(k)^* (\mathcal{X} - \langle \mathcal{X} \rangle)^2 F(k) \mathrm{d}k \tag{3.32}$$

$$= \int (x - \langle x \rangle)^2 |f(x)|^2 \mathrm{d}x. \tag{3.33}$$

mean frequency and bandwidth:
The mean frequency[3] is given by

$$\langle \mathcal{K} \rangle = \int f(x)^* \mathcal{K} f(x) \mathrm{d}x = \int k |F(k)|^2 \mathrm{d}k. \tag{3.34}$$

The mean bandwidth is defined as

$$K = \sigma_k^2 = \langle \mathcal{K}^2 \rangle - \langle \mathcal{K} \rangle^2 = \int f(x)^* (\mathcal{K} - \langle \mathcal{K} \rangle)^2 f(x) \mathrm{d}x \tag{3.35}$$

$$= \int (k - \langle k \rangle)^2 |F(k)|^2 \mathrm{d}k. \tag{3.36}$$

3.2.6 Uncertainty principle for two operators

As soon as two operators associated with two physical quantities do ***not*** commute the *uncertainty principle* enters into the game. This principle expresses the fundamental disability to obtain "localized[4]" information on the two variables concurrently, e.g. it is impossible to obtain, at the same instant, information on the velocity and momentum of a particle or to acquire optimal localization in the space and spatial frequency domain (Gabor, 1946; Cohen, 1995). This apparent ambiguity can be estimated, a notion laid down in the Heisenberg/Gabor uncertainty relation

$$\sigma_a^2 \sigma_b^2 \geq \frac{1}{2} |[\mathcal{A}, \mathcal{B}]|, \tag{3.37}$$

where $\mathcal{A}$ and $\mathcal{B}$ constitute the operators corresponding to the physical variables a and b and where the variances are defined using the results of the previous section, i.e.

$$\sigma_a^2 = \langle \mathcal{A}^2 \rangle - \langle \mathcal{A} \rangle^2 \quad \text{and} \quad \sigma_b^2 = \langle \mathcal{B}^2 \rangle - \langle \mathcal{B} \rangle^2. \tag{3.38}$$

3.3 The space representation

In chapters 2 and 8 it was shown that the multiscale analysis is based on two fundamental operations namely translations and dilatations. In this section ample attention is paid to the translation or shift operation constituting the map

$$S_\zeta : \quad f(x) \longmapsto f(x + \zeta), \tag{3.39}$$

[3] See below for the definition of $\mathcal{K}$.
[4] Localized in the sense that the variance associated with the physical quantity becomes small.

where ζ refers to either a spatial or temporal shift depending on whether x refers to space or time. Now the question is can one find an operator that conducts the translation operation? If it exists what are its properties, which eigenvalue problem is involved and what kind of other properties can one associate with it?

Intuitively speaking one can argue that the shift operator must be unitary because the norm is conserved while shifting, i.e. $\|f(x)\|_2 = \|f(x+\zeta)\|_2$. Moreover an inverse must exist, mapping the shifted function back to its original location. Clearly such an inverse operation exists because

$$S_{-\zeta}: \quad f(x+\zeta) \longmapsto f(x), \tag{3.40}$$

will certainly hold by invoking the proper substitution of variables. What remains now is to find a Hermitian operator whose action in the exponent yields the required shift, i.e.

$$S_{\zeta}: \quad f(x) \longmapsto e^{j\mathcal{A}} f(x) = f(x+\zeta). \tag{3.41}$$

To find this unknown operator $\mathcal{A}$ one can use the Taylor series, as proposed in section 3.2.2, to expand the operator as well as the function $f(x+\zeta)$. Then via comparison one can easily show that the operator $\mathcal{A}$ has to equal

$$\mathcal{A} = \zeta\mathcal{K} \tag{3.42}$$

with

$$\mathcal{K} = \frac{1}{j}\frac{\mathrm{d}}{\mathrm{d}x}. \tag{3.43}$$

This operator, $\mathcal{K}$, is known as the spatial frequency operator and it equates

$$e^{j\mathcal{A}} f(x) = \sum_{n=0}^{\infty} \frac{(j)^n}{n!}\mathcal{A}^n f(x) = \sum_{n=0}^{\infty} \frac{\zeta^n}{n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n} f(x) = f(x+\zeta),$$

where this specific choice for the spatial frequency operator becomes manifest since it equates the Taylor expansion of the operator exponential with the Taylor expansion of $f(x+\zeta)$. Notice that the function $f(x)$ is assumed to live in the proper functional space, a notion amounting to demanding that $f \in \mathcal{D}$ or $f \in \mathcal{S}$, i.e. being infinitely differentiable and of compact support or of rapid descent, see chapter 5. Of course this condition can be relaxed but still the above operator only refers to regular functions and requires a distributional interpretation when $f(x)$ is a generalized function such as the delta distribution.

To summarize, the shift operation is defined as

$$S_{\zeta}: \quad f(x) \longmapsto \mathcal{S}_{\zeta} f(x) = f(x+\zeta) \tag{3.44}$$

with $\mathcal{S}_\zeta = e^{j\zeta\mathcal{K}}$, while its inverse is found via

$$S_{-\zeta}: \quad f(x+\zeta) \longmapsto \mathcal{S}_\zeta^{-1} f(x+\zeta) = f(x), \tag{3.45}$$

where $\mathcal{S}_{-\zeta} = (\mathcal{S}_\zeta)^{-1} = e^{-j\zeta\mathcal{K}}$. Finally notice that it is not too strange that the $\frac{1}{j}\frac{\mathrm{d}}{\mathrm{d}x}$ operator acts as a shift operator in the exponent since it expresses the infinitesimal change of a function induced by an infinitesimal shift operation, an infinitesimal displacement.

3.3.1 The eigenvalue problem

To find the spectral representation for functions of operators, $\mathcal{Q}(\mathcal{A})$, one has to solve the eigenvalue problem for the operator occurring as the dependent "variable". In this case this is the frequency operator, $\mathcal{K}$. It is known that $\mathcal{K}$ is Hermitian so the eigenvalues, k, are real and the eigenfunctions complete and orthonormal. The eigenvalue problem reads

$$\mathcal{K}u(k,x) = ku(k,x), \tag{3.46}$$

for which the eigenfunctions, $u(k,x)$, equating this relation are of the form

$$u(k,x) = \frac{1}{\sqrt{2\pi}} e^{jkx}. \tag{3.47}$$

The eigenvalues, k, are referred to as the physical variable spatial frequency. It is clear that the above eigenfunctions adhere to equations (3.11) and (3.12).

3.3.2 The spectral representation

Since the eigenvalue problem is solved one can now define the expansion of the function f onto its spectral representation that reads

$$f(x) = \frac{1}{\sqrt{2\pi}} \int F(k) e^{jkx} \mathrm{d}k, \tag{3.48}$$

where

$$F(k) = \langle f(x), u(k,x) \rangle = \langle f(x), \frac{1}{\sqrt{2\pi}} e^{jkx} \rangle. \tag{3.49}$$

Here the quantity $F(k)$ refers to the spatial Fourier transform of $f(x)$ and can also be written as

$$F(k) = \mathcal{F}\{f, u(k,x)\}(k) = \langle f, u(k,x) \rangle. \tag{3.50}$$

Given the basis functions, yielded by the solution of the eigenvalue problem, one can find the Fourier spectral representation of the shift operator that is of the form

$$\begin{aligned} S_\zeta: \quad f(x) \longmapsto \mathcal{S}_\zeta f(x) &= e^{j\zeta\mathcal{K}} \int F(k) u(k,x) \mathrm{d}k \\ &= \int F(k) e^{j\zeta k} u(k,x) \mathrm{d}k = f(x+\zeta) \end{aligned} \tag{3.51}$$

which can be seen as

$$\mathcal{S}_\zeta f(x) = \mathcal{F}^{-1} e^{j\zeta k} \mathcal{F} f(x) = f(x+\zeta), \tag{3.52}$$

where the symbols $\mathcal{F}$ and $\mathcal{F}^{-1}$ refer to the forward and inverse Fourier transforms and $e^{j\zeta k}$ to the spectral representation of the shift operator. Finally the reader is referred to chapter 5 where I list some more properties of the Fourier transform especially in the context of tempered distributions.

3.3.3 Some properties of the space and spatial frequency operator

In this section I will review a selection of properties and introduce a number of useful quantities which emerge in the space and its dual, spatial frequency, representation.

Uncertainty principle for space and spatial frequency

For the specific case where ambiguity between the space and spatial frequency are are concerned the uncertainty relation of equation (3.37) becomes

$$\sigma_x^2 \sigma_k^2 \geq \frac{1}{2} \quad \text{or} \quad XK \geq \frac{1}{2}, \tag{3.53}$$

where use is made of

$$[\mathcal{X}, \mathcal{K}] = j$$

with

$$\mathcal{X} = -\frac{1}{j}\frac{\mathrm{d}}{\mathrm{d}k} \quad \text{and} \quad \mathcal{K} = \frac{1}{j}\frac{\mathrm{d}}{\mathrm{d}x}, \tag{3.54}$$

denoting the space operator in the frequency representation, and the spatial frequency operator respectively.

To summarize the uncertainty principle expresses the fact that when two operators commute one can not simultaneously resolve both quantities up to arbitrary small effective "bandwidth".

Linear shift invariant systems

The spatial derivative operator ∂_x or equivalently the spatial frequency operator $\mathcal{K} = \frac{1}{j}\frac{\mathrm{d}}{\mathrm{d}x}$ and their spectral representations are intimately related to the shift operator as became clear from the preceding sections. This observation has an important implication when systems are concerned that display an invariance with respect to this shift operation. If these systems indeed display a shift-invariance then one can find, via the principle of superposition, a link between the system's in- and output explicitly formulated in terms of the system's impulse response. Formally this relation boils down to an expression of the form

$$\mathcal{L}u = \mathcal{L}\{u\}(x) = f(x), \tag{3.55}$$

where u refers to the systems input and f to the output and $\mathcal{L}$ to the linear system operator of the form $\mathcal{L} = \sum_i (\frac{1}{j}\frac{\mathrm{d}}{\mathrm{d}x})^i$. The impulse response is defined as

$$\mathcal{L}\{\delta\}(x) = h(x) \tag{3.56}$$

and the system is denoted as shift invariant if for any x'

$$\mathcal{L}\{\mathcal{S}_{-x'}\delta\}(x) = \mathcal{S}_{-x'}h(x) \qquad \forall x' \in \mathbb{R}. \tag{3.57}$$

In that circumstance the output can be written in terms of a continuous superposition which becomes manifest in the convolution integral,

$$f(x) = \int h(x - x')u(x')\mathrm{d}x' = (h * u)(x). \tag{3.58}$$

Using the spectral representation for the shift operator, equation (3.51), one can easily derive that the convolution product of equation (3.58) becomes a simple multiplication in the spectral domain

$$F(k) = H(k)U(k) \tag{3.59}$$

where $F(k)$, $H(k)$ and $U(k)$ represent the system's output, impulse response and input respectively. Finally notice that the convolution is invariant under shifts,

$$((\mathcal{D}_\zeta h) * u)(x) = (h * (\mathcal{D}_\zeta u))(x) = \mathcal{D}_\zeta(h * u)(x), \tag{3.60}$$

and that it already emerged in the definition of the forward and inverse wavelet transform, see chapters 2 and 8.

3.4 The scale representation

Besides the translation operation the multiscale analysis of chapter 2 is based on the operation of dilation, also known as an affine transformation. Therefore I will, in a similar fashion as I did for the shift operation, review the predominant features related to the dilatation operation constituting the map

$$D_\sigma : f(x) \longmapsto \sqrt{\sigma} f(\sigma x) \quad \sigma > 0, \tag{3.61}$$

where σ refers to an arbitrary dilatation, i.e. a compression when $\sigma > 1$ or a stretching when $\sigma < 1$. The question arising concerns whether one can find an operator which does the job and what the properties are of the pertaining eigenvalue problem and the spectral representation?

From the definition of the dilatation operation one can see that it conserves energy since $\|f(x)\|_2 = \|\sqrt{\sigma} f(\sigma x)\|_2$ while on the other hand there must exist an inverse since the dilatation can be counterbalanced by the inverse operation

$$D_{\frac{1}{\sigma}} : \sqrt{\sigma} f(\sigma x) \longmapsto f(x) \quad \sigma > 0. \tag{3.62}$$

This again leads one to conclude that the ruling operator must be unitary[5], i.e. it must be of the type $e^{j\mathcal{A}}$ where $\mathcal{A}$ is the unknown but Hermitian operator. How can one find this operator?

As with the shift operation the Taylor expansion comes at hand to find the operator $\mathcal{A}$ in the exponent. The difference here lies that one has to solve a repeated eigenvalue problem in order to match

$$\sum_{n=0}^{\infty} \frac{(j)^n}{n!} \mathcal{A}^n f(x) = \sqrt{\sigma} f(\sigma x). \tag{3.63}$$

Following Cohen (1995) one can show that the operator

$$\mathcal{A} = \ln \sigma \mathcal{C}_x \tag{3.64}$$

equates the above equation when

$$\mathcal{C}_x = \frac{1}{2}(\mathcal{X}\mathcal{K} + \mathcal{K}\mathcal{X}) = \frac{1}{2}[X, K]_+ \tag{3.65}$$

with $[\cdot, \cdot]_+$ denoting the anti-commutator. In the space domain this operator is given by

$$\mathcal{C}_x = \frac{1}{2j}(x\frac{\mathrm{d}}{\mathrm{d}x} + \frac{\mathrm{d}}{\mathrm{d}x}x) \tag{3.66}$$

or, alternatively, using the commutator relation for space and spatial frequency $[X, K] = j$, by

$$\mathcal{C}_x = \mathcal{X}\mathcal{K} - \frac{1}{2}j = \mathcal{K}\mathcal{X} + \frac{1}{2}j \tag{3.67}$$

while in the spatial frequency domain the scale operator equals

$$\mathcal{C}_k = \frac{1}{2j}(k\frac{\mathrm{d}}{\mathrm{d}k} + \frac{\mathrm{d}}{\mathrm{d}k}k). \tag{3.68}$$

Why does this ansatz for $\mathcal{C}_x$, defining the map

$$D_\sigma : f(x) \longmapsto e^{j \ln \sigma \mathcal{C}_x} f(x) = \sqrt{\sigma} f(\sigma x), \tag{3.69}$$

work? To prove this one has to recognize that the scale operator solves the eigenvalue problem

$$\mathcal{C}_x x^n = -j(n + \frac{1}{2})x^n \tag{3.70}$$

[5]The careful reader may be puzzled here because I opted to define the dilatation operation along the lines of equation (3.61) rather than via the map $D_\sigma : f(x) \mapsto e^\sigma f(e^\sigma x)$ where $D_{-\sigma}$ would have yielded the inverse. The reason for this it that the wavelet transform has been defined with dilatation operator of the type I opted to use.

while the same applies to the repeated invocation of $\mathcal{C}_x$, i.e.

$$\mathcal{C}_x^k x^n = (-j)^k (n + \frac{1}{2})^k x^n. \tag{3.71}$$

Now going back to equation (3.63) one can, using equations (3.70) and (3.71), write

$$\begin{aligned} e^{j \ln \sigma \mathcal{C}_x} x^n &= \sum_{k=0}^{\infty} \frac{(j \ln \sigma)^k}{k!} \mathcal{C}_x^k x^n = \sum_{k=0}^{\infty} \frac{(j \ln \sigma)^k}{k!} (-j)^k (n + \frac{1}{2})^k x^n \\ &= e^{(n+\frac{1}{2}) \ln \sigma} x^n \end{aligned}$$

while expanding $f(x)$, around the origin, yields for the operator's action

$$\begin{aligned} e^{j \ln \sigma \mathcal{C}_x} f(x) &= e^{j \ln \sigma \mathcal{C}_x} \sum_{n=0}^{\infty} a_n x^n = e^{\frac{1}{2} \ln \sigma} \sum_{n=0}^{\infty} a_n e^{n \ln \sigma} x^n \\ &= e^{\frac{1}{2} \ln \sigma} \sum_{n=0}^{\infty} a_n (\sigma x)^n = \sqrt{\sigma} f(\sigma x), \end{aligned}$$

which proves the ansatz.

To summarize, the dilatation operation is defined as follows

$$D_\sigma : f(x) \longmapsto \mathcal{D}_\sigma f(x) = \sqrt{\sigma} f(\sigma x) \quad \forall \sigma > 0 \tag{3.72}$$

with $\mathcal{D}_\sigma \triangleq e^{j \ln \sigma \mathcal{C}_x}$ and its inverse is found via $\mathcal{D}_\sigma^{-1} \triangleq e^{-j \ln \sigma \mathcal{C}_x}$ yielding,

$$D_{\frac{1}{\sigma}} : \sqrt{\sigma} f(\sigma x) \longmapsto \mathcal{D}_\sigma^{-1} \sqrt{\sigma} f(\sigma x) = f(x) \quad \forall \sigma > 0. \tag{3.73}$$

Finally I would like to remark that intuitively speaking it is not that surprising that the $\mathcal{C}_x$ operator emerges when deriving the dilatation operator. This scale operator was namely also used to express the difference induced by invoking an infinitesimal change in the scale indicator of the smoothing function of chapter 2 and led to the definition of the wavelet transform (Holschneider, 1995). Moreover the scale operator forms the basis in the Renormalization Group theory where one is interested in solving equations of the kind (Wilson, 1983; Nottale, 1995, 1996)

$$x \partial_x f(x) = \beta(f(x)) \quad + \quad \text{initial conditions.} \tag{3.74}$$

In the pertaining sections I will review the basic properties of the scale operator. After that in the final section of this chapter I will bring the scale operator not only in the context of the multiscale analysis presented in chapter 2 being substantiated in chapter 8 but also in the context of fractals.

3.4.1 The eigenvalue problem

By solving the eigenvalue problem for the scale operator one opens the way for a spectral representation for this operator. This representation provides a better insight into the action of the scale operator and the action of functions of this operator, such as the dilatation operator given by the generator (Reed and Simon, 1978; Pearson, 1988) which is defined by the operator exponent. Furthermore the spectral representation offers a way to unravel the scale content of a signal by expanding the function into the orthonormal basis yielded by the eigenfunctions.

The eigenvalue problem for the scale operator[6] reads

$$\mathcal{C}\gamma(c,x) = c\gamma(c,x), \tag{3.75}$$

the solution of which is obtained by choosing the eigenfunctions as (Cohen, 1993, 1995)

$$\gamma(c,x) = \frac{1}{\sqrt{2\pi}}\frac{e^{jc\ln x}}{\sqrt{x}}, \quad x \geq 0, \tag{3.76}$$

where the c's refer to the eigenvalues, to the physical quantity scale and where I used a new symbol $\gamma(c,x)$ to denote the scale eigenfunctions. It is easily verified that the above eigenfunctions adhere to equations (3.11) and (3.12). So they form a complete and orthonormal basis for $x \geq 0$. This latter condition on the space x can be understood by taking into consideration that the dilatations inducing the compressions or expansions are obtained by multiplying the space x by a positive number σ while the logarithm emerges quite naturally because it puts an equal footing on the range from zero to one, linked to an enlargement, and from one to infinity, linked to a reduction (Cohen, 1993, 1995; Nottale, 1995). Before I discuss the spectral representation in terms of the basis of eigenfunctions defined in equation (3.76) I would like to compare some of their main properties compared to the eigenfunctions found for the frequency operator, i.e. the ones defined in equation (3.47).

3.4.2 The spectral representation

Given the solution of the eigenvalue problem one is now set to expand arbitrary but proper[7] functions onto the orthogonal and complete basis of eigenfunctions that belong to the scale operator $\mathcal{C}$. The expansion of the function f onto its spectral representation reads

$$f(x) = \frac{1}{\sqrt{2\pi}}\int F(c)\frac{e^{jc\ln x}}{\sqrt{x}}\mathrm{d}c, \qquad x \geq 0, \tag{3.77}$$

[6] For convenience I dropped the subscript x in the scale operator.

[7] I will refrain from issuing precise statements on the nature of the functional space in which these function have to live in order to make the expansion valid.

Scale kernel: $\gamma(c,x)$	Frequency kernel: $u(k,x)$
$\gamma(c,xx') = \gamma(c,x)\gamma(c,x')$	$u(k,x+x') = u(k,x)u(k,x')$
$\gamma(c,x/x') = \gamma(c,x)\gamma^*(c,x')$	$u(k,x-x') = u(k,x)u^*(k,x')$
$\gamma(c+c',x) = \gamma(c,x)\gamma(c',x)$	$u(k+k',x) = u(k,x)u(k',x)$
$\partial_c\gamma(c,x) = j\ln x\gamma(c,x)$	$\partial_k u(k,x) = jxu(k,x)$
$\partial_x\gamma(c,x) = \frac{jc-1/2}{x}\gamma(c,x)$	$\partial_x u(k,x) = jku(k,x)$
$e^{j\sigma c}\gamma(c,x) = e^{\sigma/2}\gamma(c,e^\sigma x)$	$e^{j\sigma k}u(k,x) = u(k+\sigma,x)$

Table 3.1 *Table with the properties of the kernels defined by the eigenfunctions of the scale and frequency operator.*

where

$$F(c) = \langle f(x), \gamma(c,x)\rangle = \langle f(x), \frac{1}{\sqrt{2\pi}}\frac{e^{jc\ln x}}{\sqrt{x}}\rangle \tag{3.78}$$

with the inner product running over the interval $x \in [0,\infty)$. Here the quantity $F(c)$ refers to the spatial Mellin transform (Dautray and Lions, 1988; Cohen, 1993; Holschneider, 1995; Cohen, 1995) of $f(x)$ and can also be written as

$$F(c) = \mathcal{M}\{f, \frac{1}{\sqrt{2\pi}}\frac{e^{jc\ln x}}{\sqrt{x}}\}(c) = \langle f, \frac{1}{\sqrt{2\pi}}\frac{e^{jc\ln x}}{\sqrt{x}}\rangle. \tag{3.79}$$

Given the spectral representation one can find an alternative proof that substantiates the action of the dilatation operator by using the scale-invariance property displayed by the eigenfunctions of the scale operator, see table 3.1,

$$e^{j\sigma c}\gamma(c,x) = e^{\sigma/2}\gamma(c,e^\sigma x)$$

or

$$\gamma(c,\sigma x) = \sigma^{jc-1/2}\gamma(c,x).$$

In chapters 2 and 7 the notion of scale-invariance displayed by homogeneous distributions and monofractals was already observed. Using this property one can easily prove the

action of the dilatation operator to be

$$\begin{aligned} e^{j\ln\sigma C} f(x) &= e^{j\ln\sigma C} \int F(c)\gamma(c,x)\mathrm{d}c \\ &= \int F(c) e^{j\ln\sigma c}\gamma(c,x)\mathrm{d}c \\ &= \int F(c)\sqrt{\sigma}\gamma(c,\sigma x)\mathrm{d}c = \sqrt{\sigma} f(\sigma x). \end{aligned}$$

The spectral representation for the dilatation operator becomes

$$\begin{aligned} D_\sigma : \quad f(x) \longmapsto \mathcal{D}_\sigma f(x) &= e^{j\ln\sigma C} \int F(c)\gamma(c,x)\mathrm{d}c \\ &= \int F(c) e^{j\ln\sigma c}\gamma(c,x)\mathrm{d}c = \sqrt{\sigma} f(\sigma x) \end{aligned} \tag{3.80}$$

which can be seen as

$$\mathcal{D}_\sigma f(x) = \mathcal{M}^{-1} e^{j\ln\sigma c} \mathcal{M} f(x) = \sqrt{\sigma} f(\sigma x), \tag{3.81}$$

where the symbols $\mathcal{M}$ and $\mathcal{M}^{-1}$ refer to the forward and inverse Mellin transforms and $e^{j\ln\sigma c}$ refers to the spectral representation of the dilatation operator. Now let me list some of the important properties connected to the scale transformation.

3.4.3 Some properties of the scale operator

Commutation relations

One of the most important observations to me made concerning the scale operator lies in the fact that it does ***not*** commute with space, i.e.

$$[\mathcal{X}, C] = j\mathcal{X} \tag{3.82}$$

in space and

$$[\mathcal{K}, C] = j\mathcal{K} \tag{3.83}$$

in spatial frequency. This means that one can not solve the eigenvalue problem for scale and space/spatial frequency operators concurrently. When the logarithm of space versus scale is concerned one finds

$$[\ln\mathcal{X}, C] = j \tag{3.84}$$

whose outcome exactly matches the commutation relation for space and spatial frequency.

Linear scale invariant systems

In a similar fashion as with the linear shift invariant systems one can find a formulation referring to scale invariant systems that are intimately related to the scale operator $\mathcal{C}$. The invariance now concerns itself with respect to the dilatation operation and makes it possible, via the principle of superposition, to establish a connection between the system's in- and output explicitly formulated in terms of the system's impulse response. Formally the relation between the input and output can be written as

$$\mathcal{L}u = \mathcal{L}\{u\}(x) = f(x), \tag{3.85}$$

where u refers to the systems input and f to the output and $\mathcal{L}$ to the linear system operator. The impulse response is now defined as

$$\mathcal{L}\{\delta(x-1)\}(x) = h(x) \tag{3.86}$$

and the system is denoted as linear scale invariant if for any x'

$$\mathcal{L}\{\mathcal{S}_{-x'}\delta\}(x) = \mathcal{D}_{1/x'}h(x) \qquad \forall x' \in \mathbb{R} \tag{3.87}$$

or equivalently

$$\mathcal{L}\{\delta(x-x')\}(x) = \frac{1}{x'}h(\frac{x}{x'}) \qquad \forall x' \in \mathbb{R}. \tag{3.88}$$

In that circumstance the output can be written in terms of a continuous superposition which becomes manifest in the scale convolution integral,

$$f(x) = \int h(\frac{x}{x'})u(x')\frac{\mathrm{d}x'}{x'} = \int h(x')u(\frac{x}{x'})\frac{\mathrm{d}x'}{x'} = (f *_l u)(x), \tag{3.89}$$

with $*_l$ denoting the scale convolution. This scale convolution, also known as a logarithmic convolution[8], becomes a simple multiplication in the domain given by the spectral representation of the scale operator, the Mellin/scale domain, i.e.

$$\mathcal{M}\{f\}(c) = \mathcal{M}\{h\}(c)\mathcal{M}\{u\}(c) \tag{3.90}$$

or

$$F(c) = H(c)U(c), \tag{3.91}$$

where $\mathcal{M}\{f\}(c)$, $\mathcal{M}\{h\}(c)$ and $\mathcal{M}\{u\}(c)$ represent, in the Mellin domain, the system's output, impulse response and input respectively. Finally notice that the scale convolution is invariant under dilatations

$$((\mathcal{D}_\sigma h) *_l u)(x) = (h *_l (\mathcal{D}_\sigma u))(x) = \mathcal{D}_\sigma(h *_l u)(x). \tag{3.92}$$

and that it already emerged in the definition of the inverse wavelet transform, see chapter 8.

[8] By setting $y = \ln x$ one obtains an ordinary convolution (Holschneider, 1995).

3.5 Numerical implementation

In this section I will review the relevant aspects related to the actual numerical implementation that goes with the space and scale representation. While doing so I will limit myself to a discussion of the eigenvalue problems for the frequency and scale operators, $\mathcal{K}$ and $\mathcal{C}$, and their associated actions of the generators, $e^{j\zeta\mathcal{K}}$ and $e^{j\ln\sigma\mathcal{C}}$, invoking the shift and dilation operations.

3.5.1 Discretization of the operator

The action of an operator runs via its kernel in the following fashion

$$g(x) = \mathcal{A}\{f\}(x) = \int_{-\infty}^{+\infty} \mathrm{A}(x,x')f(x')\mathrm{d}x', \tag{3.93}$$

where $\mathrm{A}(x,x')$ denotes the kernel of the operator over which the generalized convolution integral runs while $f(x)$ refers to a proper function undergoing the action of the operator. This convolution integral finds its discretized counterpart in the following *matrix multiplication*

$$\mathbf{g} = \mathbf{A}\mathbf{f}. \tag{3.94}$$

Here the symbols $\mathbf{g}$ and $\mathbf{f}$ refer to the vectors representing the discretized versions of the functions $g(x)$, $f(x)$, i.e.

$$\mathbf{g} = \begin{pmatrix} g(\mathrm{x}_1) \\ g(\mathrm{x}_2) \\ g(\mathrm{x}_3) \\ \vdots \\ g(\mathrm{x}_n) \end{pmatrix} \quad \text{and} \quad \mathbf{f} = \begin{pmatrix} f(\mathrm{x}_1) \\ f(\mathrm{x}_2) \\ f(\mathrm{x}_3) \\ \vdots \\ f(\mathrm{x}_n), \end{pmatrix} \tag{3.95}$$

with x denoting the discretization of x, i.e. $\mathrm{x}_i = (i-1)\Delta x, \quad i = 1\cdots n$, where Δx is the discretization interval and n the number of grid points, the size of the discretization. From now on I set, unless stated otherwise, the discretization interval to $\Delta x = 1$. After discretization the kernel A is represented by a $n \times n$ matrix

$$\mathbf{A} = \begin{pmatrix} \mathrm{A}(\mathrm{x}_1,\mathrm{x}_1) & \cdots & \mathrm{A}(\mathrm{x}_1,\mathrm{x}_n) \\ \vdots & \ddots & \vdots \\ \mathrm{A}(\mathrm{x}_n,\mathrm{x}_1) & \cdots & \mathrm{A}(\mathrm{x}_n,\mathrm{x}_n) \end{pmatrix}. \tag{3.96}$$

This matrix constitutes an approximation of the actual operator. In cases where one has to deal with a singular operation, such as differentiation, one has to interpret the discretization of the operator in the sense that it corresponds to the discretization of

an finite indexed element of the Cauchy sequence. This Cauchy sequence embodies the sequence that in the limit defines the derivative operator, see chapter 5. To cut things short one has to interpret the approximation in the sense of distributions. Before going into the specific examples let me first go through the numerical solution of the eigenvalue problem.

3.5.2 Numerical solution of the eigenvalue problem

After discretization of the operator it is now time to set up the numerical solution to the eigenvalue problem pertaining to the matrix that approximates the operator's action. Let me write the eigenvalue problem in the following form

$$\mathbf{A}\mathbf{v}_i = \mathsf{a}_i\mathbf{v}_i, \tag{3.97}$$

where the $\mathbf{v}_i$'s denote the discretized eigenfunctions, now eigenvectors, labeled by the index i and pertaining to the i^{th} eigenvalue a_i. The discretized eigenvectors are of the form

$$\mathbf{v}_i = \begin{pmatrix} v_i(\mathsf{x}_1) \\ v_i(\mathsf{x}_2) \\ v_i(\mathsf{x}_3) \\ \vdots \\ v_i(\mathsf{x}_n) \end{pmatrix}. \tag{3.98}$$

Notice that equation (3.97) corresponds to the discretized version of

$$\mathcal{A}u(a,x) = au(a,x). \tag{3.99}$$

Since the eigenvalue problems being considered here are Hermitian one knows that $\mathcal{A} = \mathcal{A}^\dagger$, a property carried over by the matrix $\mathbf{A}$ which shares this property, this self-adjointness, stating that

$$\mathbf{A} = \mathbf{A}^H \tag{3.100}$$

with ^{H} indicating the concurrent transposition and complex conjugation. This means that one interchanges the elements of $\mathbf{A}$ together with a complex conjugation, i.e. $A_{ij} = A^*_{ji} \quad \forall i,j \in [1\cdots n]$, to obtain the transposition and complex conjugation ^{H}.

As a consequence of the Hermitism the eigenvalues a_i are real while the set of eigenvectors $\mathbf{v}_i, \quad \forall i \in [1\cdots n]$, can be shown to be complete and orthogonal (Golub and van Loan, 1984). Let me now recast the eigenvalue problem of equation (3.97) into a matrix formalism

$$\mathbf{AV} = \mathbf{VU} \tag{3.101}$$

where the eigenvectors $\mathbf{v}_i$ are combined in the matrix

$$\mathbf{V} = \begin{pmatrix} v_1(\mathsf{x}_1) & \cdots & v_n(\mathsf{x}_1) \\ \vdots & \ddots & \vdots \\ v_1(\mathsf{x}_n) & \cdots & v_n(\mathsf{x}_n) \end{pmatrix} \tag{3.102}$$

while the eigenvalues, a_i, find their way into a diagonal matrix,

$$\mathbf{U} = \begin{pmatrix} \mathsf{a}_1 & 0 & 0 & \cdots & 0 \\ 0 & \mathsf{a}_2 & 0 & \cdots & 0 \\ 0 & 0 & \mathsf{a}_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathsf{a}_n \end{pmatrix}. \tag{3.103}$$

3.5.3 The spectral representation

By way of how the numerical implementation of the eigenvalue problem has been set up it is straightforward to derive the spectral representation for the operator and its cognate operator functions such as the operator exponential.

The spectral representation corresponding to

$$\mathcal{A}f = \int F(a) a v(a,x) \mathrm{d}a, \tag{3.104}$$

see section 3.2.2, reads

$$\mathbf{Af} = \mathbf{VUV}^H\mathbf{f}, \tag{3.105}$$

where use is made of equation (3.100) and the property that, after the proper normalization[9], the eigenvector matrices adhere to the unitary relation $\mathbf{V}^{-1} = \mathbf{V}^H$. Similarly one can write for the spectral representation for functions of operators, i.e.

$$Q(\mathcal{A})v(a,x) = Q(a)v(a,x), \tag{3.106}$$

$$\mathbf{Qf} = \mathbf{V}Q(\mathbf{U})\mathbf{V}^H\mathbf{f} \tag{3.107}$$

with $Q(\mathbf{U})$ is being defined as

$$Q(\mathbf{U}) = \begin{pmatrix} Q(\mathsf{a}_1) & 0 & 0 & \cdots & 0 \\ 0 & Q(\mathsf{a}_2) & 0 & \cdots & 0 \\ 0 & 0 & Q(\mathsf{a}_3) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & Q(\mathsf{a}_n) \end{pmatrix} \tag{3.108}$$

[9]The power flux normalization since I am working in $L^2(\mathbb{R})$.

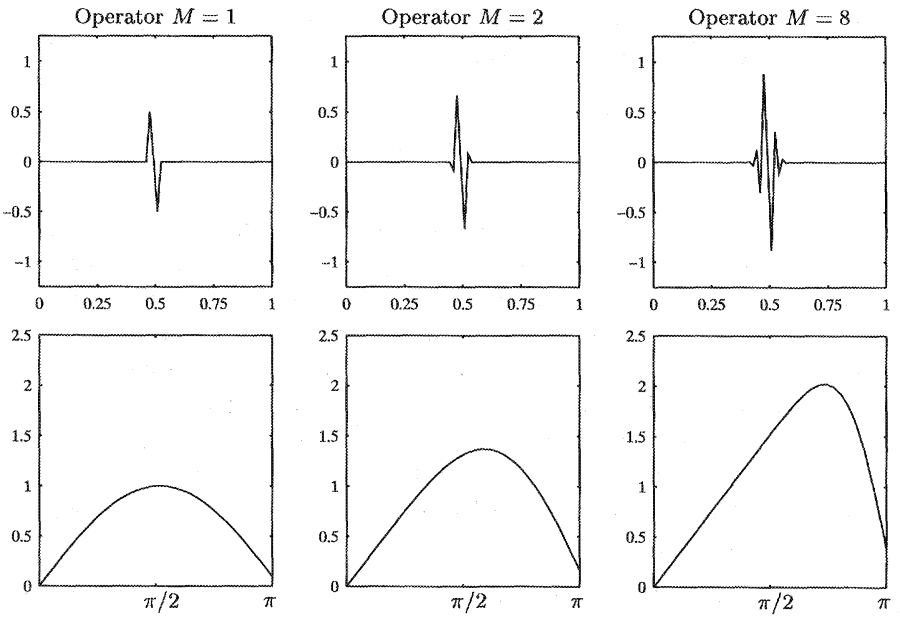

Figure 3.1 *Numerical implementation of the first order derivative operator. In the top row I displayed the approximations for the number of vanishing moments set to $M = 1$; $M = 2$ and $M = 8$. In the bottom row I displayed their spectral representation in the spatial Fourier domain.*

and where $Q(x)$ is an ordinary function.

Given this set of tools I am now set to numerically implement the shift and dilatation operators. Hereby I would like to mention that the spectrum of these operators is purely continuous so I do not expect problems there.

3.5.4 Example shift operator

In this section I would like to quickly review the numerical implementation of the shift operator that is defined in terms of the operator exponent

$$\mathcal{S}_\zeta = e^{j\zeta\mathcal{K}} \tag{3.109}$$

with

$$\mathcal{K} = \frac{1}{j}\frac{\mathrm{d}}{\mathrm{d}x}. \tag{3.110}$$

Inspection of this equation shows that one requires the numerical implementation of the first order derivative operator whose kernel can be written as

$$\mathrm{K}(x,x') = \frac{1}{j}\frac{\mathrm{d}}{\mathrm{d}x}\delta(x-x'). \tag{3.111}$$

Following Beylkin et al. (1991), section VII, it is possible to find numerical approximations to the first order derivative operator by expanding the kernel of equation (3.111) into a non-standard form for the discrete wavelet transform (Beylkin, 1992; Beylkin et al., 1992). The "free parameter" in his approach refers to the number of vanishing moments to be imposed on the discrete wavelets that are used to obtain the approximation. The higher this number the more regular the wavelets become. This gives rise to longer operators that display a linear behaviour persisting to higher frequencies in the Fourier domain[10], see figure 3.1. In the sequel I will make use of these coefficients which are symmetrically[11] put on the first l off-diagonals where l refers to the length of the approximated derivative operator.

With the help of figures 3.2 and 3.3 I make an attempt to elucidate the action of the shift operator. At first I depict the matrices[12] with the eigenvectors and the sandwiched spectral representation for the shift operator. The latter matrix is of course diagonal while the eigenvector matrices can be seen as matrices containing the numerical implementation of the unitary forward and inverse Fourier transform. So the kernel of the shift operator consists of the following entities

$$\mathbf{S}_\zeta = \mathbf{V}\mathbf{U}\mathbf{V}^H = \mathbf{F}^{-1}\mathbf{U}\mathbf{F}, \tag{3.112}$$

where the matrices $\mathbf{V}$ and $\mathbf{V}^H$ are interpreted as the inverse and forward Fourier transforms, i.e. $\mathbf{F}^{-1}$ and $\mathbf{F}$, and where the diagonal of $\mathbf{U}$ is given by $e^{j\zeta\mathrm{k}}$. In figure 3.3 I actually included the numerical implementation of the kernel which is simply obtained by raising the the matrix $j\zeta\mathbf{K}$ to the exponent. The $\mathbf{K}$ is set with the help of the numerical implementation of the spatial derivative operator multiplied by $\frac{1}{j}$. Hence one finds

$$\mathbf{S}_\zeta = e^{j\zeta\mathbf{K}}. \tag{3.113}$$

Clearly this matrix exponential has to be interpreted in the sense of equation (3.109).

[10] The derivative operator corresponds to multiplying by jk in the spatial Fourier domain.

[11] Notice that the matrix itself is not symmetrical only the values are symmetrically located around the diagonal. The matrix itself is anti-symmetric.

[12] In this I calculated the eigenfunctions and the spectral representation for the shift operator using equations (3.47) and (3.51). In figure 3.3 I directly compute the matrix exponential by numerically solving the eigenvalue problem.

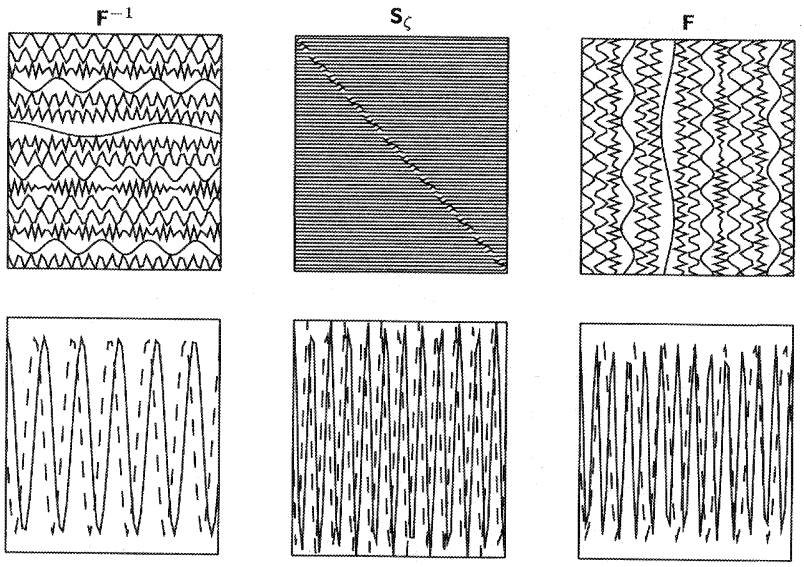

Figure 3.2 *Illustration of the different matrices involved in the spectral representation of the shift operator, see equations* (3.51) *and* (3.112) *for the numerical implementation. On the top row* (a)-(c) *one finds the real part of the matrices* $\mathbf{F}^{-1}$, $\mathbf{S}_\varsigma$ *and* $\mathbf{F}$, *containing the approximate inverse Fourier transform, the spectral representation of the shift operator and the forward Fourier transform respectively. In the second row,* (d)-(f) *I depicted a selection of traces of the matrices of* (a) *and* (c) *in the top row while in the middle one finds the diagonal of* (b). *Remark that in the lower row both the real and imaginary parts are on display.*

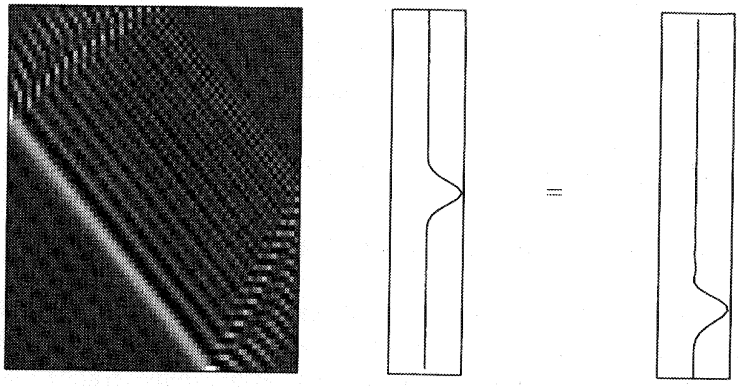

Figure 3.3 *In this figure the kernel of the shift operator is on display together with the result of its action on a Gaussian bell-shaped function* f.

3.5.5 Example dilatation operator

In a similar fashion as in the previous section 3.5.4 I will discuss the numerical implementation of the dilatation operator being defined in terms of the operator exponent

$$\mathcal{D}_\sigma = e^{j \ln \sigma \mathcal{C}} \tag{3.114}$$

with

$$\mathcal{C} = \frac{1}{2j}(x\frac{\mathrm{d}}{\mathrm{d}x} + \frac{\mathrm{d}}{\mathrm{d}x}x). \tag{3.115}$$

Inspection of this equation shows that one requires the numerical implementation of the first order derivative operator whose kernel can be written as

$$\mathrm{C}(x,x') = \frac{1}{2j}(x\frac{\mathrm{d}}{\mathrm{d}x}\delta(x-x') + \frac{\mathrm{d}}{\mathrm{d}x}\delta(x-x')x'). \tag{3.116}$$

With the help of figures 3.4 and 3.5 I make an attempt to elucidate the action of the dilatation operator. At first I depict the matrices[13] with the eigenvectors and the sandwiched spectral representation for the shift operator. The latter matrix is of course diagonal while the eigenvector matrices can be seen as matrices containing the numerical implementation of the unitary forward and inverse Mellin transform. So the kernel of the dilatation operator consists of the following entities

$$\mathbf{D}_\varsigma = \mathbf{V}\mathbf{U}\mathbf{V}^H = \mathbf{M}^{-1}\mathbf{U}\mathbf{M}, \tag{3.117}$$

where the matrices $\mathbf{V}$ and $\mathbf{V}^H$ are interpreted as the inverse and forward Mellin transforms, i.e. $\mathbf{M}^{-1}$ and $\mathbf{M}$, and where the diagonal of $\mathbf{U}$ is given by $e^{j \ln \sigma \mathrm{c}}$. In figure 3.5 I actually included the numerical implementation of the kernel. This kernel is simply obtained by raising the matrix $j \ln \sigma \mathbf{C}$ – with $\mathbf{C}$ set by means of the numerical implementation of the spatial scale derivative operator $j \ln \sigma \mathbf{C}$ – to the exponent, i.e.

$$\mathbf{D}_\sigma = e^{j \ln \sigma \mathbf{C}}. \tag{3.118}$$

Clearly this matrix exponential has to be interpreted in the sense of equation (3.114).

3.6 The scale representation versus fractals and multiscale analysis

I would like to conclude this chapter by clarifying the relevance of the scale representation with respect to the central concepts of fractals and the multiscale analysis. Indeed the motivation of this chapter primarily was to establish a robust framework for the shift

[13] Again I calculated the eigenfunctions and the spectral representation for the shift operator using equations (3.76) and (3.80). In figure 3.5 I directly compute the matrix exponential by numerically solving the eigenvalue problem.

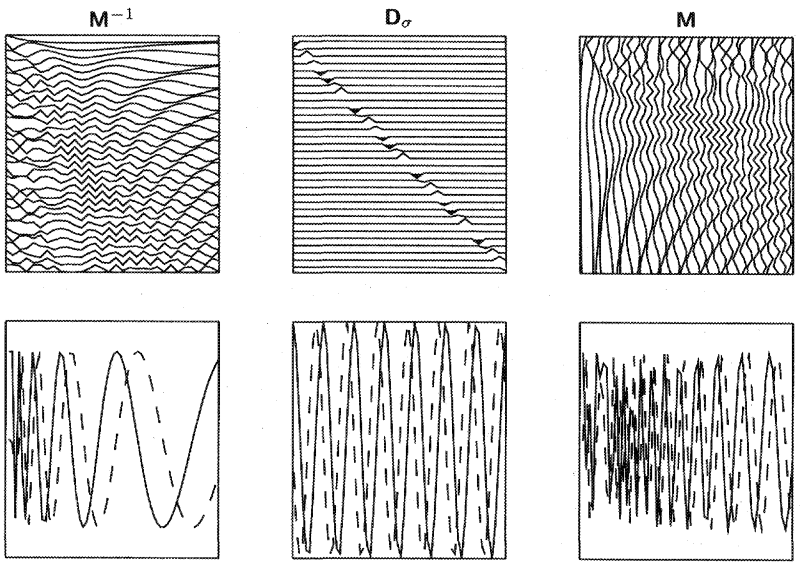

Figure 3.4 *Illustration of the different matrices involved in the spectral representation of the dilatation operator, see equations* (3.80) *and* (3.117) *for the numerical implementation. On the top row* (a)-(c) *one finds the real part of the matrices* $\mathbf{M}^{-1}$, $\mathbf{D}_\sigma$ *and* $\mathbf{M}$, *containing the approximate inverse Mellin transform, the spectral representation of the dilatation operator and the forward Mellin transform respectively. In the second row,* (d)-(f) *I depicted a selection of traces of the matrices of* (a) *and* (c) *in the top row while in the middle one finds the diagonal of* (b). *Remark that in the lower row both the real and imaginary parts are on display.*

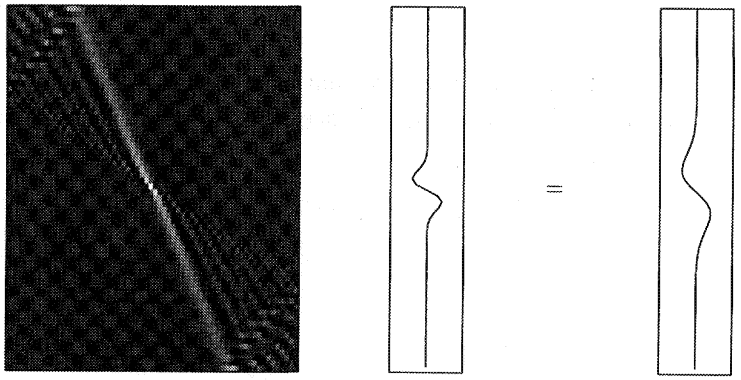

Figure 3.5 *In this figure the kernel of the dilatation operator is on display together with the result of its action on the first derivative of the Gaussian function,* f. *Notice that the dilatation actually took place.*

and dilatation operations which form the heart of the multiscale analysis and the characterization. But if one looks back to the contents of this chapter it is not hard to imagine that there must be a more fundamental correspondence between the scale representation and its cognate properties on the one hand and the fractal/singular representation, with the pertaining multiscale analysis, on the other. The reason for this observation lies in the fact that *the scaling exponents constitute eigenvalues of the scale operator.* It is this observation which inspired me to write this section.

I will start by establishing a link between the continuous wavelet transform of a homogeneous distribution, a distribution containing an isolated algebraic singularity, and the Mellin transform, the transform yielded by the basis of eigenfunctions acquired by solving the eigenvalue problem for the scale operator. Then I will continue the discussion by mentioning that the homogeneous scale-invariance displayed by monofractals must in one way or another be related to linear scale invariant systems while the heterogeneous multifractals can be associated to systems that no longer display a trivial scale-invariance. In that case the interesting question emerges on the relation between the convex singularity spectrum, from which one knows that it expresses the Hausdorff dimension to be associated with the set of points with scaling exponents on the infinitesimal interval $[\alpha, \alpha + \mathrm{d}\alpha]$, and the spectral c-domain associated with the scale operator.

3.6.1 Homogeneous distributions and the Mellin transform

Inspection of the relations pertaining to the local analysis by means of the continuous wavelet transform on homogeneous distributions – distributions containing isolated algebraic singularities – shows that one can establish the following connection (Holschneider, 1995)

$$\mathcal{W}\{f_+^\alpha, \psi(\frac{x-x'}{\sigma})\}(\sigma, x) \quad \Longleftrightarrow \quad \sigma^{1/2+\alpha}\mathcal{M}\{\psi(x-\frac{x'}{\sigma})\}(\frac{1}{j}(\alpha+1/2)), \tag{3.119}$$

where $f_+^\alpha = x_+^\alpha$ denotes the homogeneous distribution I introduced in chapter 5 and where the $\Longleftrightarrow$ is used to express the mutual equivalence between the left- and right-hand side. To see this I made use of

$$\int_0^\infty \frac{1}{\sqrt{\sigma}}\psi(\frac{x-x'}{\sigma})x^\alpha \mathrm{d}x = \sigma^{\alpha+1/2}\int_0^\infty \psi(x-\frac{x'}{\sigma})x^\alpha \mathrm{d}x \tag{3.120}$$

and

$$\mathcal{M}\{f\}(c) = \langle f, \frac{1}{\sqrt{2\pi}}\frac{e^{jc\ln x}}{\sqrt{x}}\rangle. \tag{3.121}$$

Notice that the above definition by Cohen (1995) of the forward Mellin transform deviates from the definitions given by Dautray and Lions (1988) and Holschneider (1995) in the sense that the exponents are complex rather than real.

Given the above observation a direct connection is made between the local scaling analysis by the continuous wavelet transform of the onset functions, distributions containing isolated algebraic singularities, and the Mellin transform. In fact it appears that the onset functions act as members of the Mellin transform kernel and therefore the outcome of the local analysis can be interpreted as a single component for the Mellin of the analyzing wavelet. This result is not surprising when one takes into consideration that homogeneous distributions constitute solutions to the eigenvalue problem,

$$x\frac{\mathrm{d}}{\mathrm{d}x}x^{\alpha} = \alpha x^{\alpha}, \tag{3.122}$$

which can be associated with

$$C\frac{e^{jc\ln x}}{\sqrt{x}} = c\frac{e^{jc\ln x}}{\sqrt{x}} \tag{3.123}$$

in the formulation used throughout this chapter and where c plays a similar role as the scaling exponent α. It is interesting to remark that these peculiarities point in the direction of the notion that the onset functions can be seen as functions with a large scale content(Cohen, 1993, 1995). This can be judged by the fact that the above eigenfunctions

$$F(c) = \mathcal{M}\{f\}(c) = \delta(c - c') \quad \text{when} \quad f(x) = \frac{1}{\sqrt{2\pi}}\frac{e^{jc\ln x}}{\sqrt{x}}. \tag{3.124}$$

On the other hand for signals with the lowest scale content one may refer to

$$F(c) = \frac{1}{\sqrt{2\pi}}\frac{e^{jc\ln x'}}{\sqrt{x'}}, \tag{3.125}$$

the inverse Mellin transform of which correspond to

$$f(x) = \sqrt{x'}\delta(x - x'). \tag{3.126}$$

Notice that this latter property is shared with the spatial Fourier representation where the Dirac's delta distributions produces a flat spectrum indicating a low frequency content.

3.6.2 Fractals and the scale operator

In the previous section the intricate relationship between the continuous wavelet transform of homogeneous distributions and the Mellin transform was revealed. Distributions of this type can be seen as eigenfunctions belonging to the scale operator where the scale exponents play the role of the eigenvalues. Now what can one say when fractals are concerned? It is known that fractal structures are composed of singularities. For monofractals the singularity structure is simple since they consist of singularities all of the same strength. For that reason the class of monofractals display a homogeneous scaling behaviour boiling down to a scale-invariance of the type

$$f(\sigma x) \overset{\mathrm{d}}{=} \sigma^{\alpha} f(x), \tag{3.127}$$

where α is the scaling exponent of the monofractal, f, and where the sign $\stackrel{\mathrm{d}}{=}$ refers to equality in distribution. That is to say that the probability density functions on either side of the $\stackrel{\mathrm{d}}{=}$ sign are the same yielding,

$$\langle (f(x))^q \rangle = \sigma^{-q\alpha} \langle f(\sigma x)^q \rangle. \tag{3.128}$$

In the language of this chapter these relations correspond to

$$\mathcal{D}_\sigma f = \sigma^\alpha f \tag{3.129}$$

and

$$\langle (\mathcal{D}_\sigma f)^q \rangle = e^{\ln \sigma q \alpha} \langle f^q \rangle. \tag{3.130}$$

Clearly this type of behaviour suggest that when dealing with monofractals one can think of them as linear scale invariant systems in a way that the monofractals act, in a stochastic sense, as processes that are invariant in the scale direction. This observation is quite natural since it strides well with the notion that there is only a single scaling exponent present in monofractals.

In the chapters dealing with multifractals, heterogeneous scaling, it was demonstrated that the scaling behaviour displayed by these constructs becomes much more involved because

$$\langle (\mathcal{D}_\sigma f)^q \rangle = e^{\ln \sigma \tau(q)} \langle f^q \rangle, \tag{3.131}$$

where the behaviour of the mass exponent function $\tau(q)$ is not longer linear in q. Moreover it was shown that the $\tau(q)$ function describes the scaling of the moments,

$$\langle (\mathcal{D}_\sigma f)^q \rangle \propto e^{\ln \sigma \tau(q)}, \tag{3.132}$$

and that it is related to the singularity spectrum $\mathrm{f}(\alpha)$. This singularity spectrum unravels the singularity structure of the multifractal and expresses the scaling displayed by the probability density for the scaling exponents, i.e. the probability to find a particular scaling exponent α. Its formal definition reads

$$\mathrm{f}(\alpha) = \dim_H \{x_0 \in \mathbb{R} | \alpha(x_0) = \alpha\}, \tag{3.133}$$

where $\dim_H$ denotes the Hausdorff dimension. To put it simple this definition is equivalent to expressing the following scaling for the probability to find a singularity with an strength in the interval $(\alpha, \alpha + \mathrm{d}\alpha)$, i.e.

$$\Pr[\sigma^\alpha < \check{f} < \sigma^{\alpha + \mathrm{d}\alpha}] \propto e^{-\ln \sigma \mathrm{f}(\alpha)}, \tag{3.134}$$

where $\check{f}(\sigma, x) = \mathcal{W}\{f, \psi_\sigma\}(\sigma, x) = \langle f, \psi_{\sigma,x} \rangle$. It can be shown that the singularity spectrum is related, via the Legendre transform, to the mass exponent function $\tau(q)$. In

fact what happened to arrive at this result is that the Mellin transform, relating the statistical moments to the probability density, has been approximated by a saddle point approximation. This was done by using the following proportionalities, valid within a slowly varying factor, $\mathcal{D}_\sigma\langle f,\phi_x\rangle \propto e^{\ln\sigma\alpha}$, $\Pr(\mathcal{D}_\sigma\langle f,\phi_x\rangle) \propto e^{-\ln\sigma \mathrm{f}(\alpha)}$ where ϕ_x denotes an appropriate smoothing kernel with a scale normalized to unity. For notational convenience I will replace $\langle f,\phi_x\rangle$ by f. The saddle point approximation now boils down to minimizing the exponent in

$$\langle(\mathcal{D}_\sigma f)^q\rangle = \int (\mathcal{D}_\sigma f)^q \mathrm{d}\Pr((\mathcal{D}_\sigma f)) \tag{3.135}$$

$$\propto \int e^{\ln\sigma(\alpha q-\mathrm{f}(\alpha))}\mathrm{d}\alpha \tag{3.136}$$

yielding

$$\tau(q) \quad = \quad \min_\alpha\{q\alpha-\mathrm{f}(\alpha)\}. \tag{3.137}$$

Be aware that the quantities $f(x)$ and $\mathrm{f}(\alpha)$ represent different things! Now how does this finding relate to the scale representation? To start with it is clear that the Mellin transform also comes into play when the spectral representation pertaining to the scale operator is concerned. This spectral representation reads

$$F(c) = \langle f,\gamma(c,x)\rangle. \tag{3.138}$$

It is known that this spectral quantity, when properly normalized, expresses the amount of information belonging to a specific value for the scale parameter and present in the signal $f(x)$. This interpretation is very close in line with the meaning of the singularity spectrum that expresses the relative rate of occurrence of a certain scaling exponent α. For that reason one can come forward with the argument that there has to be a direct relationship between the scale representation of the signal and its characterization by the singularity spectrum. At this point it would be too involved to robustly derive the mutual relation. Therefore I limit myself to postulate the conjecture that the following correspondences ought to hold

$$F(c) \Longleftrightarrow e^{-\ln\sigma \mathrm{f}(\alpha)} \quad \text{and} \quad c \Longleftrightarrow \alpha. \tag{3.139}$$

From this conjecture one can easily infer that the scale eigenvalues c play a similar role as the scaling exponents α. Finally I would like to remark that by way of its definition the singularity spectrum and its dependent variable α are independent of the scale indicator σ while the c depends on the scale indicator range.

3.6.3 Multiscale analysis and the scale operator

To conclude the discussion on the relevance of the scale representation in comparison to the concept of scaling used so far I like to spend a few more words on the mutual relationship between multiscale analysis and the scale representation.

Evidently there exists a close interrelation between the multiscale analysis and the scale operator judged by the fact that a slightly deviant version of the scale derivative was used to define the continuous wavelet transform or to define the family of double-indexed analyzing wavelets by means of the affine transformation. Besides the dilatation operation the wavelets required the shift operation. Therefore it is natural to think along the lines of a joint space-scale representation to be associated with a multiscale analysis based on the continuous wavelet transform. To establish this relationship consider the continuous wavelet transform as I introduced in chapter 2

$$\mathcal{W}\{f,\psi\}(\sigma,x) \triangleq -\sigma\partial_\sigma f(\sigma,x), \tag{3.140}$$

where $f(\sigma,x)$ was obtained via

$$f(x,\sigma) \triangleq (f * \phi_\sigma)(x). \tag{3.141}$$

How does this definition relate to the scale representation? To see this rewrite equation (3.140) into

$$\begin{aligned}\mathcal{W}\{f,\psi\}(\sigma,x) &\triangleq -\sigma\partial_\sigma f(x,\sigma)\\ &= -\sigma\partial_\sigma e^{-j\ln\sigma\mathcal{C}}(f*\phi)(x)\\ &= e^{-j\ln\sigma\mathcal{C}}j\mathcal{C}(f*\phi)(x)\\ &= (f*\psi_\sigma)(x),\end{aligned}$$

where the scale operator maps the smoothing kernel to the wavelet according to

$$\psi(x) = j\mathcal{C}\phi(x) \tag{3.142}$$

with σ set to unity. Notice that in the Mellin domain this mapping becomes a multiplication by c an observation well reconcilable with the notion that the wavelet transform is the result of the action of the scale derivative on the coarse grained function $f(\sigma,x)$.

Remark that Cohen (1993) also mentions that such a relationship exists. To this I would like to add that the continuous wavelet transform can be seen as a windowed scale transform allowing for the local assessment of the scale parameter annex scale exponent. One may also observe that these estimates can not be localized at will because scale and space do not commute. Concerning the global multiscale analysis the wavelet transform offered a definition of a partition function from which the singularity spectrum can readily be obtained which on its turn can be associated with the spectral scale representation. But more importantly I believe that that the wavelets provide a proper functional space on which the local/global scale analysis can safely be given a meaning especially when singular constructs such as multifractals are subject of investigation.

3.7 Concluding remarks

In this chapter I made an attempt to concisely lay down the apparatus required by the basis operations that make the multiscale analysis work. The operators responsible

for the shift and dilatation operation were derived and were given a meaning via their spectral representation. This spectral representation is found by solving the eigenvalue problem belonging to the operators. Moreover these eigenfunctions define the basis for the unitary spectral Fourier and Mellin/scale transform.

Given the spectral representation one is set to analyze and manipulate the information content of functions/signals submitted to a transform to the spectral domain. In the cases examined in this chapter the spectral domain referred to the spatial frequency and scale pertaining to the frequency and scale operators. The subsequent spectral analysis of the two physical quantities is, however, restricted because the space and scale operators do ***not*** commute withstanding a simultaneous resolution of these physical quantities up to arbitrary accuracy. The contents of this chapter will also facilitate the discussion on the spectral problem pertaining to the evolution operator ruling the acoustic wave motion.

The final goal of this chapter was, at least at the conceptual level, to reveal the possible link between scaling – as evidenced from multifractals and the local/global multiscale analysis on real well-data – and the scale representation. It is this association that will also be of assistance in the discussion on the role of the wave equation in relation to the notion of scale, the subject matter of the concluding of the epilogue.

Chapter 4

The current model for acoustic wave motion

In this chapter the very basic framework describing acoustic wave motion is presented. This framework is generally being used to describe the wave interactions in heterogeneous media. The acoustic model consists of a hyperbolic coupled system of first order partial differential equations in which the medium parameters occur as coefficients. The aim of the model is to quantitatively describe acoustic wave motion and this is done by solving the system of equations given certain profiles for the medium coefficients. One of the ways to do that is to solve the pertaining time evolution problem which comes down to solving the eigenvalue problem for a Hamiltonian. The nature of the spectrum of this Hamiltonian plays an important role and it appears that the spectrum becomes dense pure point in case the medium varies randomly along one direction with a rapid enough decaying correlation function. In that case the Lyapunov exponents become larger than zero and one can obtain global information on the dispersion via the Lyapunov spectrum. On its turn this information may be linked to the medium's complexity.
key words: *self-adjoint operators, spectral problem, localization theory, Lyapunov exponents, inversion.*

4.1 Introduction

As a wave propagates in a medium the heterogeneity of which is given, for example, by a well-log measurement, an extremely intricate wave interference mechanism comes into effect. This interaction mechanism yields a dispersion – evidenced by the wave as it propagates deeper and deeper into the heterogeneity – and specular reflections. Because of the irregularity, reflections occur at almost every depth level and these reflections may eventually lead to a complete back scattering of the probing wave. The aim of this chapter is to give a short review on what the current physical comprehension, based on the acoustic wave equation, has to offer when faced with the problem to say something on the time evolution of solutions to the one-dimensional acoustic wave problem where well-log measurements are used to define the medium coefficients. Given a good understanding of

this problem it is hoped that one is able on the one hand to come up with predictions for the time evolution of the wavefield given a well-log measurement, and on the other that one is able to come up with information on the medium complexity given the seismic data.

The setup of this chapter is as follows. I will start with a section in which I discuss a qualitative analysis conducted on a Vertical Seismic Profile (VSP). Then I will briefly introduce the ruling equations and cast them into a time evolution problem for the Cauchy data. In this time evolution problem a Hamiltonian emerges in which the medium properties occur as coefficients. Finding the solution to the eigenvalue problem pertaining to the Hamiltonian is the next topic. It is striking that solving this eigenvalue problem remains very difficult and up to today only relatively little is known on the existence and properties of the solutions. It appears, given recent results by localization theory (Souillard, 1986; Pastur and Figotin, 1991; Pastur, 1994; Carmona and Lacroix, 1990; Faris, 1995), that the spectrum pertaining to the Hamiltonian is dense pure point in cases of a one-dimensional medium that displays a moderate type of randomness in the scattering potential. This potential is, under specific conditions properly related to the acoustic medium parameters. At that point the Lyapunov exponents come in as estimates for the frequency dependent localization length, describing the attenuation, and as indicators for the emergence of the pure point spectrum. The theory proving this is very involved and it is beyond the scope of this thesis to go into detail. Besides the aforementioned information it may be postulated that the Lyapunov spectrum carries more global information, hopefully to be linked to the nature of the medium's complexity. It is that conjecture that coined me to propose an inversion type of scheme at the end of this chapter. This inversion scheme aims to capture a scaling exponent bearing the potential to be linked to the scaling of the medium. The interesting point about the powerlaw *ansatz* for the Lyapunov spectrum, is that it is directly related to the work of le Méhauté (1995) who proposes to replace the time derivative in the Maxwell equations by a fractional time derivative. The degree of this fractional derivative is related to the fractal dimension. And in this way he introduces a powerlaw behaviour for the Lyapunov spectrum.

4.2 A first look at some data

The best way to improve one's understanding on what happens when an acoustic wave impinges the earth's subsurface is to have a look at seismic data. For the problem at hand a Vertical Seismic Profile (VSP) serves this purpose well. Such a VSP data set refers to a measurement of the particle velocity induced by a source at the surface and detected along the borehole in the vertical direction. Consequently a VSP constitutes a measurement that is a function of the vertical offset, and the time coordinate, depicted horizontally, see figure 4.1.

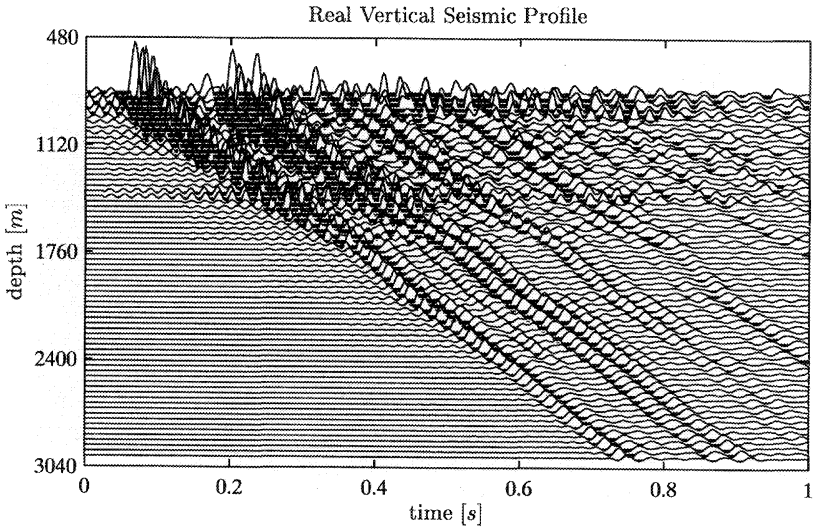

Figure 4.1 *This figure contains a real VSP. In this figure the particle velocity is depicted as a function of the time, horizontally, and depth, vertically. Remark that as the wave travels down its "impulsive" nature gradually diminishes.*

What comes to mind first when examining this real VSP data set? First of all one evidences a wave travelling downward. On its way down this wave is reflected resulting in the "V" shaped events. What also becomes apparent is that the wave gradually disperses as its propagation distance increases. This notion becomes even more manifest in a simulated VSP measurement, see figure 4.2, obtained via a numerical modelling code that uses the well-log as input and that, in this case, comes up with a prediction for the normal incident plane wave constituent. The modelling is based on a so-called layercode scheme, constituting a transfer matrix type of approach in the sense of Redheffer (1961) and Kennet and Kerry (1978).

The observation that the "impulsive" character of the probing wavefield gradually disperses as the wave travels downward becomes also apparent in figure 4.3. This figure clearly shows the apparent[1] dispersion experienced by the transmitted wavefield and induced by the heterogeneities. In this figure I included a time-frequency analysis conducted on the bottom and top most traces taken from the real VSP and the simulated normal incident plane wave impulse response[2], computed using the layercode. The time-frequency analysis consists of the computation of the joint time-frequency distribution,

[1]The term apparent dispersion is used to discern between this type of dispersion and the one induced by relaxation effects. The first is due to a conjectured redistribution of the energy due to multiple reverberations while the latter corresponds to an actual viscous loss of acoustic energy into heat.

[2]Not really the impulse response but a coarse-grained approximation of it.

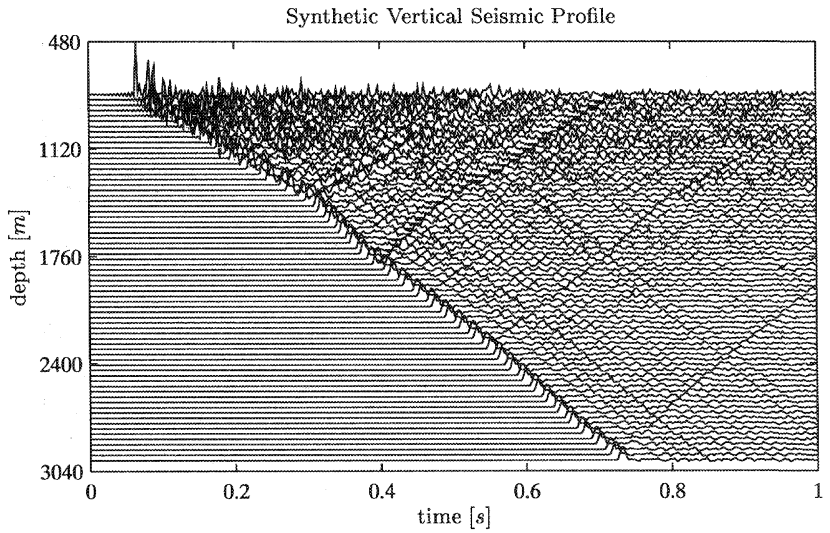

Figure 4.2 *This figure contains a synthetic VSP. In this figure the acoustic pressure is depicted as a function of the time, horizontally, and depth, vertically. Remark that as the wave travels down its "impulsive" nature gradually diminishes. Remark that a different type of source signature used as the one in the real VSP, while the vertical offset does not exactly match to the one in the real VSP, see figure 4.1*

see the middle gray-scale plot with horizontally the time and vertically the frequency, and the amplitude spectrum. The envelope of the time trace can be viewed upon as the marginals of the time-frequency distribution, see Cohen (1995). Inspection of this figure clearly shows that the frequency is confined to a specific band for the real data while the simulated data show a very irregular behaviour over the whole frequency range. This is because of the different source signatures. From both pictures it is clear that the higher frequencies are more severely attenuated, a notion corresponding to dispersion.

At this point the question is, can one account for the evidenced dispersion effects believed[3] to be caused by the heterogeneity using the current wave theory? Quite obviously this will not be a straightforward exercise because there is no clear mechanism, in a relaxation free formulation for the acoustic wave motion, that explains the apparent dispersion. So let me first briefly introduce the current formulation for the acoustic wave motion that forms the basis for the theory that tries to explain the aforementioned quantitative effects.

[3] I am tempted to more or less rule out the intrinsic dispersion induced by a viscous relaxation mechanism for the temporal frequency range inhabited by seismic waves. I realize that by assuming a relaxation one may explain for the observations. However, there is no direct evidence supporting this explanation.

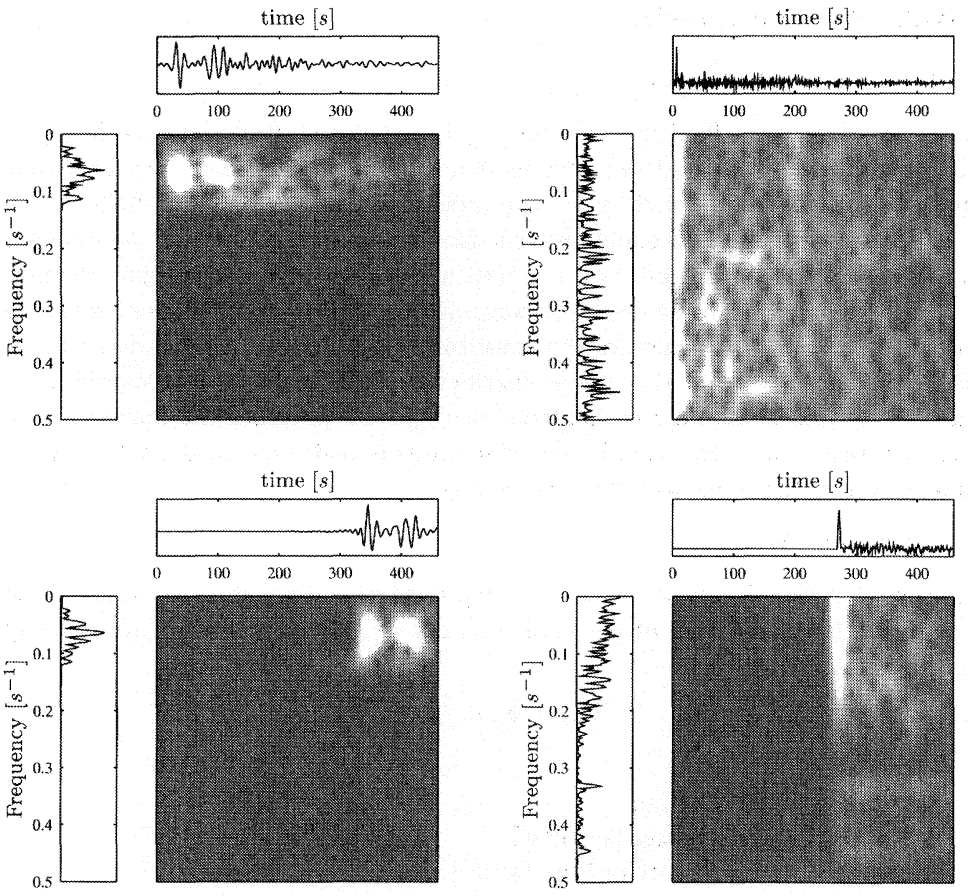

Figure 4.3 *In this figure I illustrate the time-frequency behaviour displayed by the bottom and top most traces taken from the VSP's depicted in figures 4.1 (the real VSP) and 4.2 (the simulated VSP). On the top the time-frequency analysis on the still "impulsive" source signatures are depicted for the true (left) and simulated (right) VSP's. At the bottom one sees the dispersed traces. The time-frequency analysis itself consists of the joint time-frequency distribution in the middle and its marginals, for the frequency, on the left and for the time at the bottom. Notice that the frequency is confined to a specific band for the real data while the simulated data shows a very irregular behaviour over the whole frequency range. From both pictures it is clear that the higher frequencies are more severely attenuated a notion corresponding to dispersion. Moreover the modelled data shows a very irregular behaviour for the higher frequencies judged by the spurious peaks in the amplitude spectrum.*

4.3 Basics of acoustodynamics

4.3.1 The equations of acoustics

The acoustodynamics of a fluid are determined by the combined effect of two mechanisms: first it constitutes a reaction of the wave motion – the dynamical state of matter that is superimposed on a static equilibrium state (Fokkema and van den Berg, 1993) – to a *force* given the *inertia* and, secondly, the reaction of the wave motion to a *deformation* given the *compliance*. In this discourse it is assumed that it suffices to consider the first order terms only to describe the acoustic wave motion. Since seismic waves propagate at relative low temporal frequencies the constitutive parameters are considered to be *relaxation* free and moreover the spatial distribution of the medium heterogeneities is assumed to be "frozen" during the characteristic time scale on which the transient wave phenomenon takes place. For a much more elaborate discussion on the derivation of the acoustic wave equation I like to refer to the epilogue.

The equation of motion

Application of *Newton's law of motion* to a representative elementary domain[4] of the fluid leads to the formulation of the *local equation of motion* (de Hoop, 1995) being given by

$$\partial_k p + \dot{\Phi}_k = f_k \tag{4.1}$$

with

p = acoustic pressure [Pa],
$\dot{\Phi}_k$ = mass-flow density rate [kg/m^{2}s^2],
f_k = volume source density of force [N/m^3]

and where the mass-flow density rate is a function of space-time

$$\dot{\Phi}_k(\mathbf{x}, t) = \rho(\mathbf{x}) D_t v_k(\mathbf{x}, t). \tag{4.2}$$

Here I use the summation convention for repeated subscripts. In accordance to the seismic conventions the position is defined by the boldfaced vector, $\mathbf{x}$, the coordinates $\{x_1, x_2, x_3\}$ of which are taken with respect to a Cartesian reference frame with its origin $\mathcal{O}$ located at the surface and where the vector $\{0, 0, x_3\}$ points downward. In cases where the medium varies in one direction the medium is assumed to extend to infinity in the other directions, i.e. the medium is shift invariant in the x_1 and x_2-directions. In equation (4.2) one also finds

ρ = volume density of mass [kg/m^3],
v_k = particle velocity [m/s],

and the co-moving time derivative,

$$D_t = \partial_t + v_l \partial_l, \tag{4.3}$$

[4]The notion of elementary volumes will be extensively addressed in the epilogue.

denoting the derivative experienced by an observer travelling in a co-moving frame of reference.

The deformation rate equation

The *local* temporal changes in the volume and/or shape of a portion of the fluid are characterized by the *deformation equation* reading

$$\partial_k v_k - \dot{\Theta}^i = q \tag{4.4}$$

with

$\dot{\Theta}^i$ = induced part of the cubic dilatation rate [s^{-1}],
q = volume source density of injection rate [s^{-1}].

The cubic dilatation rate is given by

$$\dot{\Theta}^i(\mathbf{x},t) = -\kappa(\mathbf{x}) D_t p(\mathbf{x},t) \tag{4.5}$$

with

κ = compressibility [Pa^{-1}].

As I stated before the fluid is assumed to be void of *relaxation effects*. It is also supposed to be *passive*, *linear*[5], *time invariant* and *locally reacting*. This latter presupposition refers to the fact that the values for the pair $\{\dot{\Phi}_k, \dot{\Theta}^i\}$ at position $\mathbf{x}$ depend *strict locally* on the values of the field $\{v_k, p\}$ at that same location $\mathbf{x}$, as is expressed by equations (4.2) and (4.5).

Finally I like to remark that when the fluid displays a *shift-invariance* within a bounded domain it is designated to be *homogeneous* while it is denoted as *inhomogeneous* in case the shift-invariance does not hold in that domain.

4.3.2 Linearization by a low-velocity approximation

Close inspection of the equation of motion, the deformation rate equation and the constitutive relations shows that the particle velocity, v_r, occurs non-linearly because it appears in the Lagrangian time derivative D_t. For seismic applications, however, one can on physical grounds[6] neglect the particle velocity's contribution in the time derivative. So the accuracy of the equations is maintained despite the replacement

$$D_t \quad \mapsto \quad \partial_t. \tag{4.6}$$

[5]To be understood in the Lagrangian sense since the constitutive relations refer to a specific region of the fluid, the representative volume (Fokkema and van den Berg, 1993; de Hoop, 1995).

[6]The variations in the quantities that are induced by the wave motion have small-amplitudes.

Now the basic equations of acoustics become, respectively, the familiar *second law of Newton* and *Hooke's law*, i.e.

$$\begin{aligned} \partial_k p &+ \rho \partial_t v_k &= f_k, \\ \partial_k v_k &+ \kappa \partial_t p &= q. \end{aligned} \tag{4.7}$$

This coupled system of first order partial differential equations, pde's, will serve as point of departure for defining the proper representation for the solution of the acoustic wave motion.

4.4 Operator formalism for acoustic wave motion

Wave theory constitutes a large branch of physics. For that reason there exists a large suite of different approaches to come up with solutions for the wave equation. However, if one takes a closer look there does not exist a large body of literature addressing the issue on the existence and asymptotic completeness of solutions, an important aspect given the nature of the problem central in this thesis. Therefore it is in my opinion beneficial to make reference, when possible, to some of the findings yielded by the rather abstract approaches developed within the fields of *scattering theory* and *spectral theory* belonging to the Schrödinger and or wave equation (Reed and Simon, 1979; Pearson, 1988; Carmona and Lacroix, 1990). The books by Reed and Simon (1979) form standard works in mathematical physics. In fact they come up with a number of theorems proving the existence and asymptotic completeness of the wave operators – under strict conditions to be imposed on the variability of the coefficients – that describe the time evolution of the acoustical states that occur in heterogeneous media. These states refer to the Cauchy data, the pressure and its time derivative.

The discussion starts by briefly setting up the self-adjoint system that can be associated with the wave equation defined by the coupled system of equations (4.7). Then I will briefly mention the relevance of proving the *existence* and *asymptotic completeness* of the wave operators. These proofs have to do with the fact that scattering theory is considered as a special branch of perturbation theory where one tries to interlink the "free" states, associated with a wave equation for the homogeneous medium, with the asymptotic given/actual states occuring in the heterogeneous environment as time goes to infinity. Read for these states vectors containing the Cauchy data. Hence the wave operators can be seen as operators that carry these vector functions, living in the proper Hilbert space, from one time instant to the other. Although I am aware of the fact that scattering theory entails an asymptotic theory, regarding the long time behaviour, it is a theory relatively well equipped with a mathematical apparatus that increases the insight. However, there remains a discussion on its applicability within the seismic context but it will certainly be of use due to its strong ties with spectral theory (Reed and Simon, 1978, 1979) – in the sense that the asymptotic completeness proofs exclude the existence of the singular continuous spectrum – and localization theory.

4.4.1 The self-adjoint system

Given the system defined in equation (4.7) one is set to prove the existence and asymptotic completeness for the solution of the wave operators. These proofs rely on the existence of a proper functional space, the space of square integrable functions, hosting the wave operators that act on the proper initial conditions, Cauchy data, and that belong to the *given* system, the *actual* physical system, as well as to the "free" state that corresponds to the *unperturbed* situation where the wave motion is free. Following Reed and Simon (1979) there are two different approaches to recast the coupled system into self-adjoint "free" and given systems. The first approach, see Reed and Simon (1979) p.g. 366 by Schulenberg-Wilcox, deals with an initial value problem for a self-adjoint first order system of partial differential equations. Reed and Simon (1979),Wilcox (1984) and Carmona and Lacroix (1990), on the other hand, reformulate the above system of equations (4.7) into an initial value problem[7] of the form

$$\begin{cases} \partial_t \mathbf{u}(\mathbf{x},t) = -j\boldsymbol{\mathcal{A}}\mathbf{u}(\mathbf{x},t) \\ \mathbf{u}(\mathbf{x},0) = \mathbf{u}_0 \triangleq \{f,g\}^T, \end{cases} \tag{4.8}$$

where the vector $\mathbf{u}(\mathbf{x},t)$ contains the proper Cauchy data. The system matrix operator is given by

$$\boldsymbol{\mathcal{A}} = j\begin{pmatrix} 0 & \mathcal{I} \\ -\mathcal{H}_2 & 0 \end{pmatrix} \tag{4.9}$$

in which the second order operator $\mathcal{H}_2$ emerges being given by

$$\mathcal{H}_2 = -c^2(\mathbf{x})\rho(\mathbf{x})\nabla\cdot\left(\frac{1}{\rho(\mathbf{x})}\nabla\cdot\right) \tag{4.10}$$

and where the acoustic wavefield vector is now defined by the Cauchy data

$$\mathbf{u}(\mathbf{x},t) = \begin{pmatrix} p \\ \partial_t p \end{pmatrix}(\mathbf{x},t). \tag{4.11}$$

The medium is parameterized by the density of mass, ρ, and the local compressional wave speed, c, which is related to the density and compressibility according to $c = (\rho\kappa)^{-1/2}$. These parameters are allowed to vary with position in all coordinate directions. Notice that the operator in equation (4.10) only becomes valid in case $\rho \in C^1$.

The system in equation (4.8) represents a wave equation of the type

$$\partial_t^2 p(\mathbf{x},t) - c^2(\mathbf{x})\rho(\mathbf{x})\nabla\cdot\left(\frac{1}{\rho(\mathbf{x})}\nabla p(\mathbf{x},t)\right) = F(\mathbf{x},t) \tag{4.12}$$

[7] Note that this initial value problem can be supplemented with a source distribution, see for example Dautray and Lions (1992).

with $p(\mathbf{x},t)$ being considered as the unknown pressure wavefield while $F(\mathbf{x},t)$ is the known source distribution. This equation is known as the "ordinary" *acoustic wave equation* describing the transient space-time characteristics of the pressure fluctuations around the equilibrium pressure and induced by a causal source. As before, one of the ways to infer information on this system is to recast equation (4.12) into an initial value problem of the form

$$\begin{cases} \partial_t^2 p(\mathbf{x},t) - c^2(\mathbf{x})\rho(\mathbf{x})\nabla\cdot\left(\dfrac{1}{\rho(\mathbf{x})}\nabla p(\mathbf{x},t)\right) = 0 \\ p(\mathbf{x},0) = f(\mathbf{x}) \quad \partial_t p(\mathbf{x},0) = g(\mathbf{x}). \end{cases} \tag{4.13}$$

Following Reed and Simon (1979), Wilcox (1984) and Carmona and Lacroix (1990) I supplement this initial value problem with the condition that the compressional velocity and density of mass converge rapid[8] enough to the values yielded by the embedding space, i.e.

$$c(\mathbf{x}) \to c_0 \qquad \rho(\mathbf{x}) \to \rho_0, \tag{4.14}$$

as $|\mathbf{x}| \to \infty$. Additionally one has to impose conditions on the boundedness of these two parameters. This implies that the conditions

$$0 < \rho_1 \leq \rho(\mathbf{x}) \leq \rho_2 < \infty \tag{4.15}$$

$$0 < c_1 \leq c(\mathbf{x}) \leq c_2 < \infty \tag{4.16}$$

have to hold for all $\mathbf{x} \in \mathbb{R}^3$ and for the constants ρ_1, ρ_2, c_1 and c_2, all bounded. Only when all these conditions are met the operator $\mathcal{H}_2$ is proper on C_0^∞ (Reed and Simon, 1979).

By way of its construction the matrix operator $\boldsymbol{\mathcal{A}}$, the infinitesimal *generator*, is self-adjoint on its domain[9], a prerequisite for finding the corresponding spectral representation, while its solution is formally given by

$$\mathbf{u}(t) = \boldsymbol{\mathcal{W}}\mathbf{u}_0 \tag{4.17}$$

with

$$\boldsymbol{\mathcal{W}} = e^{-j\boldsymbol{\mathcal{A}}t} = \begin{pmatrix} \cos\mathcal{H}_1 t & (\mathcal{H}_1)^{-1}\sin\mathcal{H}_1 t \\ -(\mathcal{H}_1)^{-1}\sin\mathcal{H}_1 t & \cos\mathcal{H}_1 t \end{pmatrix} \tag{4.18}$$

being the *wave operator*, also known as the *propagator*, governing the time evolution while

$$\mathcal{H}_1 = \sqrt{\mathcal{H}_2} \tag{4.19}$$

[8] For precise conditions the reader is referred to the references frequently occuring in the text.

[9] Remarks on the domains and other mathematical technicalities pertaining to the operators can be found in (Reed and Simon, 1979).

refers to the square root of $\mathcal{H}_2$.

One of the ways to give the square root operator a meaning is to solve the eigenvalue problem belonging to the $\mathcal{H}_2$ operator. In this way one can relative straightforwardly compute the square root in the spectral domain. Before paying attention to the spectral problem let me first go into a little more detail on setting up a scattering formalism in order to touch a little on the requirements for proving the existence and asymptotic completeness of the wave operators.

4.4.2 Existence and asymptotic completeness of the wave operators

Up to this point the discussion on the construction of a scattering mechanism for acoustic wave motion has not yet been supplied with the mathematical rigour it requires to issue statements on the existence and asymptotic completeness of the wave operators. Such an aim is beyond the scope of this thesis but I will make an effort to convey, in words, some of the ideas behind setting up the proper conditions to prove the existence and asymptotic completeness pertaining to the wave operators associated with the "free" and given dynamics for the acoustic wave motion. I do not intend to prove it and I like to refer the curious reader to the work by Reed and Simon (1979), Pearson (1988), Carmona and Lacroix (1990) and Weder (1991). In this discussion I will limit myself to which conditions one has to impose on the density of mass and local compressional wave speed distributions, in order to prove the existence and asymptotic completeness for the wave operators pertaining to the Hamiltonian defined in equation (4.10). Given this proof one can say that the wave operators converge, in the strong sense, but only in the energy norm, for those cases where the medium distributions live up to the required conditions.

As I mentioned scattering theory constitutes a special branch of perturbation theory and therefore it is not surprising that it intrinsically deals with the dynamics of two types of systems for the same system: *the given dynamics and "free" dynamics* (Reed and Simon, 1979). The basic idea behind this is that the free dynamics generally refers to those cases which are simpler than the given dynamics but sharing the same physical principles. To summarize, the main arguments favouring the use of some aspects of scattering theory lie in the fact that

- it provides the necessary theorems proving the existence and asymptotic completeness of wave operators by imposing certain condition on the coefficients occuring in the wave equation.
- it is closely related to the pertaining spectral problem (Reed and Simon, 1979) in the sense that a proof of the existence and asymptotic completeness corresponds to ruling out the existence of a *singular continuous* part of the spectrum.

The main point while proving the existence and asymptotic completeness of wave opera-

tors in a Hilbert space is that one can demonstrate that there exist given wave operators that can be proven to be asymptotically free at $t \to -\infty$ (Reed and Simon, 1979) and that converge, as time progresses to infinity, again to the "free" dynamics. In that case one only has states that are *not* bound, i.e not localized, see section 4.6.

Reed and Simon (1979) are able to prove the existence and asymptotic completeness, see theorems XI.75-XI.77, for the $\mathcal{H}_2$ by supposing that $c(\mathbf{x})$ and $\rho(\mathbf{x})$ are twice continuously differentiable functions with bounded derivatives satisfying (4.15), (4.16) and (4.14). Moreover they suppose that $c^2(\mathbf{x}) - c_0^2$, $\rho(\mathbf{x}) - \rho_0$, $D^\alpha c(\mathbf{x})$, $D^\alpha \rho(\mathbf{x})$, $0 \neq |\alpha| \leq 2$, are all[10] in $L^2_\delta(\mathbb{R}^3)$[11]. Then the wave operators associated with the given and free systems exist and are complete. Given these theorems Reed and Simon (1979) prove the completeness in the sense of generalized wave operators. Furthermore they say that it can be proven that the solution of equations (4.8) and (4.13) is asymptotic to a free solution as time progresses. In fact they have proven that $\mathcal{A}$ has *no singular continuous spectrum*, see also section 6 on pg. 112 (Reed and Simon, 1979).

Then they proceed by saying that the decay condition on $c^2(\mathbf{x}) - c_0^2$ and $\rho(\mathbf{x}) - \rho_0$, $D^\alpha \rho(\mathbf{x})$ and $D^\alpha c(\mathbf{x})$ are not very stringent in the sense that they will hold for any reasonable physical situation. On the other hand they say that the regularity conditions on the medium parameters may jeopardize the applicability of the scattering formalism and they claim that these requirements can be removed. Unfortunately this is where their elaborate treatise comes to a hold and it is beyond the scope of this thesis to come up with the proofs. Despite this deviance the results of this section will be useful since the singular continuous spectrum has been excluded. At this point it is interesting to feedback the asymptotic completeness to the empirical observations. These observations indicated that the waves in the earth disperse, a notion also found from the numerical modelling example. That means that one may have to do with a situation where there is no longer asymptotic completeness because certain physical states may be trapped in the complexity of the medium heterogeneity. In fact this is a possible indication that the wave phenomenon can not longer be treated as a perturbation with respect to the free system. It is possibly perturbed too much as a consequence of which it looses its hyperbolic behaviour. It becomes a standing kind of wave (Faris, 1995). In the epilogue I will pay some attention how one can use the findings of the multiscale analysis in relation to the above conditions for the medium.

4.5 The eigenvalue problem and the spectral representation

One way to come up with solutions, wave operators, to the initial value problem – defined by the system in equation (4.8) and belonging to the wave equation (4.13) with the Hamiltonian, $\mathcal{H}_2$, given by equation (4.10) – is to represent the operators in the

[10] Note that the D^α is defined as $D^\alpha f \triangleq \partial_1^{\alpha_1} \partial_2^{\alpha_2} \partial_3^{\alpha_3} f$ with $\alpha = \{\alpha_1, \alpha_2, \alpha_3\}$ and $|\alpha| = \alpha_1 + \alpha_2 + \alpha_3$.
[11] $L^2_\delta(\mathbb{R}^3)$ is the set of f such that $\|f\|_\delta = \|(1+x^2)^{\delta/2} f(x)\|_{L^2} < \infty$.

spectral domain pertaining to the operator $\mathcal{H}_2$. As I showed in chapter 3 it is relatively straightforward to give analytical functions – such as the square root and exponentiation – a meaning in the *spectral domain* because these functions on *operators* become simple functions on *scalars*, the eigenvalues. In chapter 3 I reviewed a relative simple class of operators all having an absolutely continuous spectrum for the eigenvalues. Moreover the eigenvalues are real and the eigenfunctions[12] are complete and orthogonal. Not to blur the basic principle I refrained in that chapter from issuing specific conditions on functional spaces to be attributed to the eigenfunction and the functions undergoing the action of the operator. In case of the operator $\mathcal{H}_2$ I can not longer do that because a lot of the physics depends on the nature of the spectrum.

Indeed the situation for the operator $\mathcal{H}_2$ is much more involved. The elaborate proofs in the literature on the existence and asymptotic completeness, associated with the wave operators, constituted already a precursor on the difficulties that emerge when solving the spectral problem for the $\mathcal{H}_2$. The spectral representation allows for the computation of the square root, $\mathcal{H}_1 = \sqrt{\mathcal{H}_2}$, and the subsequent exponentiation, yielding the expression for the propagator operator $\boldsymbol{\mathcal{W}}$. Being aware of the limited scope of this treatise I like to mention that in my opinion many of the efforts focus on proving the absence of the singular continuous spectrum (Reed and Simon, 1978, 1979; Pastur and Figotin, 1991; Pastur, 1994). This singular continuous part of the spectrum refers to that part of the spectrum that is *singular* with respect to the Lebesgue measure. This terminology comes from theorems in abstract measure theory, see Reed and Simon (1980) pg. 22, that tells one that *any measure, μ, on $\mathbb{R}$ has a canonical decomposition $\mu = \mu_{pp} + \mu_{ac} + \mu_{sing}$ where μ_{pp} is pure point, μ_{ac} absolutely continuous and μ_{sing} is continuous and singular relative to the Lebesgue measure.* Before I will pay attention to the physical interpretation of this decomposition and the nature of the associated (generalized) eigenfunctions let me present the spectral representation of $\mathcal{H}_2$ first.

The spectral representation (Wilcox, 1984; Souillard, 1986; Pearson, 1988; Carmona and Lacroix, 1990; Dautray and Lions, 1990) of the $\mathcal{H}_2$ operator is given by

$$\mathcal{H}_2 = \int_{-\infty}^{+\infty} \lambda \mathrm{d}E_\lambda, \tag{4.20}$$

where $\{E_\lambda \mid -\infty < \lambda < +\infty\}$ denotes the spectral projection over the pure point part of the measure via a summation while the continuous part of the spectral measure is covered by the integral. Given this spectral representation it is rather straightforward to compute the square root operation acted upon the $\mathcal{H}_2$ and yielding the $\mathcal{H}_1$. This computation runs, in the spectral domain, along the lines of the following representation

$$\mathcal{H}_1 = \sqrt{\mathcal{H}_2} = \int_{-\infty}^{+\infty} \sqrt{\lambda} \mathrm{d}E_\lambda. \tag{4.21}$$

[12] Care should be taken because the eigenfunctions do not always represent physical states in the sense that they do not lie in a proper Hilbert space and are therefore referred to as generalized eigenfunctions.

This spectral representation, also known as a *modal decomposition*, is not only applicable to represent operators but can also be used to decompose arbitrary functions.

Before elaborating on some specific situations and examples let me define the spectrum and subsequently describe the physical nature of the states that belong to the different parts of the spectrum.

4.5.1 The spectrum, resolvent and (physical) states

Mathematically speaking the spectrum of the operator, $\mathcal{H}_2$, is the set, $\sigma(\mathcal{H}_2)$ of real λ's such that the operator $(\mathcal{H}_2 - \lambda)$ is not *invertible* as a bounded operator. It defines, as such, the set of eigenvalues, say frequencies, allowed to the physical system, say the wave motion, that is governed by the operator $\mathcal{H}_2$. In fact this definition exactly corresponds to the definition for the *resolvent* of the operator[13] $\mathcal{H}_2$. This resolvent operator relates, when using the proper initial conditions, the operator $\mathcal{H}_2$ and its eigenvalues to the superposition integral operator whose action is defined by the Green kernel (Wilcox, 1984). Notice that by way of its construction the spectrum of $\mathcal{H}_2$, $\sigma(\mathcal{H}_2)$, is generally a larger set than the set yielded by the eigenvalue problem solution of

$$\mathcal{H}_2\psi(\lambda, \mathbf{x}) = \lambda\psi(\lambda, \mathbf{x}). \tag{4.22}$$

The eigenfunctions lie in the Sobolev space $W^{2,2}$. This latter category of eigenfunctions are denoted as generalized eigenfunctions.

The reason why these eigenfunctions have to adhere to these conditions lies in the fact that physically speaking one is only interested in those states/solutions, ψ, that represent physical plausible solutions to the eigenvalue problem. Finally notice that the operator may yield solutions that are exponentially decaying but that implies that exponentially growing solutions are also solutions which are certainly *not* physical. Finally notice that from the physical point of view the spectrum must, by convention (Souillard, 1986), be considered as a *coarse-grained* quantity[14] because it is hard to disjoin those eigenvalues that are arbitrarily close located with respect to each other.

4.5.2 Nature of the spectrum and the associated physical states

So far I only talked about the physical states without being too specific on the physical significance of the eigenvalues and their corresponding physical states. Conform the afore mentioned canonical decomposition I propose a subdivision of the spectral measure and the corresponding eigenvectors, spanning the Hilbert space or a subset of it, into three parts

$$\sigma(\mathcal{H}_2) \triangleq \sigma_{pp}(\mathcal{H}_2) \cup \sigma_{sc}(\mathcal{H}_2) \cup \sigma_{ac}(\mathcal{H}_2), \tag{4.23}$$

[13]The resolvent defines the integral operator $(\mathcal{H}_2 - \lambda)^{-1} f(\mathbf{x}) = \int \mathrm{G}(\lambda; \mathbf{x}, \mathbf{x}') f(\mathbf{x}') c^{-2}(\mathbf{x}') \rho^{-1}(\mathbf{x}') \mathrm{d}\mathbf{x}'$, where G is a Green kernel (Wilcox, 1984).

[14]The resolution of identity E_λ will never be reached, i.e. one will never resolve a "delta" functional.

where the subspaces are orthogonal and span the whole space. The subscripts of the different parts refer to *pure point* for *pp*; *singular continuous* for *sc* and *absolutely continuous* for *ac*. The combinations $\sigma_s(\mathcal{H}_2) \triangleq \sigma_{pp}(\mathcal{H}_2) \cup \sigma_{sc}(\mathcal{H}_2)$ are referred to as *singular* whereas $\sigma_c(\mathcal{H}_2) \triangleq \sigma_{sc}(\mathcal{H}_2) \cup \sigma_{ac}(\mathcal{H}_2)$ is referred to as *continuous*. Notice that these spaces may be zero but not all!

The presented subdivision is not void of any physical meaning because it reflects the presence of different physical states. On their turn these states are subdivided into

- *localized states*, states that are *exponentially* decaying. These localized states match the *pure point spectrum* and the corresponding physical states, ψ, are all of finite energy, i.e. square integrable.
- *exotic states*, states that are to be associated with the *singular continuous* spectrum. These states display a complicated hierarchy of structures that adhere to some kind of scale-invariance (Souillard, 1986). Finally notice that these states are not square integrable.
- *extended states*, states that are essentially extended over the whole system certainly for the low temporal frequencies, the large spatial wavelengths. These states share the property of being **not** *square integrable* with the singular continuous part of the spectrum.

Inspection of the above subdivision shows that the *singular continuous* part of the spectrum causes difficulties in the sense that this part of the spectrum can be attributed to either the *continuous*, $\sigma_c(\mathcal{H}_2)$, or *singular*, $\sigma_s(\mathcal{H}_2)$, part of the spectrum while it shows a complicated structure. Moreover it will appear that the eigenvalues of the pure point may be dense, a notion that also complicates matters. Probably that is why Souillard (1986) included the statement, see the end of his section 1.5 page 314 *"there is no general theorem proving the above dichotomy and it is possible that all conjectures turn out to be false due to particular counter examples."*

4.5.3 Examples

The two numerical[15] examples included in this section are aimed to qualitatively reveal the difference in appearance of the (generalized) eigenfunctions pertaining to a homogeneous and varying medium. The latter varying medium is defined in terms of the compressional wavespeeds and densities from a selected region obtained from a real well-log measurement, see figure 2.1. Inspection of figure 4.4 in relation to figure 4.5 shows quite a drastic difference, especially in the behaviour for the larger eigenvalues. Indeed the homogeneous medium displays the expected behaviour for the generalized eigenfunctions where the eigenvalues are of multiplicity two. For the heterogeneous situation, however, one observes eigenfunctions being somewhat intermittently distributed in space.

[15] Remark that within the numerical implementation, based on the first order approximation to the derivative operator, the $\mathcal{H}_2$ operator is not longer self-adjoint because of lack of symmetry.

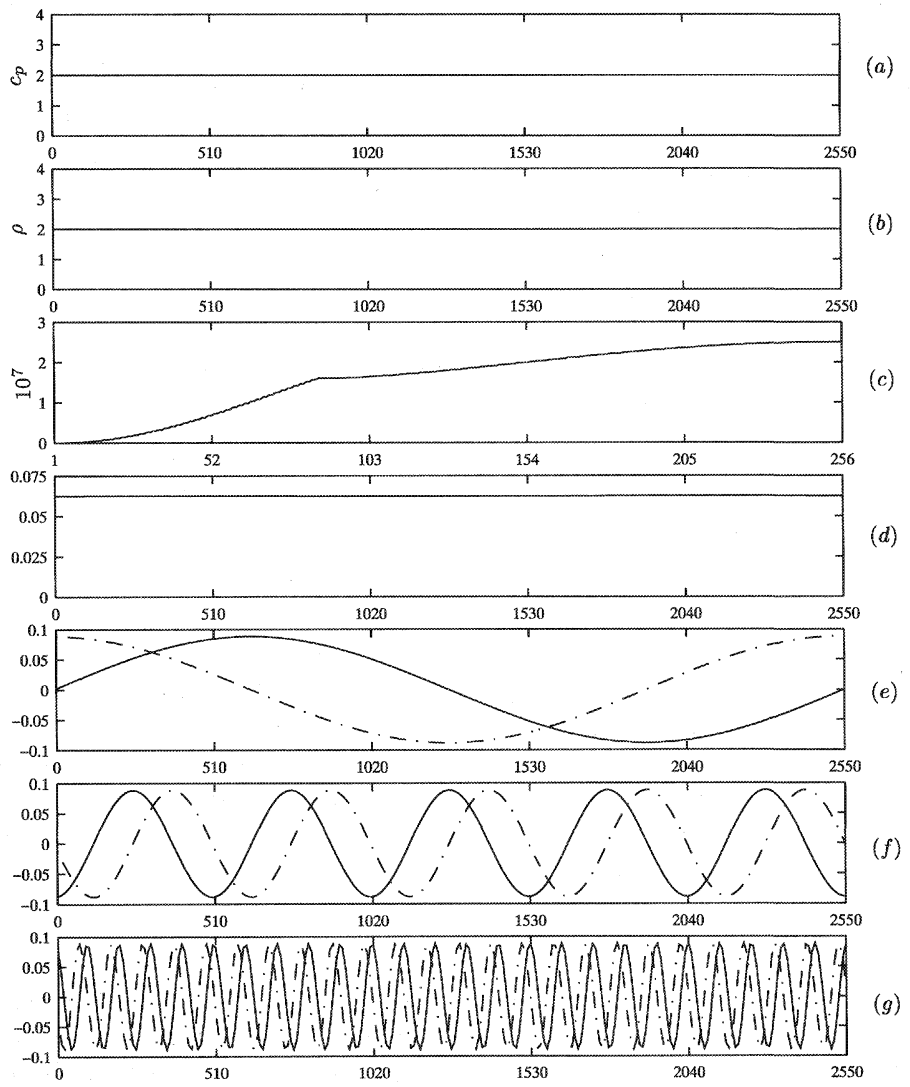

Figure 4.4 *Eigenvalue decomposition of the $\mathcal{H}_2$ for a constant velocity and density medium. (a) velocity profile; (b) density profile; (c) the eigenvalues; (d) eigenfunction corresponding to the first eigenvalue; (e) eigenfunctions corresponding to the second eigenvalue with multiplicity two; (f) eigenfunctions corresponding to the sixth eigenvalue with multiplicity two; (g) eigenfunctions corresponding to the 26th eigenvalue with multiplicity two.*

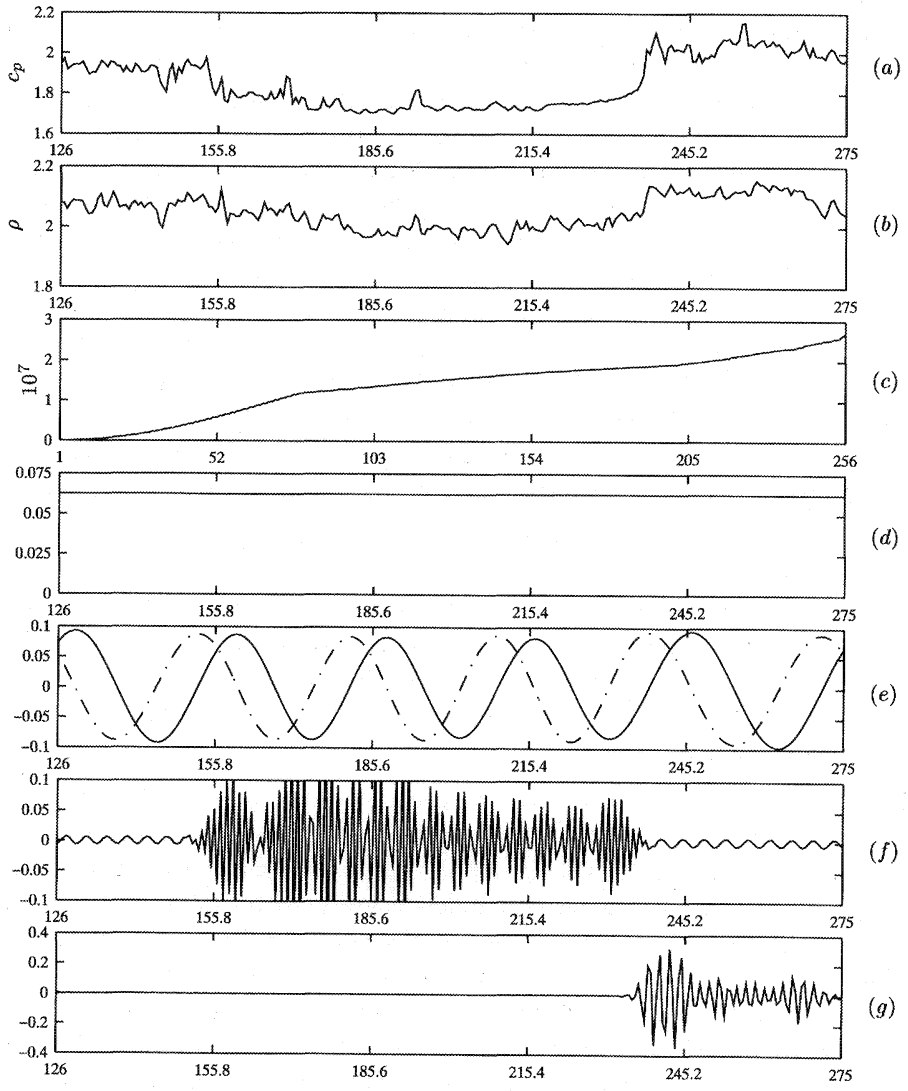

Figure 4.5 *Eigenvalue decomposition of the $\mathcal{H}_2$ for a part of the well-log of figure 2.1 ranging from* 125m *depth to* 275m *depth. (a) velocity profile; (b) density profile; (c) the eigenvalues (d) eigenfunction corresponding to the first eigenvalue; (e)-(g) other eigenfunctions.*

4.6 Localization theory

In this section I will make use of the opportunity to lay down some of the concepts of localization theory, as initially coined by Anderson (1958), that are useful in the discussion surrounding acoustic wave interactions in media that display a complex behaviour. As far as I know localization theory is the most advanced[16] theory dealing with wave interactions in media that contain many irregularities (Faris, 1995). These irregularities give rise to an intricate interference mechanism effecting the waves as they probe the medium. This interference mechanism leads either to a *constructive* mechanism, allowing wave functions to persist over long distances, or to a *destructive* mechanism completely destroying the coherency of the wave function. As will become clear below these properties of the waves depend on the spatial dimension in which the wave phenomena take place. One of the most surprising findings of localization theory is that it predicts certain situations where the wave functions completely loose their wave like behaviour. This manifests itself not only in the fact that waves do not longer propagate at all – the wave carrying medium acts as an isolator rather than a conductor (Souillard, 1986; Faris, 1995) – but also in the fact that the wave function's time evolution depends on the degree of coarse-graining effectuated on the spectrum (Souillard, 1986).

The process being responsible for the interference mechanism is, however, not yet fully understood and this is partly due to the fact that scattering theory does not suffice since: *"localization is not a simple perturbative phenomenon"* (Souillard, 1986). Moreover, for me personally, it also remains a question whether localization theory works for media containing structures on every length scale. In that case namely the "random" heterogeneity is correlated over long spatial distances so one can not exclude the existence of the singular continuous spectrum.

Originally localization theory found its roots in the field of random Schrödinger operators where the potential constitutes a random process[17] (Souillard, 1986; Siebesma, 1989; Carmona and Lacroix, 1990; Pastur, 1994). The conditions on the random process are quite severe in the sense that the correlation length must be finite (Pastur and Figotin, 1991; Pastur, 1994; Faris, 1995). The success of localization theory, as coined by Anderson (1958), sparked the application of this concept, describing the emergence of standing waves being constrained to confined spatial regions, to many different fields of research amongst which the field of seismic exploration. In this latter application the ideas were fed into the discussion on the amplitude behaviour of the seismic wavefield as it interacts with

[16] Advanced in the sense that there is a growing body of theorems substantiating the insights. The theory in itself does not impose too restrained conditions on the medium. This goes, however, at the expense of issuing specific statements on, say, the local nature of the wave interactions.

[17] Here I follow Souillard (1986) who argues: *"if the random medium is ergodic, then the Lyapunov exponent is nonzero (for almost every energy/frequency) as soon as the potential is non-deterministic; non-deterministic means here that one can not determine it completely by knowing it only on some interval, in contrast, e.g. to a quasi- periodic potential."*

the highly heterogeneous earth's subsurface (Baluni, 1985; Sheng et al., 1986; Sheng, 1995; Asch et al., 1990, 1991; Papanicolaou et al., 1990; Burridge and Chang, 1989; Burridge et al., 1992; Lawecki et al., 1994; Lawecki and Papanicolaou, 1994). However, the theory in itself is still under development judged by the fact that still there is not very much known on localization theory in spatial dimensions higher than one and therefore the rigourous results are, predominantly, limited to the one-dimensional situations pertaining to the Schrödinger operator or to the acoustic[18] wave equation (Faris, 1995; Pastur, 1994).

For the one-dimensional situation there exists a very important result that proves that, irrespective of the amount of the disorder, the waves are *exponentially localized* (Souillard, 1986; Pastur and Figotin, 1991; Pastur, 1994; Carmona and Lacroix, 1990; Faris, 1995). In higher spatial dimensions this is not the case and in three dimensional systems one may even, for a not too strong disorder, observe *phase transitions*, when lowering the temporal frequency beyond a critical frequency, the mobility edge, delineating a transgression from localized to extended states.

For the actual application of certain aspects of localization theory I will limit myself in this chapter to considering the one-dimensional situation only. The reason for this choice, despite the fact that this one-dimensionality is certainly not shared by the earth's subsurface, is three-fold:

- detailed information on the medium properties is only available in one dimension, namely along the vertical well-log profiles.
- computationally the one-dimensional case is much easier to comprehend and better tractable.
- theoretically the rigorous results are predominantly constrained to the one-dimensional situation.

I will commence the discussion by first paying some more attention to the spectrum and other quantities related to localization theory. Then I will introduce the important notion of *localization length*, the value of which governs the exponential decay for the localized waves that are proven to exist always, irrespective of the disorder, in one-dimensional media with a certain randomness. Then I will introduce the so-called Lyapunov exponents, that can be shown to be positive (Delyon et al., 1985; Souillard, 1986; Carmona and Lacroix, 1990; Pastur and Figotin, 1991; Pastur, 1994) and to be related to the localization length. Both these quantities characterize the behaviour of the localized waves and can be given an explicit meaning in the one-dimensional systems, consisting of an identically independent distributed (scattering) potential, yielding a random multiplicative chain of harmonic matrices[19]. The norm for the product of these random matrices

[18] Quite recently there has been work done on the elastic wave equation (Kohler et al., 1996).

[19] Also known in the seismic literature as propagator matrices (Gilbert and Backus, 1966; Ursin, 1983; Berkhout, 1987; Wapenaar and Berkhout, 1989).

can be shown to be a *self-averaged* quantity for which the Lyapunov exponent constitutes a frequency dependent estimate (Baluni, 1985; Souillard, 1986; Carmona and Lacroix, 1990; Pastur and Figotin, 1991; Pastur, 1994). It was shown by Osoledec (1980) that, for the above situation, the Lyapunov exponents are directly related to the reciprocal of the *localization length* and this opens the way to a direct quantitative analysis by inspecting the wavefield. I will postpone this discussion to section 4.7.

4.6.1 More on the nature of the spectrum

Before going into detail on the behaviour of products of transfer matrices I like to re-address certain important aspects related to the nature of the spectrum. This concerns, specifically, the singular continuous part of the spectrum. This singular part yields *chaotic* waves. Unfortunately this type of spectrum is very difficult to comprehend and as far as the author's knowledge is concerned there is still little known, besides certain deterministic situations described by Pearson (1988) and possibly random situations (Pastur and Figotin, 1991), on its emergence. At least it is striking that an important part of the scattering, localization and spectral theory is aimed at proving the absence of this singular continuous part of the spectrum. For example Reed and Simon (1980) write on the canonical decomposition: *"This decomposition will recur in a quantum-mechanical context where any state will be a sum of bound states (μ_{pp}), scattering states (μ_{ac}), and states with no physical interpretation (one of our hardest jobs will be to show that this last type (singular continuous) state does not occur; that is, that certain measures have $\mu_{sc} = 0$)"*. Indeed the proofs of asymptotic completeness suggest exactly this absence! The difficulty surrounding the singular continuous spectrum refer to the notion that it belongs to the continuous part but it is *singular* with respect to the Lebesgue measure. Think for example of the Cantor set I introduced in chapter 6. This Cantor set is continuous but it lives on a complete different set than the Lebesgue measure, see Reed and Simon (1980) pg. 21.

Now, what if localization theory exactly derives its mere existence from the notion that spectral measures being singular with respect to the Lebesgue measure are the ruler of the game? Unfortunately this appears to be exactly the case (Faris, 1995) and there are only two types of spectra which have this property, namely the singular continuous part and the pure point part. I find the underlying reason for this necessity very difficult to comprehend. Faris (1995) gives in his overview paper a relative simple explanation which goes along the following lines. He starts with the observation that the solutions to the eigenvalue problem for waves in random media give indication of exponential growth in both space directions. Physically speaking there is a problem here. On first sight this problem may appear to be simple. Just supplement a set of proper random initial conditions at infinity, drawn from the absolute continuous part of the spectrum, counterbalancing the exponential growth behaviour. Given the random nature of the problem at hand this is unfortunately impossible because the probability that the two

initial conditions, at the two endpoints towards $\pm\infty$, are the same is zero. In that case one has to resort to other means and it is beyond the scope of this thesis to go into detail one how this is exactly done. I will limit myself to stating that, with probability one, one can find a set of eigenfunctions that are suitable. The condition for the pertaining eigenvalues is that they have to belong to the singular continuous or pure point part of the spectrum.

For obvious reasons the pure point part of the spectrum is preferred since it excludes the emergence of chaotic type of wave functions that are attributed to the singular continuous part of the spectrum (Faris, 1995). This implies that one wants to rule out, as in scattering theory, the singular continuous spectrum. Within the random context this corresponds to imposing a condition for a sufficient degree of independence, locality, for the random fluctuations, i.e ruling out media with long range correlations. For the one-dimensional situation it appears to be sufficient to demand an exponential decaying correlation function for the heterogeneity. Think for example of the Ohrnstein-Uhlenbeck process or the random telegraph model, see chapter 2. In that case it can be proven that the wave interferences give rise to a pure point spectrum where the eigenvalues are countable and lie dense (Souillard, 1986; Pastur and Figotin, 1991; Pastur, 1994; Faris, 1995). The latter property is very interesting because in that situation the wave packet's time evolution will strongly depend on the degree of coarse graining of the spectrum, an observation not in contradiction with the observed dispersion. In this case the dispersion relation will not longer display a trivial linear relationship. This excludes the existence of a single velocity connecting space and time. Moreover, at this point the argument of Pastur (1994) can be used proving that the absolute continuous part of the spectrum is empty in case the Lyapunov exponent is larger than zero, see below.

Let me stress again that the theoretical body surrounding this theory is very difficult and abstract and is still in development. That may explain the possible lack of theoretical insight on the relation between the nature of the material properties or equivalently the scattering potential and the nature of the spectrum. However, that does not withstand an analysis of the physical states with the aim to issue statements on the nature of the spectrum, and subsequently possibly feeding this back to the nature of the medium.

4.6.2 Lyapunov exponents and localization length

If one assumes the acoustic waves to be exponentially localized then one is able to designate an exponent that expresses the exponential decay. This exponent is called the *localization length*, $\xi_l(\omega)$, which is a function[20] of the temporal frequency and which

[20]The careful reader may have noticed that the author tacitly switched from evaluating the time evolution to the space evolution. In the one-dimensional situation this can be done without causing a problem.

governs the decay of the wave functions

$$|P^{\pm}(x_3,\omega)| \sim e^{-\frac{x_3}{\xi_l(\omega)}} \quad x_3 > 0, \tag{4.24}$$

as a function of the vertical travel distance x_3. It was proven by Osoledec (1980), see for detail (Carmona and Lacroix, 1990), that the frequency dependent localization length is directly related – under certain assumptions on the randomness such as a finite correlation length for the potential/medium parameters and a translation-invariance for the mean (Pastur, 1994) – to the *Lyapunov exponents* (Baluni, 1985; Souillard, 1986; Siebesma, 1989). These Lyapunov exponents are self-averaged coefficients and are positive when the medium is random enough. They emerge in the study of products of random matrices and beyond. The relationship between these Lyapunov exponents and the localization length reads

$$\xi_l(\omega) = \gamma(\omega)^{-1}, \tag{4.25}$$

where $\gamma(\omega)$ is the Lyapunov exponent or the Lyapunov spectrum.

4.6.3 Systems of random transfer matrices

What can be said on the limiting behaviour of the transmitted wavefield in case it is the result of a cumulative multiplicative action invoked by the random transfer/propagator matrices[21]? These transfer matrices relate the acoustic wavefield[22] at either side of an interval. Scattering matrices perform a similar task by relating the up- and downward travelling wave constituents across the interval. Only in that case the ordinary matrix multiplication is replaced by Redheffer's starproduct (Redheffer, 1961). To answer the question on the limiting behaviour of the random transfer matrix products one has to go back to Anderson's tight-binding model. In that model the behaviour of an electron in a random one-dimensional potential field is described. This problem corresponds to the evaluation of a so-called disordered harmonic chain (Souillard, 1986; Siebesma, 1989; Baluni, 1985) of random transfer matrices. It can be shown that the norm for the product of the constitutive random transfer matrices grows exponentially with the number of layers, N,

$$\lim_{N\to\infty} \frac{1}{N} \ln \|\Pi_{n=1}^{N} \mathbf{M}_n \mathbf{u}_0\| = \gamma(\omega) > 0. \tag{4.26}$$

In this result, proven by Fürstenberg (1963), the symbol $\gamma(\omega)$ refers to the strict positive Lyapunov exponent. The symbols $\mathbf{M}_n$ and $\mathbf{u}_0$ refer to the consecutive propagator

[21] In the seismic context think of the propagator method (Gilbert and Backus, 1966) or of the so-called layercode schemes (Kennet and Kerry, 1978). All these matrix methods bear in common that they ensure continuity of the pressure and of the (normal component of the) particle velocity across the interfaces separating the consecutive layers.

[22] It does not matter whether one looks at the one- or two-way wavefields since it is assumed that the medium is homogeneous beyond the level x_3. So one only has to do with the downgoing pressure constituent.

matrices and the initial conditions for the wavefield respectively. Since the harmonic transfer matrices depend on the temporal frequency it is clear that the Lyapunov exponents themselves will also depend on the frequency. It was proved by Fürstenberg (1963) that the above result holds for almost every realization and that expresses the notion that the Lyapunov exponents refer to a spatial *self-averaging* process. With other words, the outcome of equation (4.26) does not depend on the specific details of one realization. Hence, the definition of the Lyapunov exponent does not require an indispensable ensemble averaging. Now how can one use this result to one's advantage? At this point the theorem of Osoledec (1980) comes at hand since it relates the Lyapunov exponent to the localization length and the latter can readily be obtained from wavefield data. In case the Lyapunov exponent is positive one has to do with exponentially decaying eigenmodes, while the corresponding eigenvalues lie, under the condition of sufficient locality of the randomness, dense in the pure point part of the spectrum. This implies that in this situation one no longer has to do with ordinary extended wave modes which are the solution of the wave equation with constant or almost constant coefficients. Note again that the possible emergence of chaotic type of waves can only be circumvented by excluding the singular continuous part of the spectrum and hence imposing conditions on the correlation displayed by the medium fluctuations. Finally remark that in a disordered medium *ad infinitum* under these conditions the one-dimensional "waves" will always be exponentially localized irrespective of the amount of disorder. Only the weaker disorder the longer the localization length becomes for a fixed frequency.

4.6.4 Asymptotic behaviour

Given the above relationships and observations one is now set to issue some statements on the qualitative behaviour of the frequency dependent localization. Following Baluni (1985) one can assess the asymptotic behaviour for two regimes namely the high and low temporal frequency regimes. According to Baluni (1985) these two regimes refer to the weak and strong disorder regimes. A system in its weak order regime is a system where the fluctuations in the quantity $k_3(\omega)\Delta x_3$ are regarded as being small. The variances of the fluctuations, with respect to the mean, become small in case the temporal frequency becomes small. This is because the fluctuations consist of the product between randomly varying layer thicknesses, Δx_3, and the alternating[23] wavenumbers $k_3(\omega)$, which are fixed and deterministic. For weak disorder, $\langle (k_3(\omega)\Delta x_3)^2 \rangle \ll 1$, Baluni (1985) shows that the frequency dependent Lyapunov exponents converge to a frequency squared dependency, $\gamma(\omega) \sim \omega^2$ for $\omega \to 0$. For the strong disorder regime on the other hand, $\langle (k_3(\omega)\Delta x_3)^2 \rangle \gg 1$, the Lyapunov exponent becomes a constant, $\gamma(\omega) \sim C$ for $\omega \to \infty$, an observation striding well with the fact that in the geometrical optical regime the "wave" only sees the consecutive transmission coefficients. The multiplicative action of these transmission

[23] Baluni (1985) uses an alternating sequence of layers consisting of two different vertical wave numbers and with randomly distributed layer thicknesses. Notice that this is exactly the random telegraph model.

coefficients results in a frequency independent exponential decay as a function of the propagation distance.

To interpret the above results it helps to take into consideration that the primary energy of the wavefront becomes, in the course of the propagation, completely redistributed due to the scattering induced by the heterogeneities. Eventually this mechanism kills the primary wavefront, the $\omega \rightarrow \infty$ "contribution", because the magnitude of the product of the consecutive transmission coefficients tends to zero as the number of encountered layer transitions increases. Now the common conjecture is that it is still possible to have a coherent pulse propagating in case of a *constructive* interference between the even multiple reflection events and the vanishing primary (Banik et al., 1985a,b; Resnick et al., 1986; Burridge et al., 1988; Herrmann, 1991; Herrmann and Wapenaar, 1992, 1993). To better understand this complex mechanism many attempts have been made to recognize – e.g. via approximate descriptions for the wave motion – the ruling quantities responsible for this intricate interference mechanism.

The interesting point of the above analysis, is that it delivers the same asymptotes as obtained for the O'Doherty-Anstey formula in combined action with the random telegraph model (Banik et al., 1985b). This random telegraph model yields an exponential decay for the correlation of the medium fluctuations and this behaviour is shared by most random processes that are being used to characterize the randomness in the context of wave interactions (Burridge and Chang, 1989; Banik et al., 1985a,b).

From my experience, see chapter 2, I do not expect to find an exponential decay for the correlation function. This would namely imply a break in the scaling, a notion *not* substantiated by the scaling displayed by the well-log data. On its turn this may suggest that *the singular continuous part of the spectrum can not be excluded!*

4.7 An alternative approach

Even though one can probably not rule out the existence of the singular continuous part of the spectrum, one can postulate equations (4.24)-(4.26) to hold. In that case one can try to infer information on the general properties displayed by the Lyapunov spectrum. To start the discussion I included figure 4.6 where I depicted the amplitude and Lyapunov spectrum for the bottom most traces of the real and synthetic VSP data sets depicted in figures 4.1 and 4.2. From these plots it is clear that these spectra display an irregular behaviour, an observation certainly applying to the synthetic VSP, generated using a transfer matrix modelling scheme. At this point let me postulate the ansatz that the Lyapunov spectrum potentially displays a powerlaw type of frequency behaviour. This conjecture is mainly based on two motivations.

The first refers to the existence of a relative large body of seismic literature describing

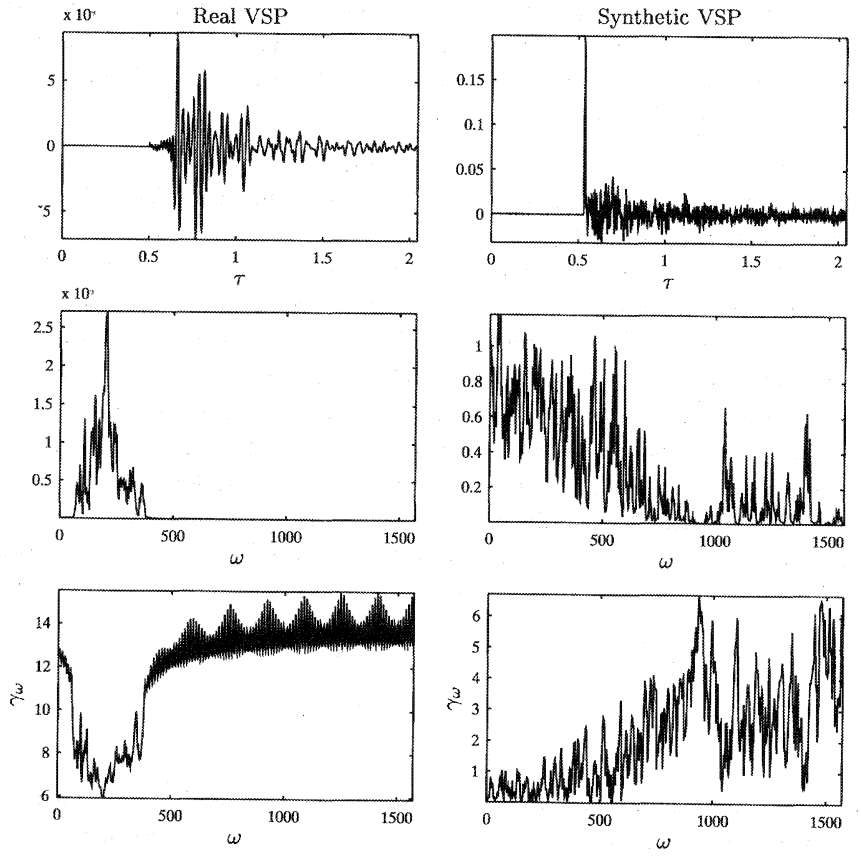

Figure 4.6 *This figure contains the two bottom most traces taken from the VSP's depicted in figure 4.1 and 4.2. From these traces I computed the amplitude spectrum, middle row and the Lyapunov spectrum, bottom row. The real VSP related quantities are depicted in the left column. The simulated quantities in the right column.*

a (quasi) constant-Q type[24] of behaviour for the dispersion. In case the Q factor is independent of frequency one has, inevitably, to do with a powerlaw type of frequency dependent attenuation, or in other words a powerlaw for the Lyapunov spectrum. Although this type of behaviour is generally attributed to a powerlaw type of behaviour for Boltzmann's stress-strain relations, yielding a viscous relaxation mechanism, it is an important statement. It refers namely to the empirical observations concerning seismic data and beyond, reported on by Kjartansson (1981). He demonstrates that the

[24]Here Q is defined as $Q = \frac{\langle \Delta E \rangle}{\langle E \rangle}$, where $\langle E \rangle$ is the mean energy stored in on period ($T = \frac{2\pi}{\omega}$) and $\langle \Delta E \rangle$ the energy loss during one cycle.

Q is related to the exponent α_{ss}. This α_{ss} rules the stress-strain relation according to $Q^{-1} = \tan \pi\alpha_{ss}$. Moreover he already mentions: *"The empirical observations that Q in solids varies much slower than even the square root of frequency, is thus an expression of the statistical nature of the heterogeneities."*[25].

The second motivation refers to the findings of the multiscale analysis and characterization reported on in chapter 2. From this analysis it became clear that a large number of scaling exponents are required to capture the scaling. However, one can speculatively imagine that the scaling of the Lyapunov spectrum is dominated by one of the scaling exponents. Of course, if so, it remains to be shown which one. In the current literature there are already some references towards a scaling type of behaviour (Alexander, 1989; le Méhauté, 1991; Lapidus and Pomerance, 1993; le Méhauté, 1995), displayed by the Lyapunov spectrum and its via the Hilbert transform related integrated density of states (Souillard, 1986; Carmona and Lacroix, 1990; Pastur, 1994). First of all le Méhauté (1991) proposes to replace the time derivative by a fractional differentiation, the degree of which is related to the fractal dimension. This replacement implies a powerlaw behaviour for the Lyapunov spectrum and yields a generalized type of diffusion as coined by Mandelbrot (1982). In that case the space-time behaviour of the field that is given by

$$P(x_3, \omega) \sim e^{-(j\omega)^{\alpha_d} x_3}, \quad 0 < \alpha_d \leq 2, \tag{4.27}$$

where α_d is, for obvious reasons, related to the conjectured scaling for the Lyapunov spectrum. It is interesting to note that equation (4.27) can also be recognized as being generated by a generalized Laplacian, yielding a Hamiltonian of the type $\mathcal{H} = F(-j\nabla) = |\mathbf{p}|^{\alpha}$, $0 < \alpha \leq 2$, where $\mathbf{p}$ is the momentum. In this context this momentum can be interpreted as the temporal frequency. Finally it was Bickel (1993), followed by Hargreaves (1992), who introduced this type of solution, the functional form of which corresponds to the *stable*[26] and infinitely divisible Levy distribution[27] (Samarodnitsky and Taqqu, 1994; Carmona and Lacroix, 1990; Reed and Simon, 1979). Remark that the solution of this generalized type of diffusion are not longer hyperbolic; the time becomes *irreversible* and the solutions are *smooth*. With other words, the solutions do not longer constitute a *group*, but form a *semi-group* instead. Moreover the solutions do not longer propagate but dilatate (le Méhauté, 1995).

[25] Kjartanson argues that: *"in materials with sharply defined heterogeneities (e.g. grain boundaries or pores), absorption controlled by diffusion, such as phase transformations or thermal relaxation, leads to a Q proportional to $\omega^{\frac{1}{2}}$ and $\omega^{-\frac{1}{2}}$ at high and low frequencies, even for uniform distributions of identical pores or crystals."*

[26] Stable under the operation of convolution.

[27] Also known as Pareto-Levy distribution.

4.7.1 The proposed method

The aim of this section is to introduce a way to infer information on the nature of the Lyapunov spectrum. The idea is to capture the scaling exponent which is conjectured to rule the Lyapunov spectrum and that is potentially related to the scaling displayed by the medium fluctuations. Let me remark again that the anticipated scaling exponent derives its mere existence from the scaling ansatz.

From the multiscale analysis presented in chapters 2 and 8 it became clear that the continuous wavelet transform constitutes a perfect vehicle to locally estimate the Hölder exponents. Given this notion together with the fact that the powerlaw shaped Lyapunov spectrum maps to a singularity at the origin one is able to set up an "inversion scheme" for this exponent. The main advantage of this type of approach is that the Lyapunov spectrum itself displays a very erratic behaviour while its inverse Fourier transform, see figure 4.7, behaves rather "smoothly", displaying a nice singularity around the origin. The scaling ansatz for the Lyapunov exponents yields the following behaviour

$$\gamma(\omega) \sim \omega^{\alpha_d}, \tag{4.28}$$

which translates to the following behaviour in the "time" domain

$$\gamma_\tau(\tau) \sim \tau^{-(\alpha_d+1)}, \tag{4.29}$$

yielding the following behaviour for the maximum of the wavelet coefficients

$$\max |\mathcal{W}\{\gamma_\tau, \psi\}(x, \sigma)| \sim \sigma^{-(\alpha_d+\frac{1}{2})}. \tag{4.30}$$

The procedure I propose comprises the following main steps:

- a Vertical Seismic Profile (VSP) – gather with the propagated pressure wavefield as a function of depth – are Fourier transformed followed by taking the logarithm of the real part and a subsequent inverse Fourier transform. With these operations the Lyapunov exponents $\gamma_\tau(\tau)$ are obtained for each depth level, i.e. $\gamma_\tau(\tau; x_3)$.
- these $\gamma_\tau(\tau; x_3)$-gathers are wavelet transformed.
- the maxima of the wavelet transform are picked for a selection of scales.
- linear regressions are computed, for all traces, on the logarithm of the maxima versus the logarithm of the scale indicator σ. The slope of these regressions is related to localization scaling exponent α_d for the vertical offset x_3.

4.8 Concluding remarks

In this chapter an attempt has been made to briefly sketch the current state of science concerning wave propagation phenomena in media with a large amount of heterogeneity.

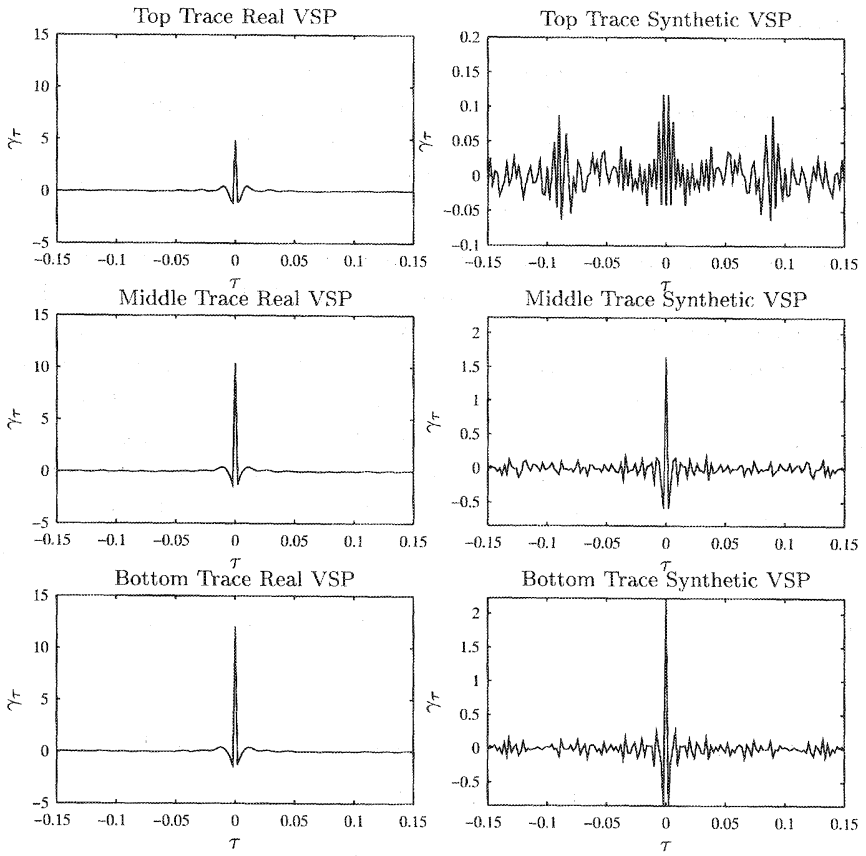

Figure 4.7 *This figure contains the inverse Fourier transform of the Lyapunov spectrum. Again the traces are taken from the real and simulated VSP's of figures 4.1 and 4.2 but now they concern the top, middle and lower traces. On the left the real data is depicted while on the right simulated data is depicted.*

It was mentioned that there exists a rigorous but very involved theory that predicts exponential localization to occur in one-dimensional media and for one-dimensional acoustic waves. The proof for this result is based on the assumption that the correlation function displays a rapid enough decay. This implies that the medium fluctuations must be independently distributed or the correlation function must at least be exponentially decaying. When this condition is not met, the singular continuous spectrum can not longer be excluded and this may give rise to a type of wave functions being void of a physical meaning. However, when sufficient independence for the fluctuation is guaranteed then the spectrum becomes pure point with its eigenvalues lying dense. Physically speaking this corresponds to exponentially localized waves which are strictly speaking not longer

waves. At this point it is interesting to link up with the Anstey-O'Doherty formula which predicts a similar kind of exponential localization. This approximate Anstey-O'Doherty formula is expressed in terms of the correlation function of the reflectivity and is only accurate, when applied properly, in its description of the generalized[28] primary in a co-moving frame of reference. This co-moving frame is defined in terms of the specifics of a realization while the generalized primary refers to the coherent part of the transmitted signal. This coherent part consists of a mutual interference of the vanishing primary with the multiple reflections. Given this aim of the Anstey-O'Doherty formula it is not surprising that its predictions combine well with the findings of localization theory because the short range correlations rule the behaviour of the generalized primary.

Despite the fact that – due to the empirically found long tailed correlations, exhibited by well-log measurements – the singular continuous part of the spectrum can not be excluded, the Lyapunov spectrum may carry information on the scaling of the medium. Indeed the example discussed in his chapter gives at least some indication of scaling and possibly exponential localization. The actual value of the estimated exponent, however, does not match with the value found for the synthetic VSP. The explanation for this difference remains open. However, qualitatively one can say that the real VSP data indeed disperse. But they do not disperse as strongly as a generalized diffusion predicts. In that case namely waves do not propagate at all and that observation does not combine well with the evidenced behaviour. This leaves me to postulate that there must be some kind of combined effect. This combination consists of a wave like hyperbolic behaviour and a non-hyperbolic diffusion kind of behaviour.

To summarize: only a very limited view on waves interactions with complex media is presented in this chapter. The local wave interaction responsible for the reflections has not been addressed at all. For the interested reader I present in the *Epilogus* some specular reflections concerning my own premature ideas on how one can possibly look at waves in scaling media.

[28]The term generalized primary was coined by Resnick et al. (1986). The term pseudo primary is also used and was first coined by Burridge et al. (1988).

Part II

Capita Selecta

Chapter 5

Selected topics on distribution theory

5.1 Introduction

One of the main motivations for this thesis is formed by Schwartz's distribution theory (Schwartz, 1957; Gel'fand and Shilov, 1964; Zemanian, 1965; Griffel, 1981; Dautray and Lions, 1988; Strichartz, 1994). Besides the fact that this theory provides a formal justification – within the realm of functional analysis – of the original ideas on a symbolic calculus coined by Heaviside and Dirac, it also draws a transparent parallel between the detection of a physical *event*, a shock or a pulse, and the detection of a *singularity*. Both these mathematical and physical mechanisms bear in common that they are based on a non-linear averaging – over a non-vanishing interval – by an instrument or test function. From the mathematical point of view this method boils down to generalizing the concept of a *function*, which *locally* assigns a value to its independent parameters, to that of a *functional*. A functional is a rule that assigns a value to each function in a set of so-called *test functions* of non-vanishing support[1]. In this treatise a special type of functionals will be treated solely, namely that of the *distributions*. These distributions – being obtained after imposing the additional constraints of *continuity* and *linearity* on the functional – provide a powerful framework to analyze, synthesize and characterize physical processes. Such a framework not only refers to the formal definition of *operators* in terms of distributions, capturing the dynamics of the phenomenon, e.g. the impulse response of a system, but it also provides the fundamental basis for the multiscale analysis techniques[2] presented in this thesis.

The common denominator of this treatise is to select and define a proper vector space of test/analyzing functions on which the multiscale analyzing technique is given a meaning

[1] With a non-vanishing support is meant that the support of the function can not become of measure zero.

[2] The multiscale analysis techniques are introduced in chapter 8 and are based on the continuous wavelet transform. This scale transformation conducts the multiscale decomposition and adheres to the conditions imposed by the distribution theory discussed in this chapter.

by the mathematical method of *testing*, i.e. measuring them in the proper functional space. Obviously, the proposed testing procedure, the mapping of the test function to an *observable*, a finite number, by the *functional* introduces an inevitable *uncertainty* being a direct reflection of the intrinsic boundedness in the experiment's resolution or the non-vanishing condition imposed on the test function's support.

In the sequel I convey the basic ideas behind distribution theory and the associated concepts such as *observables* and *uncertainty*. I will commence the discussion by formally defining distributions and the associated test functions. Then I will review a special type of *singular* distributions, Hadamard's pseudo functions or functions with algebraic singularities (Gel'fand and Shilov, 1964; Zemanian, 1965), followed by a discussion on how to regularize them. Then I will show that it is possible, via the concept of the *adjoint identity*, to define certain operations on these distributions. Amongst these operations there is also the operation of *convolution* that plays an all important role in capturing the spatial-temporal dynamics of physical phenomena. This convolution performs an equally important role in distribution theory where it allows for the method of *regularization*, the spatial-temporal smoothing process by the instrument.

So in the pertaining sections I will briefly touch upon the method of regularization, fractional integration/differentiation as examples of operations based on convolution. In the end I will conclude this chapter by providing a number of definitions allowing for the characterization of the degree of differentiability annex regularity by means of the so-called Hölder exponents.

5.2 Measuring, test functions and distributions

The theory of distributions is a reflection of the observation that the concept of a *function* $f(x)$ – *locally* assigning a value to a physical variable at a point x – is difficult to comprehend. This difficulty arises from the fact that, without exception, the measuring process is being intervened by an inevitable bound on the obtainable resolution in space-time. With other words, there is always an intrinsic *uncertainty* being a direct consequence of the fundamental inability of the measuring device to come up with an instantaneous and/or strict spatially confined measurement in space-time. For example, consider measuring the transient temperature distribution at varying positions x, $x \in [x_0, x_1]$ and over a temporal interval $t \in [t_0, t_1]$. From physical arguments it is clear that in this situation the perception of a function fails since the space-time localizing capabilities of the thermometer are constrained by the physical size of its bulb and by its inertial characteristics. So in effect, it can be stated that a measurement in a physical experiment is always the result of some sort of averaging procedure mapping the unknown physical quantity to an *observable*. This mapping, the measuring process, corresponds mathemat-

ically to evaluating the following *scalar product*[3], when only the spatial characteristics are concerned,

$$\langle f, \varphi \rangle \triangleq \int_{-\infty}^{+\infty} f(x)\varphi(x)\mathrm{d}x, \tag{5.1}$$

where f symbolizes the unknown, locally integrable physical quantity, $f \in L^1_{\mathrm{loc}}$, whereas $\varphi(x)$ symbolizes the characteristics of the measuring device which is assumed to be of non-vanishing support, the diameter of its bulb is finite, and with, $\varphi(x) \geq 0 \ \wedge \forall x \in \mathbb{R}$ and $\int \varphi \mathrm{d}x = 1$. By repeating this measuring procedure at different locations, i.e. using shifted versions of φ, it is possible to obtain a finite-resolution spatial profile for the temperature distribution.

Given the above considerations it is evident that the rather limited conjecture of a *function* needs to be replaced by that of a *functional*, a rule that assigns a number to each function in a set of so-called *test functions* (Zemanian, 1965). Such a formalism allows for the mathematical justification of the measuring procedure as depicted in equation (5.1) and leads, after imposing the additional technical conditions of *linearity* and *continuity* to the concept of a *distribution*. These distributions constitute the mapping of a suitable test function of non-vanishing support to a number,

$$\varphi \quad \mapsto \quad \langle f, \varphi \rangle, \tag{5.2}$$

where the physical entity has been represented by the distribution which calculates the "average" over the test function's support.

By way of the above construction, the duality of the test function and the entity f, the distribution, facilitates a much larger class of entities to be associated with the symbol f than the conventional class of functions can. Of course, such an extension of the function entity to the entity of a *distribution* or *generalized function*, entails the vector space of test functions – by imposing specific conditions on it – depending on the "smoothness" of the distribution annex generalized function.

5.2.1 The space $\mathcal{D}$ of test functions

In preparation to the introduction of distributions it will be necessary to define a *linear* vector space of test functions on which these distributions are going to operate meaningfully. Consider for this purpose the set $\mathcal{D}$ of complex valued functions φ which are of *compact support* and which are infinitely smooth on their support as well as on their endpoints, i.e. $\mathcal{D} \subseteq C_0^\infty$ (other alternatives will be considered later). It was Schwartz

[3]Note the use of the word *scalar*. This is to make a distinction between equation (5.1) and inner products, where complex conjugates are used. However, since equation (5.1) symbolizes a measurement, the use of complex conjugates is not necessary and therefore the equation will be referred to as scalar product.

(1957) who proposed this class of test functions $\mathcal{D}$ of which the function,

$$\varphi(x) = \begin{cases} e^{\frac{1}{x^2-1}} & \text{for } |x| < 1, \\ 0 & \text{otherwise,} \end{cases} \tag{5.3}$$

is a well known member with all its derivatives vanishing at the endpoints and beyond, $|x| \geq 1$.

At this point it can be stated, following Zemanian (1965), that *any complex valued function $f(x)$ that is continuous, not necessarily differentiable, for all x and zero outside a finite interval can be approximated uniformly by test functions.* This corresponds to the existence of a test function, $f_\sigma \in \mathcal{D}$, for which given an $\epsilon > 0$, $|f(x) - f_\sigma(x)| \leq \epsilon$. Let me illustrate how such an uniform approximation can be obtained. First, construct a set of properly normalized test functions, $\int \psi_\sigma(x)\mathrm{d}x = 1$,

$$\psi_\sigma(x) \triangleq \frac{\varphi(\frac{x}{\sigma})}{\int_{-\infty}^{+\infty} \varphi(\frac{x}{\sigma})\mathrm{d}x} \qquad \sigma > 0. \tag{5.4}$$

Subsequently, use this sequence, the sequence of test functions of decreasing support (for $\sigma \downarrow 0$), to uniformly approximate the function $f(x)$ by means of the following integral representation,

$$f_\sigma(x) \triangleq \int_{-\infty}^{+\infty} f(x')\psi_\sigma(x - x')\mathrm{d}x'. \tag{5.5}$$

Clearly, this integral can be recognized as a *convolution*, an important concept in the mathematical formulation of the spatial-temporal dynamics in a physical measuring procedure. In sections 5.5 and 5.7 I will come back to this important issue.

Taking the limit $\sigma \downarrow 0$ in the integral on the right hand side in equation (5.5) leads to the exact recovery of the function $f(x)$ but this limiting procedure is not *trivial* and requires the introduction of the Dirac δ-distribution, an example of a *singular* distribution.

5.3 Regular distributions

Before going into detail on the properties of singular distributions I would like to provide the formal definition for a functional and quickly touch upon issues related to *linearity* and *continuity*. First let me restate that a functional is a rule that assigns a number to every member of a set of functions, in this case the space $\mathcal{D}$. The functional is denoted by the symbol f and the number it assigns to $\varphi \in \mathcal{D}$ is given by $\langle f, \varphi \rangle$. Now a distribution can formally be defined:

Definition 5.1: Distributions
A continuous linear functional on the space $\mathcal{D}$ is called a distribution.

In this definition the *linearity* refers to the property, $\langle f, a_1\varphi_1 + a_2\varphi_2\rangle = a_1\langle f, \varphi_1\rangle + a_2\langle f, \varphi_2\rangle$, where $\varphi_1, \varphi_2 \in \mathcal{D}$ and $a_1, a_2 \in C$ and the *continuity* refers to the notion that, for any sequence of test functions $\{\varphi_n(x)\}_{n=1}^{\infty}$ that converges in $\mathcal{D}$ to $\varphi(x)$, the sequence of numbers $\{\langle f, \varphi_n\rangle\}_{n=1}^{\infty}$ converges to the number $\langle f, \varphi\rangle$ (Zemanian, 1965). The convergence of the φ_n to φ means that the supports of the φ_n are in a fixed interval and that the φ_n and all their derivatives converge uniformly to φ and all its derivatives. The space of these distributions is indicated $\mathcal{D}'$, the *dual* or *conjugate* space of $\mathcal{D}$.

Before elaborating more extensively on the mathematical aspects of distributions let me briefly review the concept of a distribution, its linearity, continuity and its convergence, in the physical context of the thermometer experiment. In this experiment the actual spatial-temporal temperature distribution is the subject measurement and this, of course, exactly corresponds to the role of the functional when, for example, the spatial characteristics of the temperature distribution are of interest. In this case the *linearity* refers to the notion that it is possible, in most cases, to combine two separate measurements into one by means of taking the average of the two measurements obtained from two locations with not necessarily the same thermometers. In that case the following expression,

$$a_1 \int_{-\infty}^{+\infty} f(x)\varphi_1(x)\mathrm{d}x + a_2 \int_{-\infty}^{+\infty} f(x)\varphi_2(x)\mathrm{d}x = \int_{-\infty}^{+\infty} f(x)[a_1\varphi_1(x) + a_2\varphi_2(x)]\mathrm{d}x \tag{5.6}$$

holds and where $a_1 = a_2 = \frac{1}{2}$ in case the same thermometer is used.

The *continuity*, on the other hand, refers to the notion that measurements taken at one spot but with different thermometers will not differ much in case the thermometers do not deviate too much. This latter property opens the way to refer to the concept of convergence, the convergence of the measured temperature distribution to the actual one, as the spatial accuracy of the thermometer is being increased, in terms of a sequence of test functions. In order to illustrate this conjecture let me first define again the integral, under the afore mentioned conditions and the condition of local integrability, in the Lebesgue sense, $f \in L^1_{\mathrm{loc}}$,

$$\langle f, \varphi\rangle \triangleq \int_{-\infty}^{+\infty} f(x)\varphi(x)\mathrm{d}x, \tag{5.7}$$

where the limits in the integral can be replaced by finite values since φ is of finite support.

Clearly, the above integral defines a *linear* functional and its continuity can be made explicit by consideration of the sequence $\{\varphi_n\}_{n=1}^{\infty}$ that converges to $\varphi(x)$ in $\mathcal{D}$. This convergence refers to the fact that for a given $\epsilon > 0$ one can find a $N \in \mathbb{N}$ such that, for all $n \geq N$, $|\varphi(x) - \varphi_n(x)| \leq \epsilon$. On its turn this property leads to the observation

$$\lim_{n\to\infty} |\langle f, \varphi\rangle - \langle f, \varphi_n\rangle| = 0. \tag{5.8}$$

since

$$\begin{aligned} |\langle f,\varphi\rangle - \langle f,\varphi_n\rangle| &= \left|\int_{-\infty}^{+\infty} f(x)[\varphi(x)-\varphi_n(x)]\mathrm{d}x\right| \\ &\leq \int_a^b |f(x)||\varphi(x)-\varphi_n(x)|\mathrm{d}x \\ &\leq \epsilon\int_a^b |f(x)|\mathrm{d}x. \end{aligned} \tag{5.9}$$

Remark that the identity in equation (5.8) is made possible by selecting the bounds of integration in equation (5.9) in accordance to the supports of $\varphi(x)$, $\varphi_n(x)$ and by invoking the local integrability condition, $f \in L^1_{\mathrm{loc}}$, yielding a finiteness of the quantity $\int_a^b |f(x)|\mathrm{d}x$. So as a consequence of the two afore mentioned properties it can be stated that the integral in equation (5.7) defines a type of distribution which is called *regular*. The *identity* for this type of distribution states that the distributions $f, g \in \mathcal{D}$ differ at most on a *set of measure zero* in case $\langle f,\varphi\rangle = \langle g,\varphi\rangle$ and this corresponds to one of the basic properties of Lebesgue integration theory where sets of measure zero are unimportant. But, on the other hand, it can be stated that in case f, g differ on a set not of measure zero, then $\langle f,\varphi\rangle \neq \langle g,\varphi\rangle$ for at least one test function, $\varphi \in \mathcal{D}$. From this argument it is deduced that each test function of $\mathcal{D}$ uniquely determines a regular distribution of $\mathcal{D}'$ and this corresponds to $\mathcal{D} \subset \mathcal{D}'$ (Zemanian, 1965).

Besides the fact that the space of distributions $\mathcal{D}'$ contains the space of test functions, $\mathcal{D} \subset \mathcal{D}'$, it may also contain a much wider class of distributions which are not necessarily *regular*. However, this extension, the accommodation of a much "wilder" class of objects, goes, in certain cases, at the expense of loosing operations such as $f(x)g(x)$ or $f(g(x))$. But fortunately most other operations, such as integration/differentiation, can be given a meaning via a limiting procedure, they are to be defined in the *weak sense* or in the *sense of distributions*, a notion to be clarified below.

5.4 Singular distributions

The Dirac δ-distribution serves as an excellent example of a distribution that is *not* regular, that is not confined to the class of ordinary regular functions, but instead it is a member of the class of *generalized functions* (Gel'fand and Shilov, 1964). This Dirac δ-distribution is implicitly defined by the linear functional on $\mathcal{D}$,

$$\langle \delta,\varphi\rangle \triangleq \varphi(0), \tag{5.10}$$

where the quantity on the left hand side has only been given a meaning by virtue of the right hand side. A close inspection of this definition reveals that this functional is *continuous* (Zemanian, 1965). It is also evident that it is impossible to link the definition in equation (5.10) to that of a locally integrable function. Because, in that case the Dirac

δ-distribution would have to comply to the local integrability condition and this would have yielded – as a consequence Lebesgue's integration theory – a convergence to *zero* everywhere as the support of the test function is approaching zero, $\sigma \downarrow 0$. Evidently this observation is in contradiction with the behaviour for the δ-distribution as defined in equation (5.10) and therefore it is argued that the δ-distribution is *not* a regular distribution. To summarize it is stated that all distributions that are not regular are *singular* distributions or *generalized* functions, where the term generalized refers to the notion that $\mathcal{D}'$ can accommodate a wider class of objects than the class of conventional functions can (Schwartz, 1957; Gel'fand and Shilov, 1964; Zemanian, 1965; Strichartz, 1994).

The best way to illustrate the rather peculiar properties of the Dirac δ-distribution and its derivatives[4] is to regard their construction by means of a limiting procedure generated by a sequence of regular functions that converges, in the *sense of distributions*, to the δ-distribution, or its derivatives, as the limit $\sigma \downarrow 0$ is being taken. While composing such a converging sequence, also known as a *delta convergent sequence*, there are only two conditions to be met. Following Gel'fand and Shilov (1964) these conditions are

- For any $M > 0$ and $|a| \leq M$ and $|b| \leq M$, the quantity

$$\left| \int_a^b f_\sigma(x) \mathrm{d}x \right|$$

must be bounded by a constant independent of a, b or σ, only depending on M.

- For any fixed non-vanishing a and b one must have

$$\lim_{\sigma \downarrow 0} \int_a^b f_\sigma(x) \mathrm{d}x = \begin{cases} 0 & \text{for} \quad a < b < 0 \quad \wedge \quad 0 < a < b \\ 1 & \text{for} \quad a < 0 < b. \end{cases}$$

Obviously, the examples

$$f_1(x,\sigma) = \begin{cases} \dfrac{1}{\sigma} & \text{for } 0 < x < \sigma, \\ 0 & \text{otherwise,} \end{cases} \tag{5.11}$$

$$f_2(x,\sigma) = \frac{1}{\sigma}\Gamma(x)e^{-\frac{x}{\sigma}}, \quad x > 0 \tag{5.12}$$

$$f_3(x,\sigma) = \frac{1}{\pi}\frac{\sin(\frac{x}{\sigma})}{x} \tag{5.13}$$

$$f_4(x,\sigma) = \frac{1}{\sigma\sqrt{\pi}}e^{-\left(\frac{x}{\sigma}\right)^2}, \tag{5.14}$$

[4]Later to be defined in a more formal manner in section 5.5.

where $\Gamma(x)$ is the gamma function defined by

$$\Gamma(x) = \begin{cases} \int_0^{+\infty} y^{x-1} e^{-y} \mathrm{d}y & \text{for } x > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{5.15}$$

meet these requirements. Indeed it can be shown that – irrespective of the details of the regular functions – the following limit holds, $\lim_{\sigma \downarrow 0} f_i(x) \to \delta(x) \quad \forall i = 1 \ldots 4$.

Invocation of a delta convergent sequence, see the examples in equation (5.11)-(5.14) into the definition of a regular distribution,

$$\langle f_i(x,\sigma), \varphi(x) \rangle = \int_{-\infty}^{+\infty} f_i(x,\sigma)\varphi(x)\mathrm{d}x, \tag{5.16}$$

while taking the limit of σ towards zero finally yields

$$\lim_{\sigma \to 0} \langle f_i(x,\sigma), \varphi(x) \rangle = \lim_{\sigma \to 0} \int_{-\infty}^{+\infty} \delta_i(x,\sigma)\varphi(x)\mathrm{d}x \triangleq \varphi(0). \tag{5.17}$$

By way of this limiting procedure one has been able to define the Dirac δ-distribution as a limiting form of a regular distribution. This method – based on the construction of a converging sequence – is not limited to the δ-distribution but also allows for a similar definition for its derivatives via the differentiation of the elements in the Delta convergence sequence towards x. Since this approach has not been trivial it is understandable that reversing the argument of equation (5.10) to

$$\varphi(0) = \int_{-\infty}^{+\infty} \delta(x)\varphi(x)\mathrm{d}x. \tag{5.18}$$

cannot be assigned any meaning in the conventional sense of functions since the right-hand side constitutes an *improper* integral. With other words the δ-distribution can not be considered as an ordinary function, a regular distribution. On the contrary it should be interpreted either symbolically or as a limiting procedure converging, in a *weak sense* or in the *sense of distribution*, to the δ-distribution. So, in a way, this type of definition opens the way to define certain manipulations on these constructs that are formally only valid for ordinary functions.

To conclude this rather informal section on the definition of singular distributions I would like to express another way of looking at the non triviality of the δ-distribution and its derivatives. This can be done in the context of Lebesgue integration theory which states that it can not discern "functions" of measure zero, functions the support of which is confined to a single point. So, in effect, the singular distribution has to be measured by a proper test function. This observation corresponds to the operation of effectively blurring (thickening) the singularity and thereby giving it a meaning, making it an *observable*.

Of course, certain conditions, such as continuity in case of the δ-distribution, has to be imposed on the test functions in order to give the singular distributions a meaning. Such an observation calls for the definition of functional spaces which admit a certain irregular behaviour of the distribution. Before going into detail on these aspects I would like to review a somewhat larger class of singularities first, followed by a discussion on how to formally define operations on them.

5.4.1 Algebraic singularities, Hadamard's finite part and regularization

One of the main motivations of distribution theory is that it allows one to deal with "functions" containing singularities. In words: "*a singularity can be defined as a point at which the derivative of a given function of a complex variable does not exist but every neighbourhood of which contains points for which the derivative exists*" (Webster, 1988). A function is *singular* when it contains a *singular* point and this results in a breakdown of the expansion into a Taylor series, i.e. the function contains an "edge", a point of rapid variation.

The only singular distributions (functionals) – a terminology more appropriate than that of singular functions – reviewed so far were the δ-distribution and its derivatives. In this section I will present another type of singular distributions namely "functions" with algebraic singularities or distributions generated by Hadamard's "finite part" of an algebraically[5], divergent integral, the so-called *pseudo functions* (Zemanian, 1965).

Consider the category of functions the derivative of which contains a singularity that does not grow faster – when the singular point is being approached – as the reciprocal of a polynomial of arbitrary order. Due to Hadamard it is then possible to associate a singular distribution to these non-integrable singularities[6], singularities which cause the integral,

$$\int_{-\infty}^{+\infty} f(x)\varphi(x)\mathrm{d}x,$$

to diverge as x approaches the singular point of f in $\operatorname{supp}\varphi$. The crux of this method, also called a regularization by Gel'fand and Shilov (1964), is simple. It just defines a distribution $f \in \mathcal{D}'$ such that for all $\varphi \in \mathcal{D}$, vanishing in the neighbourhood of the singularity, f coincides with $f(x)$ everywhere except at the abscissa of the singularity, as will be explained later.

The afore mentioned algebraic type of singularities naturally arises when the derivatives of

$$f(x) = x_+^\alpha = \begin{cases} 0 & x \leq 0 \\ x^\alpha & x > 0 \end{cases} \qquad -1 < \alpha < 0 \tag{5.19}$$

[5]Notice that the pseudo functions are generally not limited to this algebraic type of singularities.
[6]So, in effect, this is a reflection of the divergence for the derivative of the original function.

are carefully analyzed. Despite the fact that this function is locally integrable, $f(x) \in L^1_{\text{loc}}$, its ordinary first derivative, $f'(x) = \alpha x^{\alpha-1}$, is not. In this case the now divergent integral,

$$\int_0^{+\infty} \alpha x^{\alpha-1} \varphi(x) \mathrm{d}x \tag{5.20}$$

needs to be regularized and that can be accomplished by Hadamard's definition, see Gel'fand and Shilov (1964), Zemanian (1965), of the "finite part", or the "principal value" (Dautray and Lions, 1988; Duistermaat, 1993),

$$\langle \mathrm{Pf}\, f, \varphi \rangle \triangleq \mathrm{FP} \int_0^{+\infty} \alpha x^{\alpha-1} \varphi(x) \mathrm{d}x \triangleq \int_0^{+\infty} \alpha x^{\alpha-1} [\varphi(x) - \varphi(0)] \mathrm{d}x, \tag{5.21}$$

where $\mathrm{Pf}\, f$ refers to the *pseudofunction* f rather than to the ordinary function $f(x)$ and FP refers to taking the "finite part". Equation (5.21) actually defines the desired rule giving the divergent integral in equation (5.20) a meaning. In fact this rule, this functional, boils down to replacing $\varphi(x)$ by $\varphi(x) - \varphi(0)$ where $\varphi(0)$ compensates for the divergence at $x = 0$, without affecting its behaviour at infinity.

In short, the proposed method allows for the definition of a functional, $f' = \alpha x_+^{\alpha-1}$, to be associated with the diverging integral yielded by the ordinary derivative of the function $f(x)$. So, in a way, it extends the admissible range for α in equation (5.19) to $-2 < \alpha < -1$. This extension can readily be pushed even further leading to,

$$\langle x_+^\alpha, \varphi \rangle \triangleq \int_0^{\infty} x^\alpha [\varphi(x) - \varphi(0) - \cdots - \frac{x^{n-1}}{\Gamma(n-1)} \varphi^{(n-1)}(0)] \mathrm{d}x, \tag{5.22}$$

where the allowable range for the α is now $-n-1 < \alpha < -n$.

Let me conclude this section by defining the important normalized generalized functions f_+^α as

$$f_+^\alpha = \begin{cases} \dfrac{1}{\Gamma(\alpha+1)} x_+^\alpha & \alpha \text{ non-integer} \\ \delta^{(l)} & \alpha = l,\ l \text{ integer} \end{cases} \tag{5.23}$$

with

$$x_+^\alpha = \begin{cases} 0 & x \leq 0 \\ x^\alpha & x > 0 \end{cases} \tag{5.24}$$

and where the $\delta^{(l)}$ refers to the l^{th}-order derivative of the δ-distribution taken in the sense of distributions. Finally note that the Pf has been omitted for convenience.

5.5 Definition of some elementary operations on distributions

Up-to this point two fundamentally different convergence mechanisms have been used. First of all it was shown that it is possible to construct a sequence of test functions *uniformly* converging to a test function that approximates a continuous compact support function, an ordinary function, up to arbitrary accuracy. On the other hand it was also shown that it is possible to construct a sequence of test functions point wise converging to a *singular* distribution, a *generalized* function, f. This latter convergence refers to the notion that the distribution f is being constructed as a limiting procedure on a sequence $\{\varphi_n\}_{n=1}^{\infty}$ yielding the sequence[7] $\{\langle\varphi_n,\phi\rangle\}_{n=1}^{\infty}$ converging to $\langle f,\phi\rangle$, with both $\varphi,\phi\in\mathcal{D}$.

The afore mentioned point wise convergence to a singular distribution has the advantage that it allows for an elegant definition of certain elementary operations on these singular distributions. Such an approach necessitates the introduction of the *adjoint identity* relating two operators $\mathcal{T}$ and $\mathcal{U}$ in the following way

$$\int_{-\infty}^{+\infty}\mathcal{T}\{\psi\}(x)\varphi(x)\mathrm{d}x=\int_{-\infty}^{+\infty}\psi(x)\mathcal{U}\{\varphi\}(x)\mathrm{d}x, \tag{5.25}$$

where $\varphi,\phi\in\mathcal{D}$ and where $\mathcal{T}$ symbolizes the operator of interest whereas its *adjoint*, $\mathcal{U}$, is an operator equalizing the left and right hand sides of equation (5.25).

In this way one has, via the proxy of the adjoint operator, circumvented the definition of the operations directly on the distribution f itself. But instead it is implicitly defined by means of the adjoint operator $\mathcal{U}$, acting on the vector space of test functions $\varphi\in\mathcal{D}$. Of course, in the formulation it is tacitly assumed that the operator $\mathcal{U}$ exists and has a meaning. So the whole procedure corresponds to finding the adjoint operator satisfying the identity $\langle\mathcal{T}\{f\},\varphi\rangle=\langle f,\mathcal{U}\{\varphi\}\rangle$, where the functional f has to be interpreted as the result of a limiting sequence converging to f, i.e.

$$\lim_{n\to\infty}\langle\mathcal{T}\{\varphi_n\},\varphi\rangle=\lim_{n\to\infty}\langle\varphi_n,\mathcal{U}\{\varphi\}\rangle, \tag{5.26}$$

or

$$\langle\mathcal{T}\{f\},\varphi\rangle=\langle f,\mathcal{U}\{\varphi\}\rangle. \tag{5.27}$$

A careful examination of the definition of the operator $\mathcal{T}$ via the proxy $\mathcal{U}$ reveals the important implication that the adjoint operator's action must yield a solution being an element of $\mathcal{D}$ too, hence $\mathcal{U}\{\varphi\}\in\mathcal{D}$! Otherwise, it will not be possible to define the operator $\mathcal{T}$ consistently and meaningfully. Despite this constraint, restricting the scope for the permissible operations, it is still possible to define certain operations although no standard recipe exists to find the adjoint operator $\mathcal{U}$ in terms of $\mathcal{T}$. If it is possible to

[7] Note that a direct relation exists between this sequence and the delta convergent sequence discussed in section 5.4.

find $\mathcal{U}$, then the operation $\mathcal{T}\{f\}$ is defined by equation (5.27).
As a first example take $\mathcal{T}$ to be the translation operator, i.e. $\mathcal{T}_a\psi(x) = \psi(x-a)$[8]. Substitution into equation (5.25) yields

$$\begin{aligned}\int_{-\infty}^{+\infty} \mathcal{T}_a\{\psi\}(x)\varphi(x)\mathrm{d}x &= \int_{-\infty}^{+\infty} \psi(x-a)\varphi(x)\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} \psi(x)\varphi(x+a)\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} \psi(x)\mathcal{T}_{-a}\{\varphi(x)\}\mathrm{d}x. \end{aligned} \tag{5.28}$$

So $\mathcal{U}_a = \mathcal{T}_{-a}$ and from this

$$\langle f(x-a), \varphi(x)\rangle = \langle f(x), \varphi(x+a)\rangle \tag{5.29}$$

is obtained. As a second example take $\mathcal{T}$ to be the derivative operator $\mathrm{d}/\mathrm{d}x$, $\mathcal{T}\{\psi\}(x) = \mathrm{d}\psi(x)/\mathrm{d}x$. Again substitution in equation (5.25) gives

$$\int_{-\infty}^{+\infty} \left(\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\right)\varphi(x)\mathrm{d}x = -\int_{-\infty}^{+\infty} \psi(x)\frac{\mathrm{d}}{\mathrm{d}x}\varphi(x)\mathrm{d}x, \tag{5.30}$$

yielding, $\mathcal{U} = -\mathcal{T}$. This last equation is easily checked by partial integration and taking into account that the stock term vanishes since both φ and ψ are of compact support and infinitely smooth, also at the endpoints of the support, hence $\psi, \varphi \in \mathcal{D}$. So the derivative of a distribution is given by

$$\langle \frac{\mathrm{d}}{\mathrm{d}x} f, \varphi\rangle = -\langle f, \frac{\mathrm{d}}{\mathrm{d}x}\varphi\rangle. \tag{5.31}$$

Now by taking again for f the δ-distribution one finds

$$\langle \frac{\mathrm{d}}{\mathrm{d}x}\delta, \varphi\rangle = -\langle \delta, \frac{\mathrm{d}}{\mathrm{d}x}\varphi\rangle = -\varphi^{(1)}(0), \tag{5.32}$$

where $^{(1)}$ stands for the first derivative and where $\delta, \delta^{(1)} \in \mathcal{D}'$. This operation can readily be extended to the k^{th} derivative of a distribution and equation (5.31) generalizes in that case to

$$\langle \frac{\mathrm{d}^k}{\mathrm{d}x^k} f, \varphi\rangle = (-1)^k\langle f, \frac{\mathrm{d}^k}{\mathrm{d}x^k}\varphi\rangle. \tag{5.33}$$

As a next example consider the multiplication of a distribution by a function $q(x)$. From equation (5.25) it is obvious that $\mathcal{T} = \mathcal{U} = q(x)$. However, care must be taken that

[8] I use here $\mathcal{T}$ for the translation operator rather than $\mathcal{S}$ as defined in chapter 3 to avoid confusion with the Schwartz class $\mathcal{S}$.

$q(x)\varphi(x) \in \mathcal{D}$ whenever $\varphi \in \mathcal{D}$. This means that $q(x)$ should be infinitely differentiable[9]. If $q(x)$ satisfies this condition then

$$\langle qf, \varphi \rangle = \langle f, q\varphi \rangle. \tag{5.34}$$

Note that if q would be discontinuous at $x = 0$ and if f is the δ-distribution, then the product $q\delta$ is **not** defined. This is because $\langle q\delta, \varphi \rangle = \langle \delta, q\varphi \rangle$ is not defined since $q\varphi \notin \mathcal{D}$. Take $q \in C^\infty$ and $f = \delta^{(1)}$, then

$$\begin{aligned} \langle q\delta^{(1)}, \varphi \rangle &= \langle \delta^{(1)}, q\varphi \rangle \\ &= -\langle \delta, (q\varphi)^{(1)} \rangle \\ &= -\langle \delta, q^{(1)}\varphi + q\varphi^{(1)} \rangle \\ &= -q^{(1)}(0)\varphi(0) - q(0)\varphi^{(1)}(0) \\ &= \langle -q^{(1)}(0)\delta + q(0)\delta^{(1)}, \varphi \rangle. \end{aligned} \tag{5.35}$$

So $q\delta^{(1)} = -q^{(1)}(0)\delta + q(0)\delta^{(1)}$ and not $q^{(1)}(0)\delta$ as one would maybe expect at first glance. As a last example of this section, before going into Fourier transforms, the convolution of two distributions is considered. It must be noted that the convolution of two distributions of $\mathcal{D}'$, is only possible if at least one of the distributions is of bounded support, or if both distributions have supports bounded on the left, or if both distributions have supports bounded on the right.

For two continuous and integrable functions f and g the convolution produces a new function $h(x)$, which is defined by

$$h(x) = (f * g)(x) = \int_{-\infty}^{+\infty} f(x')g(x - x')\mathrm{d}x'. \tag{5.36}$$

Now, letting $\varphi \in \mathcal{D}$ and f and g still continuous functions, consider

$$\langle h, \varphi \rangle = \langle f * g, \varphi \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x')g(x - x')\mathrm{d}x'\varphi(x)\mathrm{d}x. \tag{5.37}$$

Change of variables yields

$$\begin{aligned} \langle f * g, \varphi \rangle &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x)g(x')\varphi(x + x')\mathrm{d}x'\mathrm{d}x \\ &= \langle f(x), \langle g(x'), \varphi(x + x') \rangle \rangle. \end{aligned} \tag{5.38}$$

Under the restrictions for the distributions made above, equation (5.38) can be used to define the convolution of two distributions. Taking $g = \delta$ gives

$$\langle f * \delta, \varphi \rangle = \langle f(x), \langle \delta(x'), \varphi(x + x') \rangle \rangle = \langle f(x), \varphi(x) \rangle. \tag{5.39}$$

[9]This is the case when a test function $\varphi \in \mathcal{D}$ is used and can, to certain extent, be relaxed by choosing a vector space of test functions being at least one time differentiable. That will imply that q has to be one time differentiable also.

So the convolution of a distribution with the δ-distribution gives the distribution again. Taking $g = \delta^{(k)}$, so g is the k^{th} derivative of the δ-distribution, gives a very interesting result,

$$\begin{aligned} \langle f * \delta^{(k)}, \varphi \rangle &= \langle f(x), \langle \delta^{(k)}(x'), \varphi(x+x') \rangle \rangle \\ &= \langle f(x), \langle \delta(x'), (-1)^k \varphi^{(k)}(x+x') \rangle \rangle \\ &= \langle f, (-1)^k \varphi^{(k)} \rangle \\ &= \langle f^{(k)}, \varphi \rangle. \end{aligned} \tag{5.40}$$

So taking the k^{th} derivative of a distribution is the same as convolving the distribution with the k^{th} derivative of the δ-distribution.

5.6 The Schwartz class $\mathcal{S}$

It is not possible to define the Fourier transform for all distributions of $\mathcal{D}'$, but it is possible to do this for a subset of the distributions of $\mathcal{D}'$. This subset is called the space of tempered distributions or distributions of slow growth and is denoted by $\mathcal{S}'$. Associated with $\mathcal{S}'$ is a class of test functions which are called test functions of rapid descent. This space of test functions is denoted by $\mathcal{S}$ and is called the Schwartz class. The test functions of $\mathcal{S}$ are called of rapid descent because they must satisfy

$$|\varphi(x)| \leq M_n |x|^{-n} \quad \text{as} \quad x \to \infty, \tag{5.41}$$

with M_n a constant which may depend on n and $n = 1, 2, 3, \ldots$. So the test functions of $\mathcal{S}$ do not necessarily have to be of compact support. A test function φ is said to belong to $\mathcal{S}$ if $\varphi \in C^\infty$ and if it and all its derivatives are of rapid descent. The test functions of $\mathcal{D}$ clearly belong to $\mathcal{S}$, so $\mathcal{D} \subset \mathcal{S}$. On the other hand, the distributions which are in $\mathcal{S}'$ are certainly in $\mathcal{D}'$ and so $\mathcal{S}' \subset \mathcal{D}'$. Therefore the operations on distributions of $\mathcal{D}'$, which were defined in the previous section, can also be defined for the distributions of $\mathcal{S}'$.

An example of a test function of the Schwartz class is the Gaussian

$$\varphi(x) = \frac{1}{\sqrt{\pi}} e^{-x^2} = f_4(x, 1). \tag{5.42}$$

The δ-distribution is an example of a distribution which belongs to $\mathcal{S}'$. And e^{x^2} is an example of a distribution belonging to $\mathcal{D}'$ but not to $\mathcal{S}'$. The reason why $\mathcal{S}$ is useful in studying the Fourier transform, is that the Fourier transform maps $\mathcal{S}$ onto itself, i.e. $\hat{\varphi} \in \mathcal{S}$[10] whenever $\varphi \in \mathcal{S}$ and vice versa. This will not be proved, and for proofs the reader is referred to Zemanian (1965) and Schwartz (1959).

[10] In this chapter I denote the Fourier transform by a hat above the variable. Furthermore, for the frequency I will use ω since the independent variables do not correspond to physical quantities here, i.e. x does not necessarily mean a spatial coordinate

5.6.1 The Fourier transform of tempered distributions

As was already mentioned, if $\varphi \in \mathcal{S}$ then $\hat{\varphi} \in \mathcal{S}$ and vice versa. So using the adjoint identity of equation (5.25),

$$\int_{-\infty}^{+\infty} \mathcal{T}\{\psi\}(x)\varphi(x)\mathrm{d}x = \int_{-\infty}^{+\infty} \psi(x)\mathcal{U}\{\varphi\}(x)\mathrm{d}x, \tag{5.43}$$

and taking $\mathcal{T} = \mathcal{F}$,

$$\int_{-\infty}^{+\infty} \hat{\psi}(\omega)\varphi(\omega)\mathrm{d}\omega = \int_{-\infty}^{+\infty} \psi(x)\mathcal{U}\{\varphi\}(x)\mathrm{d}x, \tag{5.44}$$

is obtained. Substituting $\hat{\psi}(\omega) = \int_{-\infty}^{+\infty} \psi(x)e^{-i\omega x}\mathrm{d}x$ and interchanging the order of integration gives Parseval's equation

$$\begin{aligned} \int_{-\infty}^{+\infty} \hat{\psi}(\omega)\varphi(\omega)\mathrm{d}\omega &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \psi(x)e^{-i\omega x}\mathrm{d}x\varphi(\omega)\mathrm{d}\omega \\ &= \int_{-\infty}^{+\infty} \psi(x)\int_{-\infty}^{+\infty} \varphi(\omega)e^{-i\omega x}\mathrm{d}\omega\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} \psi(x)\hat{\varphi}(x)\mathrm{d}x. \end{aligned} \tag{5.45}$$

This shows that $\mathcal{U} = \mathcal{F}$. It is clear that whenever $\varphi \in \mathcal{S}$, then $\mathcal{U}\{\varphi\} = \hat{\varphi} \in \mathcal{S}$. So this allows to define the Fourier transform of any distribution of slow growth as

$$\langle \hat{f}, \varphi \rangle = \langle f, \hat{\varphi} \rangle. \tag{5.46}$$

Using the same arguments it is also possible to define the inverse Fourier transform of distributions $f \in \mathcal{S}'$ as

$$\langle \mathcal{F}^{-1}\{f\}, \varphi \rangle = \langle f, \mathcal{F}^{-1}\{\varphi\} \rangle, \tag{5.47}$$

which makes again sense because if $\varphi \in \mathcal{S}$, then $\mathcal{F}^{-1}\{\varphi\} \in \mathcal{S}$.

A remark must be made why it is not possible to define the Fourier transform for all distributions of $\mathcal{D}'$. This is because the Fourier transform of the test functions φ of $\mathcal{D}$ are not in $\mathcal{D}$, i.e. the Fourier transform of a infinitely smooth function of compact support is not an infinitely smooth function of compact support. The only test function of $\mathcal{D}$ that has a Fourier transform in $\mathcal{D}$ is the test function $\varphi = 0$. Since $\mathcal{F}\{\varphi\} \notin \mathcal{D}$ and $\mathcal{F}^{-1}\{\varphi\} \notin \mathcal{D}$ when $\varphi \in \mathcal{D}$, it is not possible to give a sense to the right hand sides of equations (5.46) and (5.47) for distributions of $\mathcal{D}'$ (Zemanian, 1965).

This section is concluded with a few examples. Consider the Fourier transform of the

k^{th} derivative of the shifted δ-distribution $\mathcal{F}\{\delta^{(k)}(x-a)\}$. This yields

$$\begin{aligned}\langle \mathcal{F}\{\delta^{(k)}(x-a)\}(\omega), \varphi(\omega)\rangle &= \langle \delta^{(k)}(x-a), \hat{\varphi}(x)\rangle \\ &= \langle \delta(x-a), (-1)^k \hat{\varphi}^{(k)}(x)\rangle \\ &= \langle \delta(x), (-1)^k \hat{\varphi}^{(k)}(x+a)\rangle \\ &= (-1)^k \hat{\varphi}^{(k)}(a), \end{aligned} \tag{5.48}$$

where $\hat{\varphi}^{(k)}$ means the k^{th} derivative of $\hat{\varphi}$ and not the Fourier transform of the k^{th} derivative of φ, so $\hat{\varphi}^{(k)} = (\mathcal{F}\{\varphi\})^{(k)}$ and not $\mathcal{F}\{\varphi^{(k)}\}$. The last term of equation (5.48) becomes

$$\begin{aligned}(-1)^k \hat{\varphi}^{(k)}(a) &= (-1)^k \left[\frac{\mathrm{d}^k}{\mathrm{d}x^k} \int_{-\infty}^{+\infty} \varphi(\omega) e^{-i\omega x} \mathrm{d}\omega\right]_{x=a} \\ &= \int_{-\infty}^{+\infty} (i\omega)^k \varphi(\omega) e^{-i\omega a} \mathrm{d}\omega \\ &= \langle (i\omega)^k e^{-i\omega a}, \varphi(\omega)\rangle. \end{aligned} \tag{5.49}$$

This shows that

$$\mathcal{F}\{\delta^{(k)}(x-a)\} = (i\omega)^k e^{-i\omega a}, \tag{5.50}$$

which has the following special cases

$$\mathcal{F}\{\delta\} = 1, \tag{5.51}$$

$$\mathcal{F}\{\delta(x-a)\} = e^{-i\omega a}, \tag{5.52}$$

$$\mathcal{F}\{\delta^{(k)}\} = (i\omega)^k. \tag{5.53}$$

In a similar way it can be shown that

$$\mathcal{F}\{1\} = 2\pi\delta(\omega), \tag{5.54}$$

$$\mathcal{F}\{e^{-iax}\} = 2\pi\delta(\omega+a), \tag{5.55}$$

$$\mathcal{F}\{(ix)^k\} = (-1)^k 2\pi\delta^{(k)}(\omega), \tag{5.56}$$

$$\mathcal{F}\{(ix)^k e^{-iax}\} = (-1)^k 2\pi\delta^{(k)}(\omega+a). \tag{5.57}$$

It is a well known fact that the convolution of two functions is a multiplication in the Fourier domain. This is a property which also holds for two distributions if at least one of them is of compact support. So if at least one of the distributions f, g is of compact support, then

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\}\mathcal{F}\{g\}. \tag{5.58}$$

5.7 Regularization

5.7.1 Regularization by convolution

In the example of section 5.2, the temperature in a room was measured with a thermometer. This was done by placing the bulb of the thermometer at the "point" x. Suppose now that the temperature depending on the space variable x is of interest and not just the temperature at one point. Obviously, the way to follow is to measure the temperature at every point x. This gives the *measured* temperature as a *function* of x. By placing the bulb of the thermometer at every point x, what is actually done is a convolution of the temperature distribution f with the test function φ. So the measured temperature in the room is given by

$$h(x) = \langle f(x'), \varphi(x - x') \rangle = \int_{-\infty}^{\infty} f(x')\varphi(x - x')\mathrm{d}x'. \tag{5.59}$$

Here h represents the measured quantity as function of x, f is the unknown process and φ is the measuring device, which is shifted along the spatial coordinate. It is possible to prove that the function $h(x)$ of equation (5.59) is an infinitely smooth function, so it has infinitely many derivatives. This results in the next theorem which is taken from Zemanian (1965):

Theorem 5.1: Regularization with a test function (Zemanian, 1965)
Let f be in $\mathcal{D}'$ and let φ be in $\mathcal{D}$. Then $h = f * \varphi$ is an ordinary function which is given by

$$h(x) = \langle f(x'), \varphi(x - x') \rangle. \tag{5.60}$$

Moreover, $h(x)$ is infinitely smooth and

$$\frac{\mathrm{d}^k}{\mathrm{d}x^k} h(x) = \langle f(x'), \frac{\partial^k}{\partial x^k} \varphi(x - x') \rangle. \tag{5.61}$$

A similar theorem can be given for $f \in \mathcal{S}'$ and $\varphi \in \mathcal{S}$. Theorem 5.1 shows that a convolution is a smoothing process and so describing measuring in terms of distributions and test functions shows that a measurement is a smoothed representation of the quantity under investigation. Regularizing the δ-distribution yields

$$h_1(x) = \langle \delta(x'), \varphi(x - x') \rangle = \varphi(x), \tag{5.62}$$

where the property $\langle \delta(x - a), \varphi(x) \rangle = \langle \delta(x), \varphi(x + a) \rangle = \varphi(a)$ has been used. Here $\varphi(x)$ can be the Gaussian or the test function defined by equation (5.3) or any other test function. These functions can be regarded as a smoothed version or an approximation of the δ-distribution. In fact, they provide a measurement of the δ-distribution by giving the distribution of support zero a thickness. Note that the δ-distribution acts as a

unity operator for convolution. Analogously, the derivative of the δ-distribution can be regularized, giving

$$\begin{aligned} h_2(x) &= \langle \delta^{(1)}(x'), \varphi(x - x') \rangle \\ &= \langle \delta(x'), -\frac{\partial}{\partial x'} \varphi(x - x') \rangle \\ &= \langle \delta(x'), \frac{\partial}{\partial x} \varphi(x - x') \rangle = \varphi^{(1)}(x). \end{aligned} \tag{5.63}$$

So $h_2 = h_1^{(1)}$, which was already expected, since equation (5.40) showed that taking the derivative is the same as a convolution with the derivative of the δ-distribution. Equation (5.63) shows that a smoothed version of the derivative of the δ-distribution is given by the derivative of the Gaussian. So, in a sense, the derivative of the Gaussian can be seen as an approximation of the derivative of the δ-distribution. But before going further into this matter, the meaning of regularization by convolution is viewed in the Fourier domain.

5.7.2 Smoothness and rapid descent

Consider an integrable function $f(x)$ which has also an integrable derivative. This function can be written as

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{i\omega x} \mathrm{d}\omega. \tag{5.64}$$

The requirements on $f(x)$ demand that $f(x)$ is bounded. This means that

$$|f(x)| \leq M < \infty. \tag{5.65}$$

When substituting this in equation (5.64), it is seen that equation (5.65) is certainly satisfied if

$$\int_{-\infty}^{+\infty} |\hat{f}(\omega)| \mathrm{d}\omega < \infty. \tag{5.66}$$

This last equation will only hold if $\hat{f}(\omega)$ goes fast enough to zero when $\omega \to \pm\infty$. Similarly, if $f(x)$ has N continuous derivatives, which are all integrable, then

$$\int_{-\infty}^{+\infty} |\omega^N \hat{f}(\omega)| \mathrm{d}\omega < \infty, \tag{5.67}$$

which only holds when $\omega^N \hat{f}(\omega)$ goes fast enough to zero for $\omega \to \pm\infty$. This shows that there is a connection between the differentiability of a function and the decay of its Fourier transform. And vice versa, there is a connection between the decay of a function and the smoothness of its Fourier transform. As is seen above, the more continuous

derivatives $f(x)$ has, the faster the decay of its Fourier transform. More precisely, if f has derivatives up to N that are all integrable, then $\hat{f}$ decreases at infinity as ω^{-N}. And of course, if $\hat{f}$ has derivatives up to N that are all integrable, then f decreases at infinity as x^{-N}, expressing the intrinsic duality of the Fourier transform. Note, however, that the reverse of these arguments need not to be true. In fact, it can be proved that if f ($\hat{f}$) decreases at infinity like x^{-N} (ω^{-N}) then $\hat{f}$ (f) has continuous and bounded derivatives of all orders up to $N-1$ (Strichartz, 1994).

Because the Fourier transform of distributions has been defined, it is also possible that the Fourier transform does not decrease at infinity, but increases. An example is the N^{th} derivative of the δ-distribution, which equals $(j\omega)^N$ (see equation (5.53)). This can be interpreted that the N^{th} derivative of the δ-distribution has no smoothness at all. So when the Fourier transform of a function increases for $\omega \to \pm\infty$, it is not smooth. Knowing that smoothness in the place domain corresponds to decay in the Fourier domain, consider again the process of convolution. The test functions of $\mathcal{D}$ and $\mathcal{S}$ are infinitely smooth, so their Fourier transforms decay faster than any inverse polynomial. This means that multiplying $\hat{\varphi}$, when $\varphi \in \mathcal{S}$ or $\varphi \in \mathcal{D}$, with the Fourier transform $\hat{f}$ of any distribution of $\mathcal{S}'$, gives a function in the Fourier domain which still decays faster than any inverse polynomial. This means that the inverse Fourier transform of $\hat{f}\hat{\varphi}$ is infinitely smooth, or $f * \varphi$ is infinitely smooth, what was already stated in theorem 5.1. The convolution of a distribution with a test function can therefore be viewed as making a band-limited approximation of the distribution. Of course, the higher the cut-off frequency of $\hat{\varphi}(\omega)$, i.e. the frequency where $\hat{\varphi}(\omega)$ is going to decay fast, the better the approximation becomes. The higher the cut-off frequency of $\hat{\varphi}(\omega)$ means that $\varphi(x)$ becomes more concentrated.

Take for example the Gaussian function $f_4(x,\sigma)$ of equation (5.14). If $\sigma \to 0$ then $f_4(x,\sigma)$ approximates the δ-distribution. The function $f_4(x,\sigma)$ becomes concentrated around 0 when $\sigma \to 0$, but its Fourier transform is spread out over the frequency axis, and it approximates a flat spectrum. How well does the Gaussian approximate the δ-distribution? Well, it depends on the scale where it is examined. In the next chapter it will be shown that a Gaussian of a certain scale σ_0 behaves as an infinitely smooth function if it is examined for scales smaller than σ_0, but that it behaves as a δ-distribution for scales larger than σ_0. So for large scales this Gaussian behaves as a singular distribution.

5.8 Fractional integration and differentiation

Before shifting attention to providing definitions allowing for a categorization of the degree of regularity of a functional it is beneficial to quickly review the operation of *fractional* integration and differentiation. This generalized operation coined by Riesz and others (Mandelbrot, 1982; le Méhauté, 1991; Duistermaat, 1993) will come at hand since it naturally emerges in fractal theory. For example it is used in chapter 7 to define

fractional Brownian motion[11] or fractional Levy motion and in chapter 7 where the fractional integration is used to smooth the universal multifractal density.

Consider, following Duistermaat (1993), the space of distributions bounded from the left, i.e. the space $\mathcal{D}'(\mathbb{R})_+$. This space consists of those distributions the support of which is given by supp $[x_0, \infty)$, with $x \in \mathbb{R}$ and $x \neq -\infty$. Given this condition the convolution for $u, v \in \mathcal{D}'(\mathbb{R})_+$ can be given a meaning yielding $(u * v)(x) \in \mathcal{D}'(\mathbb{R})_+$ (Zemanian, 1965). Define

$$\partial^\alpha f = \mathcal{D}^\alpha\{f\}(x) = \mathcal{I}^{-\alpha}\{f\}(x) \triangleq \frac{1}{\Gamma(-\alpha)}(f * x_+^{-\alpha-1}). \tag{5.68}$$

For $\alpha > 0 \ \wedge \ \alpha \neq \mathbb{Z}^+$ $\mathcal{D}^\alpha\{\cdot\}$ denotes the α^{th} order fractional derivative, whereas for $\alpha < 0 \ \wedge \ \alpha \neq \mathbb{Z}^-$ $\mathcal{I}^{-\alpha}\{\cdot\}$ denotes the α^{th} order fractional integration. For $\alpha = 0$, $\mathcal{D}^\alpha\{\cdot\} = \mathcal{I}^{-\alpha}\{\cdot\} = Id\{\cdot\}$ with $Id\{\cdot\}$ the identity operator. Notice that these fractional operators $\mathcal{D}^\alpha\{\cdot\}$ and $\mathcal{I}^{-\alpha}\{\cdot\}$ become ordinary N^{th} order derivatives or primitives when $N = \alpha \in \mathbb{Z}^+$ or when $N = \alpha \in \mathbb{Z}^-$ respectively. Also note that the convolution kernel in the definition of these operations is strongly related to the singular distributions with algebraic singularities defined in section 5.4.1. In fact the type of generalized functions generated by these algebraic singularities are known as *homogeneous functions* of degree α and they adhere to the dilatation equation

$$f(\sigma x) = \sigma^\alpha f(x) \tag{5.69}$$

or equivalently

$$\langle f(x), \varphi(\frac{x}{\sigma})\rangle = \sigma^{\alpha+1}\langle f(x), \varphi(x)\rangle. \tag{5.70}$$

The order α homogeneous functions, also known as the order α tempered distributions, constitute a *group* since they support the *identity* operator, the inverse operator and a "chain" rule, i.e. $x_+^\alpha * x_+^\beta = x_+^{\alpha+\beta}$ and where

$$\frac{\mathrm{d}}{\mathrm{d}x} x_+^{\alpha+1} \sim x_+^\alpha. \tag{5.71}$$

So for example,

$$x_+^{-1} = \frac{\mathrm{d}}{\mathrm{d}x} x_+^0 = \frac{\mathrm{d}}{\mathrm{d}x}\theta = \delta, \tag{5.72}$$

where θ is the Heaviside function being a homogeneous function of order 0 and where the δ-distribution is a homogeneous function of order -1.

[11] Note that in this definition the fractional integration/differentiation is done in the sense of distributions, via the first increment or a test function of mean zero, because the measure $\mathrm{d}B(z)$ is non-integrable.

5.9 Categorizing singularities

Generalized functions, functions containing a singularity, can be categorized by their (ir)regularity. The degree of regularity is expressed by the Lipschitz or Hölder exponent[12] which characterizes either the *local* regularity, the Hölder exponent at a point, $x = x_0$, or the *global* regularity, uniform regularity, denoted by the dominating, the smallest, Hölder exponent to be associated with the closed interval $x \in (a, b)$. The discussion will start by considering functionals which are positive Hölder, $\alpha > 0$, followed by an extension to generalized functions, the tempered distributions, to which negative, $\alpha < 0$, Hölder exponents can be associated.

Now what does it mean when a function is Hölder α with $\alpha > 0$? It means that when $n < \alpha < n + 1$ the function is n times differentiable but *not* $(n + 1)$ times differentiable, i.e. the $(n + 1)^{\text{th}}$ does not exist because it is not bounded. In that case the function is *singular* in its n^{th} derivative and the derivative can be only be taken in the sense of distributions. Now let me reiterate the formal definition for Hölder regularity.

Definition 5.2: Regularity (Mallat and Hwang, 1992)

- *Let n be a positive integer such that $n \leq \alpha \leq n + 1$. A function $f(x)$ has a local Hölder exponent α at x_0 if and only if there exist two constants C and $h_0 > 0$ and a polynomial of order n, $P_n(x)$, such that for $h < h_0$*

$$|f(x_0 + h) - P_n(h)| \leq C|h|^{\alpha}. \tag{5.73}$$

- *The function $f(x)$ has a global Hölder exponent α on the interval (a, b) if and only if there is a constant C and there is a polynomial of order n, $P_n(h)$, such that equation (5.73) is satisfied if $x_0 + h \in (a, b)$.*

- *The Hölder regularity of $f(x)$ at x_0 is the superior bound of all values α such that $f(x)$ is Hölder α at x_0.*

- *The function $f(x)$ is singular at x_0 if it has a Hölder exponent $\alpha < 1$ at x_0.*

From the last item of this definition it is clear that the n^{th}-order derivative of a functional being Hölder α, with $n < \alpha < n + 1$, is singular because the $(n + 1)^{\text{th}}$ derivative diverges but when this derivative is bounded then is the n^{th} derivative of f is regular.

Integrating the function $f(x)$ increases the Hölder exponent by 1, i.e.

$$\partial^{-1} f(x): \quad \alpha \mapsto \alpha + 1 \quad \text{at} \quad x = x_0 \tag{5.74}$$

[12] In literature either one of these two names are used but it may be fair to say that the Lipschitz exponent refers to integer values whereas the Hölder exponent refers to non-integer values. In this thesis I will use the Hölder exponent.

where ∂^{-1} denotes the anti-derivative. However, the opposite, differentiating f, does not necessarily lead to a decrease by one of the *local* Hölder exponent, see Mallat and Hwang (1992), where the authors attribute this observation to oscillatory phenomena that can also yield the emergence of a singularity.

On the other hand, in case $\alpha > 1$ and non-integer, it is possible to appoint a global Hölder exponent α to the non-vanishing open interval $x \in (a, b)$, if and only if its derivative is uniformly Hölder $\alpha - 1$ on the same interval. As a consequence of the "if and only if" it is possible to reverse the argument and that results in the formulation of definition 5.3 where the range of Hölder exponents is extended to negative values.

Definition 5.3: Negative Hölder exponents(Mallat and Hwang, 1992)
Let $f(x)$ be a distribution of finite order, a tempered distribution see sections 5.6 and 5.8. Let α be a real non-integer number and $[a, b]$ a compact interval of $\mathbb{R}$. The distribution $f(x)$ is said to have a global *Hölder exponent α on (a, b), if and only if its primitive has a* global *Hölder exponent $\alpha + 1$ on (a, b).*

In this definition the term function has intendedly been replaced by that of a distribution because for $\alpha \leq 0$ the concept of a function is not longer valid because of the divergence at the singular point $x = x_0$. In fact the distribution refers to Hadamard's finite part as introduced in section 5.4.1. Unfortunately it is not possible to provide a similar definition for local negative Hölder exponents or to functionals of integer order. This latter category necessitates a more sophisticated theory as proposed by Bony (1983) and Jaffard (1991) who are able to make an extension to point wise negative Hölder exponents using microlocalization theory. This advanced theory is, however, beyond the scope of this thesis. Despite this deviancy it is possible to come up with an alternative definition for isolated singularities with negative exponents.

Definition 5.4: Isolated negative Hölder exponents
A distribution $f(x)$ is said to have an isolated singularity with Hölder exponent α at x_0, if and only if $f(x)$ has a global *Hölder exponent α over an interval (a, b), with $x_0 \in (a, b)$, and if $f(x)$ has a* global *Hölder exponent 1 over any open subinterval (a, b) that does not include x_0.*

Application of definitions 5.2 and 5.4 to generalized function x_+^{-1}, i.e. to the δ-distribution leads to the reasoning that the second primitive of x_+^{-1} is of global Hölder $\alpha = 1$ in the direct neighbourhood of $x = 0$ and has therefore a global Hölder exponent α with $\alpha < 1$ in the neighbourhood of $x = 0$. Due to the supp in the third item of definition 5.2 one can then, in the neighbourhood of $x = 0$, appoint a $\alpha = -1$ as a uniform Hölder exponent to the δ-distribution. In chapter 8 I will pay extensively attention to the actual measurement of Hölder exponents by means of the continuous wavelet transform. Finally

note that there exists a direct relationship between the order of a homogeneous function, see section 5.8, and the Hölder regularity.

5.10 Concluding remarks

In this chapter an attempt has been made to lay down the basic framework of *distributions*. This attempt is, however, unfortunately limited in scope. The most important action taken was the replacement of the conventional concept of a function by that of a *functional*. This latter concept is, by way of its construction, not only capable to relate mathematical operations to physical measurements but it also provides the formalism in which solutions of particular systems of (partial) differential equations can be given a meaning (Zemanian, 1965). The hyperbolic system of first order partial differential equations denoting the wave equation forms an example.

The established link between the mathematics and the physics can at best be summarized in the following way

$$\begin{aligned} \text{functional} &\leftrightarrow \text{physical variable} \\ \text{test function} &\leftrightarrow \text{measurement device} \\ \text{tested functional} &\leftrightarrow \text{outcome measurement.} \end{aligned}$$

where the inevitable physical bound on the obtainable spatial-dynamic scale range is incorporated by the concept of the *test function*.

But above all distribution theory postulates a methodology on how to deal with certain objects that do not have a meaning in the classical mathematical sense. The distributional interpretation roughly boils down to interpreting mathematical operations in the sense of sequences which embody successive approximations, without ever reaching the "exact" approximation, to the operation which originally had an exact meaning in the conventional mathematics. It is this way of "sequence thinking" that also reveals the intrinsic duality existing in the distribution theory. This duality refers to the notion that the test function themselves can, depending on how far they proceeded down the line of the pointwise converging sequence, also be interpreted as singular distributions. This far reaching observation is a direct consequence of the intrinsic ambiguity suffered by the observing test function, which, by way of its definition, must have of larger support, i.e. it must have proceeded less far in the sequence so that it "sees" the "singular" other test function, i.e. it sees it as being of measure zero. In this way the observing test function can not discern the other "test" function from say the singular δ-distribution. This latter observation can be attributed to the fact that for example a Gaussian shaped test function can be seen as an element of the Delta convergent sequence, it constitutes a coarse approximation to the δ-distribution.

Chapter 6

Selected topics on deterministic mono- and multifractals

6.1 Introduction

In order to understand the basic ideas behind the concept of mono- /multifractals I will start the discussion by reviewing the iterative construction and properties of two simple *deterministic* fractal sets. At first the triadic Cantor Set will be examined followed by a discussion on a binomial multifractal. The first set is an example of a *monofractal* – a homogeneous fractal characterized by one *single* scaling exponent (Mandelbrot, 1982) – whereas the latter constitutes a generalization to a *multifractal* set, being inhomogeneous and consisting of a hierarchy of scaling exponents, to be captured by the *singularity spectrum* (Mandelbrot, 1974; Siebesma, 1989; le Méhauté, 1991; Lichtenberg and Lieberman, 1992; Bacry et al., 1993; Holschneider, 1995).

6.2 A monofractal example, definition of the fractal dimension

The first example of generating a fractal set is depicted in figure 6.1 where I included the construction mechanism for the triadic Cantor Set (Mandelbrot, 1982; Siebesma, 1989). The iteration starts with an *initiator* – a closed interval $C_0 = [0, 1]$ in this case – which is cut[1] into three equal pieces and where the middle part is deleted. This first iteration C_1 – the *generator* – is then repeated on the two remaining pieces. While proceeding the iteration procedure each line segment will be divided by two as a consequence of the action of the generator and finally, after repeating the reductions *ad infinitum*, the triadic Cantor set is being defined as the limit $C = \lim_{k\to\infty} C_k$. Now clearly, by construction, each piece of the set C_k can be acquired from the set C_l after applying the proper enlargement, when $l > k$, or reduction, when $l < k$, to this set C_l. This apparent invariance to dilatations, the

[1] Note that this is a *singular* operation.

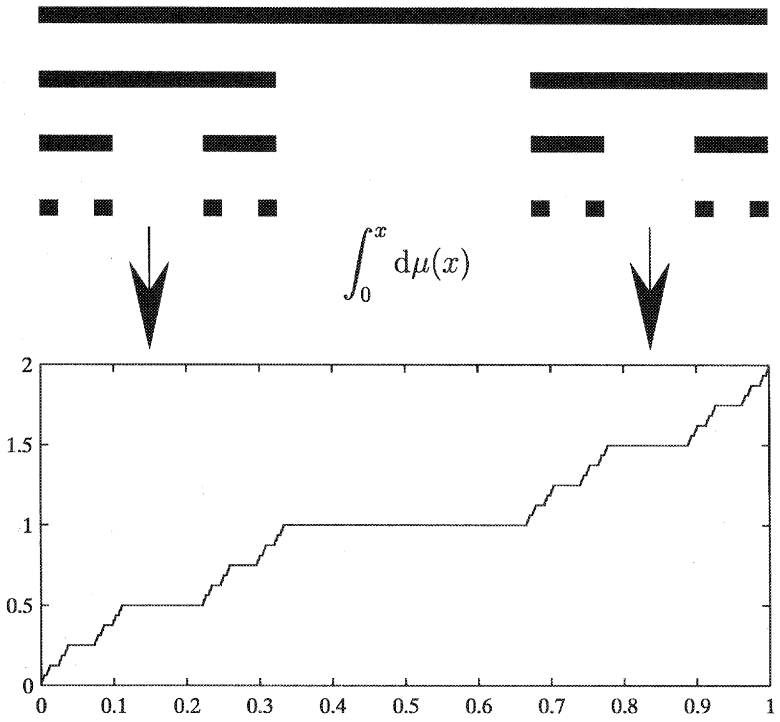

Figure 6.1 *Example of the iterative generation of the triadic Cantor Set. Below I depicted the "running" sum over the singular measure on which I supplemented a unit density, i.e.* $\int_0^x \mathrm{d}\mu(x)$ *where the support of the measure is given by the above Cantor set.*

invariance to scale transformations, is known as the *self-similarity* property expressing a *scale-invariance* which is in one way or another characteristic for all fractal structures.

As a result of the above construction mechanism, see figure 6.1, also known as a *cascade process*, one obtains an *irregular*, singular, set of points from the initial regular, smooth, one-dimensional unit length line segment. The number of points – the length of the segments vanishes in the limit – is infinite and they seem to span a structure that has a dimension somewhere in between that of a *line* and that of a *point*. This intuitive observation has been given a more precise meaning by the introduction of the *fractal*

dimension – a generalization of the metric dimension being used to characterize smooth structures – which can at best characterize complex sets such as the Cantor set.

Definition 6.1: Fractal box-dimension or capacity (Kolmogorov, 1941; Mandelbrot, 1974, 1982)
The box dimension of a set $A \subset \mathbb{R}^p$ is given by,

$$D_B \triangleq \lim_{\sigma \downarrow 0} \frac{\log N_\sigma}{\log \sigma^{-1}}, \tag{6.1}$$

where N_σ is the minimum number of p-dimensional σ-sized neighbourhoods needed to cover the set A. The parameter σ refers to the size of the boxes, the length of their "gauge".

Application of this definition to smooth objects such as lines in $\mathbb{R}$, surfaces in $\mathbb{R}^2$ and volumes in $\mathbb{R}^3$ leads to trivial values for the dimension $D_B = 1, 2, 3$ which exactly correspond to the Euclidean dimension because the number N_σ is inversely proportional to the σ to the power $1, 2, 3$, i.e. $N_\sigma \propto \sigma^{-1}$, $N_\sigma \propto \sigma^{-2}$, $N_\sigma \propto \sigma^{-3}$. On the contrary, application of this definition to the Cantor set reveals the interesting property that the dimension is of *non-integer* value. This observation follows from the fact that one needs $N_\sigma = 2^j$ line segments of size $\sigma = (\frac{1}{3})^j$ to fully cover the set, resulting in a box-dimension for the Cantor set that is given by,

$$D_B = \lim_{j \to \infty} \frac{\log 2^j}{\log 3^j} = \frac{\log 2}{\log 3} = 0.6309 \cdots .$$

In fact the box-dimension can be interpreted as an *exponent* expressing the amount of information required to specify the set up-to a scale σ, i.e. its *complexity*. The box-dimension is a simplified form of the Hausdorff dimension.

Definition 6.2: Hausdorff dimension and Hausdorff measure (Hausdorff, 1918; Mandelbrot, 1974, 1982)
Consider a covering of the subset $A \subset \mathbb{R}^p$ by p-dimensional neighbourhoods of linear size σ_i. The Hausdorff dimension $\dim_H$ is the critical dimension for which the Hausdorff measure $H_d(\sigma)$ takes a finite value,

$$H_d(\sigma) \triangleq \lim_{\sigma \downarrow 0} \inf \sum_i \sigma_i^d = \begin{cases} 0 & d > \dim_H, \\ \text{finite} & d = \dim_H, \\ \infty & d < \dim_H, \end{cases} \tag{6.2}$$

and where the infimum extends over all possible coverings subject to the constraint that $\sigma_i \leq \sigma$.

Following (le Méhauté, 1991) one can, from the practical point of view issue the statement: " ... It is of course the way in which this upper-bound (in equation (6.2)) approaches infinity that determines the fractal dimension, a property we shall meet when studying physical problems ... ", a statement with which I wholeheartedly agree.

So by this definition the integral

$$\int_A \mathrm{d}^{\dim_H(A)} = H_d(\sigma), \tag{6.3}$$

has been given a meaning by the Hausdorff dimension $\dim_H$, i.e. it represents a finite length.

6.3 A multifractal example, definition of the partition function, generalized dimensions and the singularity spectrum

At this point it is beneficial to generalize the class of *homogeneous* monofractals to the much richer class of *multifractals* which display a highly irregular, intermittent behaviour (Mandelbrot, 1974; Parisi and Frisch, 1985). Again start with the interval [0.1] but now with a *density* $\rho(x)$ superimposed on it, yielding a total mass $\int_0^1 \rho(x)\mathrm{d}x = \int_0^1 \mathrm{d}\mu(x) = 1$, where $\mu(x)$ is the *measure*. Let the generator consist of a division of the unit interval into two segments, followed by a redistribution of the mass – under the condition that the total mass is preserved $\int \mathrm{d}\mu(x) = 1$ – across the two reduced line segments, according to $p_1 = 0.25$ for the first and $p_2 = 0.75$ for the second and with the normalization, $p_1 + p_2 = 1$. On its turn this procedure is also repeated *ad infinitum* to finally yield a *binomial* multifractal measure[2]. By construction such a iterative process can be seen as a *multiplicative cascade* in which the weights at the $(k+1)^{\text{th}}$ iteration are the result of successive multiplications by the multiplicative increments (by either p_1 or p_2).

As for the initial definition of the box-dimension, describing monofractals such as the triadic Cantor set, the above multifractal can also be characterized but now by an infinite set of generalized dimensions.

Definition 6.3: Generalized or Renyi dimensions D_q (Renyi, 1970; Mandelbrot, 1974; Hentschel and Procaccia, 1983; Parisi and Frisch, 1985)
To define the generalized dimensions D_q partition the measure using a grid with a lattice constant σ and subsequently introduce the multiscale partition function as

$$Z_\sigma(q) \triangleq \sum_{\forall x_i \in A} \mu[B_\sigma(x_i)]^q = \sum_{\forall x_i \in A} {p_i}^q, \tag{6.4}$$

where $p_i = \mu[B_\sigma(x_i)] \triangleq \int_{B_\sigma(x_i)} \mathrm{d}\mu(x)$ is the total measure within the i^{th} box, at position

[2] These multifractals are examples of a so-called Baker's map (Bacry et al., 1993; Holschneider, 1995).

x_i *and of linear size* σ. *The set of generalized dimensions* D_q *is then defined by,*

$$(q-1)D_q \triangleq \tau(q) \triangleq \lim_{\sigma \downarrow 0} \frac{\log Z_\sigma(q)}{\log \sigma}, \tag{6.5}$$

implying a scaling behaviour for the partition function $Z_\sigma(q)$, *for small* σ,

$$Z_\sigma(q) \propto \sigma^{\tau(q)} \tag{6.6}$$

and where $\tau(q)$ *is the mass exponent function.*

For the binomial multifractal – $i = 1, 2$, see for its construction figure 6.2 – the *partition function* for the k^{th} iteration is simply given by

$$Z_{\sigma=2^{-k}}(q) = (p_1^q + p_2^q)^k = \chi(q)^k, \tag{6.7}$$

with $\chi(q)$ being the *generator*, the partition function after the first iteration. Application of definition 6.3, yields for the generalized dimensions D_q,

$$(q-1)D_q = -\log_2 \chi(q) = -\log_2(p_1^q + p_2^q), \tag{6.8}$$

which are – by nature of the self-similar construction – fully determined by the generator.

More generally – beyond the binomial case and optionally p-dimensional – a generator, composed of b^p boxes of linear size b^{-1} can be constructed where,

$$\chi(q) = \sum_{i=1}^{b^p} p_i^q, \tag{6.9}$$

and defines a generalized Cantor set embedded in p dimensions. Comparing the triadic Cantor set with the binomial multifractal, see figure 6.2, shows that the latter displays a highly *intermittent, non-stationary,* behaviour, a characteristic property of multifractal measures where the scaling properties vary from location to location. This heterogeneity makes it necessary to define a *pointwise* dimension or *singularity exponent* α.

Definition 6.4: Local scaling exponent of a measure (Collet, 1986; Siebesma, 1989; Bacry et al., 1993; Holschneider, 1995)
The local scaling exponent or local fractal dimension of a measure μ *is defined as*

$$\alpha(x_0) \triangleq \liminf_{\sigma \downarrow 0} \frac{\log \mu(B_\sigma(x_0))}{\log \sigma} \tag{6.10}$$

where $B_\sigma(x_0)$ *is a* σ*-box centered at* x_0.

Now the interesting question arises whether the generalized dimensions provide information on the distribution of the scaling exponents α.

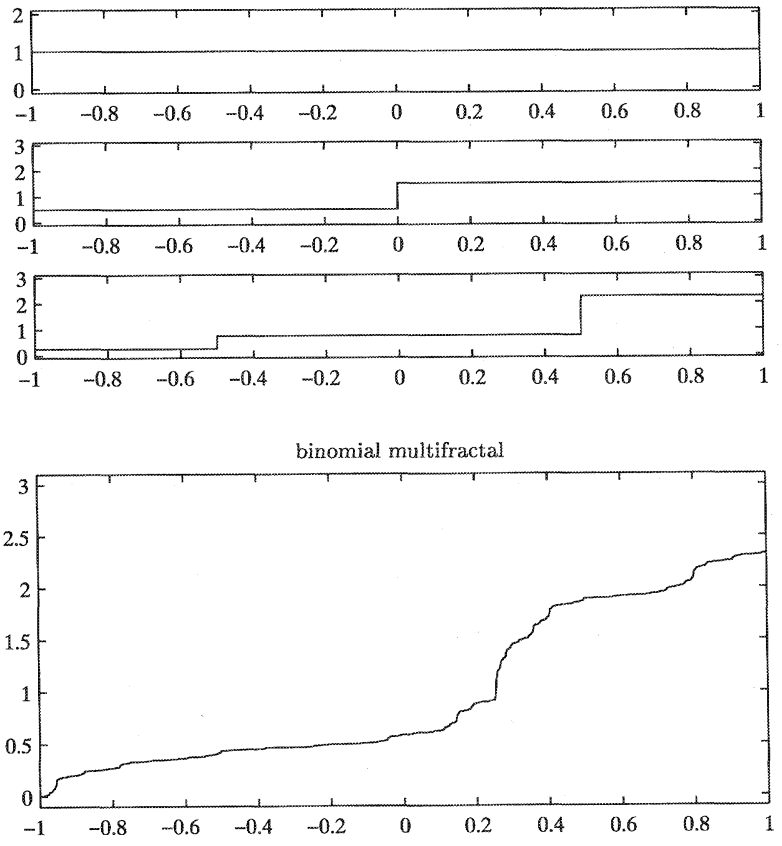

Figure 6.2 *Example of the iterative generation of a binomial multifractal with the measures set to* $p_1 = 0.25$ *and* $p_2 = 0.75$.

Theorem 6.1: Singularity spectrum and generalized dimensions (Collet, 1986; Holschneider, 1995)
The singularity spectrum and the generalized fractal dimensions are related via a Legendre transform

$$\frac{\mathrm{d}}{\mathrm{d}q}\tau(q) = \alpha, \qquad \tau(q) = \alpha q - \mathrm{f}(\alpha). \tag{6.11}$$

For the proof of this theorem I would like to refer to Collet et al. (87) but instead I will follow Parisi and Frisch (1985) who give a rather intuitive derivation which links the

singularity spectrum[3] – expressing the distribution of the singularities – to the generalized dimensions D_q or mass exponent function $\tau(q)$.

For this purpose regard the distribution of a measure A by boxes of size σ. Let $N_\sigma(\alpha)$ be the number of boxes the *local* scaling exponent of which corresponds to $\mu[B_\sigma(x)] \propto \sigma^\alpha$ and where the scaling exponent varies over the range α to $\alpha + d\alpha$. In this way Hausdorff dimensions $\mathrm{f}(\alpha)$ can be associated with the subsets $A_\alpha \subset A$ according to

$$\mathrm{f}(\alpha) \triangleq \lim_{\sigma \downarrow 0} \frac{-\log N_\sigma(\alpha)}{\log \sigma} \tag{6.12}$$

which is equivalent to the following scaling behaviour

$$N_\sigma(\alpha) \propto \sigma^{\mathrm{f}(\alpha)}\Big|_{\sigma \downarrow 0}. \tag{6.13}$$

Rewrite the partition function, see equation (6.4), in terms of the *local* scaling exponent α and the singularity spectrum $\mathrm{f}(\alpha)$ by summing over the subsets A_α rather than over A,

$$Z_\sigma(q) = \sum_\alpha \sum_{x \in A_\alpha} \mu[B_\sigma(x)]^q \propto \int \sigma^{q\alpha - \mathrm{f}(\alpha)} \mathrm{d}\alpha, \tag{6.14}$$

i.e. summing over all boxes with singularities within the interval $(\alpha, \alpha + d\alpha)$ for varying α. Using equation (6.5) and taking the limit for small σ one can approximate the integral of equation (6.14) by a *saddle point approximation*. In this approximation the integral is assumed to be dominated for small σ by the singularities α for which the exponent $q\alpha - \mathrm{f}(\alpha)$ is minimal, i.e.

$$Z_\sigma(q) \propto \sigma^{q\alpha - \mathrm{f}(\alpha)} = \sigma^{\tau(q)} \tag{6.15}$$

The minimal value of $q\alpha - \mathrm{f}(\alpha)$ is attained for those α for which $q - \partial_\alpha \mathrm{f}(\alpha) = 0$, consequently

$$\begin{cases} \tau(q) &= q\alpha - \mathrm{f}(\alpha) \\ q &= \partial_\alpha \mathrm{f}(\alpha) \end{cases} \tag{6.16}$$

and conversely,

$$\begin{cases} \mathrm{f}(\alpha) &= q\alpha - \tau(q) \\ \alpha &= \partial_q \tau(q). \end{cases} \tag{6.17}$$

These equations establish the relationship – via the Legendre transform – between the generalized or Renyi dimensions D_q to the singularity spectrum $\mathrm{f}(\alpha)$. The singularity spectrum can be formally defined by

[3] Also known as the scaling exponent function (Lichtenberg and Lieberman, 1992).

Definition 6.5: Singularity spectrum of a measure (Collet, 1986; Parisi and Frisch, 1985; Bacry et al., 1993; Holschneider, 1995)
The singularity spectrum of a measure $\mu(x)$ is the function $\mathrm{f}(\alpha)$ such that

$$\mathrm{f}(\alpha) \triangleq \dim_H\{x \mid \mu[B_\sigma(x)] \propto \sigma^\alpha|_{\sigma\downarrow 0}\}, \tag{6.18}$$

where $\dim_H$ denotes the Hausdorff dimension and $B_\sigma(x)$ is a σ-box centered at x.

The parameter q acts as a *selector* for the different singularities. For $q > 0$ it hunts for the *strong* singularities, i.e. the *high* density boxes, whereas for $q < 0$ the *weak* singularities, the boxes with *low* density, are emphasized. So in effect the set A, consisting of intertwined fractal subsets A_α, has been decomposed by this partition function.

For the rather simplistic example of the binomial multifractal it is straightforward to derive an analytical expression for the mass exponent $\tau(q)$, the generalized dimensions D_q and the singularity spectrum $\mathrm{f}(\alpha)$. The expression for the $\tau(q)$ reads (recall (6.8))

$$\tau(q) = -\log_2[p_1^q + p_2^q], \quad D_q = \frac{\tau(q)}{q-1} \tag{6.19}$$

and the singularity spectrum reads

$$\mathrm{f}(\alpha) = -c(\alpha)\log_2 c(\alpha) - (1-c(\alpha))\log_2(1-c(\alpha)), \tag{6.20}$$

with $c(\alpha) = (\alpha - \alpha_{\min})/(\alpha_{\max} - \alpha_{\min})$ and $\alpha_{\min} = -\log_2 p_1$, $\alpha_{\max} = -\log_2 p_2$.

In figure 6.3 I have depicted plots of the generalized dimensions D_q, the mass exponent $\tau(q)$ and the corresponding singularity spectrum $\mathrm{f}(\alpha)$ for the binomial fractal with $p_1 = 0.25$ and $p_2 = 0.75$. Inspection of the $\tau(q)$ and D_q curves shows that the $(q \to -\infty)$- and $(q \to \infty)$-asymptotes correspond to the end points of the singularity spectrum $\mathrm{f}(\alpha)$, $\alpha_{\min} = -\{\partial_q \tau(q)\}_{q\to\infty} = D_\infty$ and $\alpha_{\max} = -\partial_q\{\tau(q)\}_{q\to-\infty} = D_{-\infty}$ for the strongest and weakest singularity respectively. The maximum for the $\mathrm{f}(\alpha)$ spectrum is given by $\mathrm{f}(\alpha(q=0)) = D_0$ and equals unity. This value corresponds to the fact that all boxes contain a singularity as the limit $\sigma \downarrow 0$ is being approached, i.e. the probability of a box containing a singularity is one. As a consequence of the normalization, $\int \mathrm{d}\mu(x) = 1$, the mass exponent for $q = 1$ becomes, $\tau(1) = 0$ and that allows one to find the *information dimension* which is defined by $D_1 \triangleq \mathrm{f}(\alpha(q=1)) = \alpha(q=1)$. The information dimension specifies the subset, $A_{\alpha(q=1)}$ on which the measure is concentrated, carrying the measure.

6.4 Estimation of the $\tau(q)$ function and the singularity spectrum $\mathrm{f}(\alpha)$ by the method of box-counting

Several different alternatives exist to estimate the mass exponent $\tau(q)$ and the singularity spectrum $\mathrm{f}(\alpha)$ from *sampled* data. In this chapter I will only review the box-counting

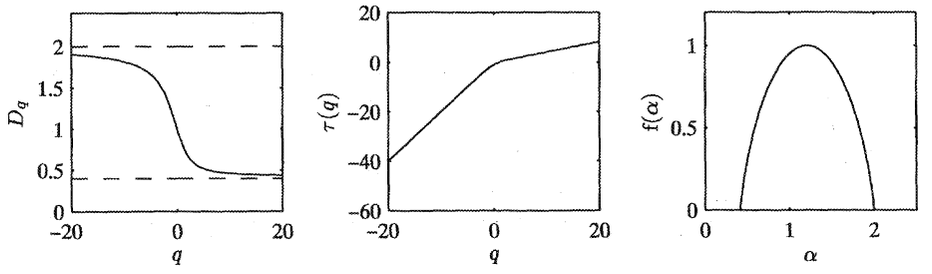

Figure 6.3 *The generalized dimensions D_q (left), the $\tau(q)$ function (middle) and singularity spectrum* $\mathrm{f}(\alpha)$ *(right) for a binomial multifractal measure with $p_1 = 0.25$ and $p_2 = 0.75$ as displayed in figure 6.2.*

analysis technique, while the superior technique based on the continuous wavelet transform will be treated in chapter 8. This box-counting technique is, in fact, the direct embodiment of the definition of the Hausdorff dimension. The following main steps can be recognized.

Procedure 6.1: Measurement of the singularity spectrum $\mathrm{f}(\alpha)$
The procedure to measure the mass exponent function $\tau(q)$ and the singularity spectrum $\mathrm{f}(\alpha)$ *from singular measures.*

1. *Partition the measure into an uniform grid with step size σ, conduct a coarse graining, and subsequently compute the mass in all segments,*

$$\mu(B_\sigma(x_i)) \triangleq \int_{B_\sigma(x_i)} \mathrm{d}\mu(x) = \int_{x_i-\frac{1}{2}\sigma}^{x_i+\frac{1}{2}\sigma} \mathrm{d}\mu(x) = \mu(x_i + \frac{\sigma}{2}) - \mu(x_i - \frac{\sigma}{2}) \quad \forall x_i \in A, \tag{6.21}$$

where x_i are the grid points. This action can be interpreted as a convolution where the smoothing kernel is given by the indicator function, the box-car function

$$\int_{-\infty}^{+\infty} I_\sigma(x - x')\mathrm{d}\mu(x), \tag{6.22}$$

where $I_\sigma(x) = 1$ on the interval $(x - \frac{1}{2}\sigma, x + \frac{1}{2}\sigma)$ and zero otherwise.

2. *Compute the partition function*

$$Z_\sigma\{I\}(q) = \sum_{\forall i} \mu[B_\sigma(x_i)]^q \sim \sigma^{\tau(q)} \mid_{\sigma \downarrow 0} \tag{6.23}$$

and conduct a linear regression in the $\log(\sigma)$-$\log[Z_\sigma(q)]$ *space, exploiting the property that a powerlaw becomes linear in* log-log *space.*

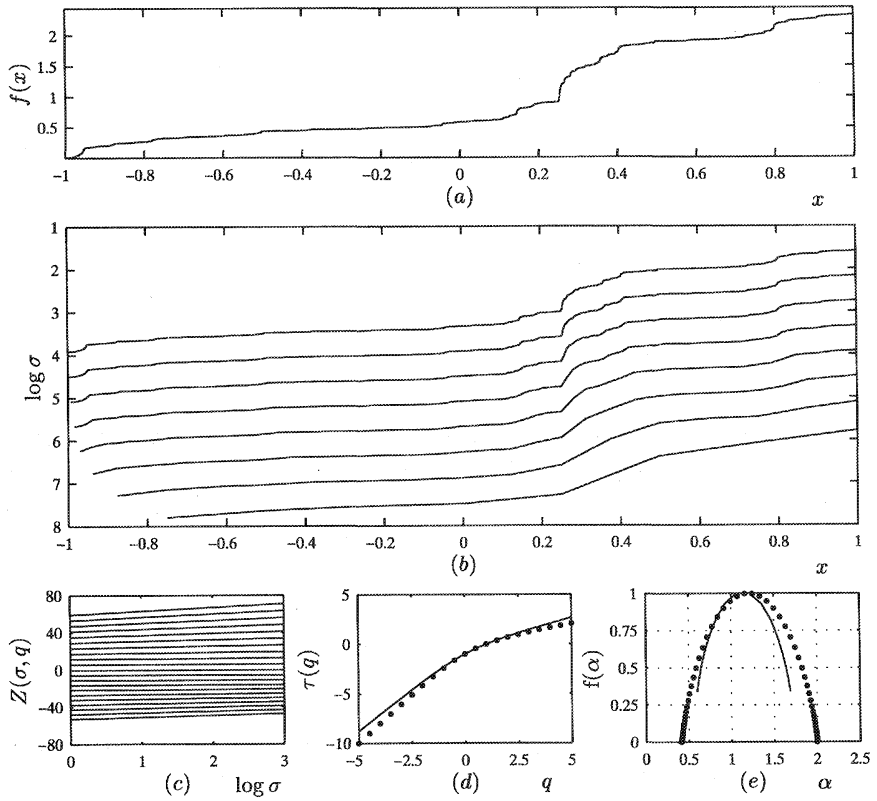

Figure 6.4 *Example of the estimation of the mass density functions $\tau(q)$ and singularity spectrum $f(\alpha)$ via the method of box-counting. (a) the binomial multifractal. (b) the logarithmically down-grained, with the box-car, binomial multifractal. (c) the partition function $Z_\sigma(q)$. (d) the estimated mass density function $\tau(q)$ obtained from the linear fits conducted on the partition function displayed in (c). (d) the singularity spectrum $f(\alpha)$ obtained from the $\tau(q)$ function via the Legendre transform.*

3. *Compute the Legendre transform to acquire the estimate for the singularity spectrum,*

$$\begin{cases} \alpha & = & \partial_q \tau(q) \\ f(\alpha) & = & \alpha q - \tau(q) \end{cases} \tag{6.24}$$

For the purpose of validating the successive steps of the proposed box-counting method yielding sample estimates for the mass density function $\tau(q)$ and singularity spectrum $f(\alpha)$ I included figure 6.4. In figure 6.4 (b) the successive smoothings, partitionings of the measure, acquired by smoothing the data with a box-car function of varying support, while in figure 6.4 (c) the partition function $Z_\sigma(q)$ plotted on log-log scale. Figure 6.4 (d) contains the estimated $\tau(q)$ function being obtained via linear regressions for every q on the partition function. These estimated values for the $\tau(q)$ function are subsequently used as input for the Legendre transform in order to compute the estimate for the singularity spectrum $f(\alpha)$. Both estimates for the $\tau(q)$ and $f(\alpha)$ are plotted jointly with the expected curves for the $\tau(q)$'s and $f(\alpha)$'s, given by equations (6.19) and (6.20). It is clear that the proposed estimation procedure gives the correct answers. The method is, however, limited in its application since it requires densities as input and moreover its singularity detection range is bounded, due to the poor regularity and integrability properties of the box-car.

6.5 Concluding remarks

This section has been written with the intention to provide the reader with a preliminary introduction on the concepts that will play an important role in capturing the complexity displayed by certain types of geophysical measurements. Amongst these rather superficially reviewed concepts the notion of scaling is prominent. It was shown that the complexity – for instance, the complexity generated by the multiplicative cascades defining the binomial multifractals – can elegantly be unraveled by inspection of the scaling behaviour displayed by a partition function. For the binomial multifractals this partition function can explicitly be defined in terms of the generator that was used to create the fractal set. In fact the scaling characteristics of the partition function are completely expressed by the mass exponent function from which then the singularity spectrum can readily be derived. This latter quantity expresses the relative occurrence of the different scaling exponents – that are appointed to the different singularities appearing in the measure – in terms of a Hausdorff dimension. So in a way the singularity spectrum delineates the hierarchy of scaling exponents to be associated with the singularities emerging in the measure as a result of the singular construction mechanism that is responsible for the generation of the binomial multifractals. To illustrate the mutual intertwinement of the yielded complexity with respect to the specific choices made in the construction of the measure I separate, following Siebesma (1989), the constructed measures into the following classes:

- A non-fractal/regular measure. In that case the singularity spectrum reduces to one point because one can appoint one trivial scaling exponent, $f(\alpha) = \alpha = p$, to all points in the measure yielding, $D(q) = p$ for the generalized dimensions and $\tau(q) = (q-1)p$ for the mass exponent function. Example: A uniform distribution on the

unit interval in one dimension, $p = 1$. This corresponds to a generalized Cantor set with a generator where all b boxes receive the same weight, $p_i = b^{-1}$, $i = 1, \ldots, b$.

- A homogeneous/mono fractal measure: In that case a homogeneous measure is defined on top of a fractal set with dimension $D = \dim_H$. This homogeneity refers to the fact that the scaling characteristics do not vary with position. Consequently, $D = D_q$ and $f(\alpha) = \alpha = D$. Example: A generalized Cantor set with a generator where the measure of n, $n < b$, boxes take the same non-zero value and where the remaining b^d boxes are deleted, $p_i = \frac{1}{n}$, $i = 1, \ldots, n$.

- A inhomogeneous/multi fractal measure. In that case one has an inhomogeneous measure defined where the scaling properties are varying with position. Consequently one finds a curve for the generalized dimensions, D_q, that change as a function of q. As a result of this is the associated singularity spectrum, $f(\alpha)$ is of a parabolic shape. Example: A generalized Cantor set with a generator where all boxes receive different weights p_i.

It is the extension to the inhomogeneous case that is of great importance and it is a direct reflection of the method of *curdling*, as proposed by Mandelbrot (1974), that introduces the required *intermittency*. Especially this *intermittency* – manifesting itself by the occurrence of relative violent periods/regions superseded by relatively calm periods/regions in the system's activity – that reflects a general property so characteristic for many (geo)physical processes involving dissipation, e.g in hydrodynamic turbulence (Mandelbrot, 1974; Parisi and Frisch, 1985; Schertzer and Lovejoy, 1987b; Schmitt et al., 1992), and justifies the generalization to multifractals. In fact the multifractal formalism stands for a decomposition of a fractal measure into its constituents, that is into its interwoven sets each of which is characterized by singularity strength[4] α and its Hausdorff dimension $f(\alpha)$. It was shown that this singularity spectrum, $f(\alpha)$, is directly related to the generalized or Renyi dimensions. These generalised dimensions can readily be obtained from the asymptotic powerlaw behaviour displayed by the multiscale partition function and in this way yielding estimates for the mass exponent function a quantity that is related to the generalised dimensions according to, $D_q = \frac{\tau(q)}{q-1}$.

By way of the construction of the partition there exists an intrinsic analogy between phenomenological/statistical thermodynamics and the multifractal formalism (Mandelbrot, 1974, 1982; Hentschel and Procaccia, 1983; Parisi and Frisch, 1985; Schertzer and Lovejoy, 1987b) annex the dynamical system theory (Lichtenberg and Lieberman, 1992). This analogy becomes apparent because it can be shown[5] that the pair of quantities, q and $\tau(q)$, play the same role as the *inverse of the temperature* and the *Gibb's free energy*

[4] In later chapters I will refer to the scaling exponent α as the Hölder exponent.

[5] I do not exactly know who established this relationship first but I think I have to refer to (Collet, 1986; Mandelbrot, 1989).

whereas their conjugates can be recognized as the quantities α and $f(\alpha)$ that can be linked to the *energy* and the *entropy* respectively.

The author is aware of the fact that the concepts reviewed in this chapter are not amongst the most common in the exploration geophysical world. However, they are indispensable in the quest to characterize the complexity of well-log measurements for example. In the pertaining chapters I will try to construct a more general framework in support of these concepts. This framework will not only refer to how to measure a fractal construct by means of a multiscale partitioning, coarse graining, by a proper family of analyzing functions, see chapter 5 where the basics of distribution theory was given and chapter 8 where wavelets are proposed as multiscale analyzing functions.

Chapter 7

Selected topics on stochastic monofractals

7.1 Introduction

In one way or the other, scale invariant stochastic processes – processes whose statistical behaviour is similar under judicious scaling operations on time and/or space, think of a Brownian motion trajectory – play an important role in the characterization of irregular processes (Mandelbrot and Wallis, 1969; Tartarskii, 1971; Yaglom, 1987; Samarodnitsky and Taqqu, 1994). This similarity under dilatations, scale transformations, is in case of *homogeneous* monofractal scaling indexed by the Hurst exponent[1] H and can directly be associated to the fundamental renormalization group property stating that the random process $X(z)$ and its dilatated version $\sigma^H X(\sigma z)$ have the same stochastic properties, i.e. they share the same probability density function. This type of behaviour, the renormalization group property, is commonly encountered for irregular processes such as $\frac{1}{f}$ noises, pink or "shot" noises and forms one of the important building blocks in statistical mechanics (Wilson, 1983).

A well known example of a self-similar random process is Brownian motion (Einstein, 1905). This process, also known as a Wiener process, is stochastically *non-stationary*, a notion which corresponds to the *infrared* catastrophe, the divergence of the power spectrum as the frequency tends to zero. The first increments, however, are stationary, Gaussian and independent. Mandelbrot and Wallis (1969) refer to the non-stationarity of Brownian motion as the "Joseph effect", a biblical metaphor for the apparent long tailed correlations which can be associated with the extensive periods of prosperity and hardship experienced by Joseph. Besides this effect, there is the so-called "Noah effect" which reflects the fact that Noah survived an exceptional high flood, which on its turn, can be associated with a random process supporting very large outliers rather than having the mass of its fluctuations concentrated around the mean. This latter effect constitutes

[1] Notice that numerous names exist for this scaling exponent. The choice of the name I will use will be context driven.

an extension to the non Gaussian α_l indexed Levy noises (Samarodnitsky and Taqqu, 1994; Schertzer and Lovejoy, 1987a; Montroll and Schlessinger, unknown) yielding a much larger *variability*[2].

At first a formal definition for homogeneous or monofractal scaling will be given, followed by a discussion on the properties of Brownian motion, the classical *non-differentiable* but *continuous* random process. Then, in section 7.3 Brownian motion is generalized to index-H fractional Brownian motion, fBm, via the method of fractional integration/differentiation[3] allowing for a true invocation of the "Noah" effect, where either periods of prosperity and hardship alternate on a limited time range, the process behaves *antipersistent*, or where the periods of wealth and hardship *persist* over longer periods. A similar generalization can be applied to the α_l stable Levy motion (Samarodnitsky and Taqqu, 1994; Schertzer and Lovejoy, 1987a; Montroll and Schlessinger, unknown) yielding a process with both the "Joseph" and the "Noah" effects incorporated.

7.2 Brownian motion

Brownian motion (a special limiting case of the Wiener-Markov or Ohrnstein-Uhlenbeck process (Priestley, 1981) is the most commonly known example of a *non-differentiable*, *self-similar* and stochastic *non-stationary* process being associated to the "classical" diffusion process. This non-differentiability is a direct consequence of the fact that the process diverges as the *spatial* and/or *temporal* resolution – spatial for the trajectory and temporal for say the x coordinate of the trajectory as a function of time – is increased, because the process reveals more and more spatial and temporal detail. This divergence is known as the *ultraviolet* catastrophe, whereas on the other hand the non-stationarity of Brownian motion can be linked to the *non-integrability* of the process, the non-integrable pole, singularity, at origin of the frequency spectrum, yielding the so-called *infrared* catastrophe. Despite, these apparent difficulties Brownian motion has the advantage over the stationary Markov process that it does not display a break in its scaling behaviour. There is no characteristic scale, a notion especially beneficial when complex phenomena such as the ones being studied in this thesis are of interest. Because of divergences Brownian motion has to be interpreted in the sense of distributions because only in this way it can become an observable by measuring, testing the process with the proper test/analyzing function, i.e. mapping it to a finite scale/resolution σ.

By setting the damping factor to zero of the Wiener process one finds the following first order stochastic differential equation

$$\frac{\mathrm{d}X(z)}{\mathrm{d}z} = \epsilon(z), \tag{7.1}$$

[2]See Klafter et al. (1996) for an informal discussion on the physical relevance and historical perspective of this subject.

[3]Defined in the sense of distributions, see chapter 5.

the solution of which is given by

$$X(z) = \int_0^z \epsilon(z')\mathrm{d}z' \qquad \text{with } X(0) = 0, \tag{7.2}$$

and defines a record of Brownian motion, $B(z) \triangleq X(z)$, in case the driving force, ϵ, is white Gaussian noise or defining a Levy motion, a Levy flight, when ϵ is an α_l stable Levy noise[4]. As a consequence of the limit procedure on the correlation length annex damping factor the process itself behaves as,

$$B(z) \sim z^{\frac{1}{2}} \qquad \forall z \in \mathbb{R}^+ \tag{7.3}$$

where the symbol $\sim$ has been used to indicate equality within a slowly varying factor, whereas the variance behaves as,

$$\sigma^2(z) = \langle X^2(z) \rangle \sim z \qquad \forall z \in \mathbb{R}^+ \tag{7.4}$$

and this, on its turn, implies the following behaviour for the covariance function

$$s(\sigma, z) = \langle X(z+\sigma)X(z) \rangle \sim \sigma^2(|\sigma| + |z|). \tag{7.5}$$

It is evident that all above quantities have become functions of the position z, they grow with z, and that observation unfortunately leads to a divergence of these quantities as the limit $z \to \infty$ is being approached. So, strictly speaking, these quantities can not longer be considered as *observables* and in practical circumstances this refers, as le Méhauté (1991) puts it, to the way in which, say, the variance approaches infinity as the length of the sample record is increased. In fact it expresses a lack of convergence for, say, the sample estimates for the variance.

In order to circumvent the above problem of *non-integrability*, the divergence of the variance and covariance as $z \to \infty$, one can introduce a new process, defined in terms of the *first increments*, $\Delta X(z, \sigma) \triangleq X(z+\sigma) - X(z)$. In such a definition the σ plays the role of an effective gauge delineating the scale/resolution range over which the original function is being studied. In this way the long period diverging cycles are omitted and the new process can shown to be *stationary*. Such an approach, a *regularization*, maps the process onto an *observable* and finds its roots in distribution theory, see chapter 5, and amounts, in case the second order statistics are of interest, to the definition of a variogram/structure function analysis. From this one finds, by evaluation of the mean square increments,

$$D(\sigma) = \langle |\Delta X(z, \sigma)|^2 \rangle \sim |\sigma|, \tag{7.6}$$

that the variogram displays a normal shift invariant behaviour since the variogram only depends on the scale indicator σ. Clearly this behaviour corresponds to the stochastic

[4]See Mandelbrot (1967, 1982), Tartarskii (1971) and Yaglom (1987) for a detailed discussion.

stationarity yielded by the first increments. Summarizing one can say that the stationarity constraints imposed by the structure function are less severe, it only demands a process with stationary first increments, a notion exactly met by Brownian motion.

Another way of understanding the non-stationarity issue can be given by the argument that the inverse Fourier transform of the power spectrum of Brownian motion,

$$S(k) \sim \frac{1}{k^2} \tag{7.7}$$

is not defined, the integral diverges due to the *infrared catastrophe* at the origin, the divergence for $k \downarrow 0$. This type of singularity at the origin – due to the fact that the solution of equation (7.1) involves a division by $-(jk)^2$ in the frequency domain – can easily be circumvented by considering the first increments rather than the process itself. In this way the integral is made to converge since the Fourier transform of the increment acts as a multiplication by $\frac{\sigma}{4}\sin^2(\frac{\sigma}{2}k)$, it regularizes the divergent inverse Fourier transform of (7.7), i.e. $\lim_{\sigma\downarrow 0}\mathcal{F}\{\Delta X\}(k) \sim k^2$. This procedure is one example of how to deal with the divergence of singularities, to deal with algebraic singularities in the sense of distributions, see section 5.3.

7.3 Fractional Brownian motion

Mandelbrot and Wallis (1969) proposed to extend the ordinary Brownian motion process to the index-H fractional Brownian motion, fBm, a process already recognized for $H = \frac{1}{3}$ by Kolmogorov (1941). They accomplish this extension by applying a *fractional* integration/differentiation to the ordinary Brownian motion process as defined in section 7.2, i.e.

$$B_H(z) \triangleq \mathcal{I}^\alpha\{B\}(z) \qquad \alpha = H - \frac{1}{2} \qquad 0 < H < 1. \tag{7.8}$$

This operation $\mathcal{I}^\alpha$ – see section 5.8 for a detailed discussion – acting on an arbitrary[5] function $f(z)$ is defined by,

$$\mathcal{I}^\alpha\{f\}(z) \triangleq \int_0^{+\infty} \frac{(z-z')^{-\alpha-1}}{\Gamma(-\alpha)} f(z')\mathrm{d}z' \tag{7.9}$$

and performs a fractional integration for $\alpha > 0$ or a fractional anti-integration (differentiation) for $\alpha < 0$. Under similar arguments as for divergence of the Fourier integral, this procedure has to be regularized since Brownian motion is not strict differentiable/integrable, yielding the following proposition

[5] When this function is not fractionally integrable/differentiable then this operation has to be interpreted in the sense of distributions.

Proposition 7.1: Fractional Brownian motion (fBm) (Samarodnitsky and Taqqu, 1994)
Let $0 < H < 1$. Then standard fractional Brownian motion $\{B_H(z),\ l \in \mathbb{R}\}$ has the integral representation

$$B_H(z) = \frac{1}{C(H)} \int_{-\infty}^{+\infty} K(z, z')\mathrm{d}B(z'), \quad z \in \mathbb{R} \tag{7.10}$$

where the Schwartz's kernel is given by

$$K(z, z') = [((z - z')_+)^{H-\frac{1}{2}} - ((-z')_+)^{H-\frac{1}{2}}]$$

and where

$$C(H) = \left\{ \int_0^{\infty} ((1 + z')^{H-\frac{1}{2}} - (z')^{H-\frac{1}{2}})^2 \mathrm{d}z' \right\}^{\frac{1}{2}} + \frac{1}{2H}.$$

When $H = \frac{1}{2}$, $C(\frac{1}{2}) = 1$ and the representation (7.10) is to be interpreted as the integral of white noise.

Notice that in this definition again an increment is being used to regularize the integral of equation (7.9). As already mentioned the idea behind this is very important and the reader is referred section 5.8 for a more elaborate discussion on this issue.

The generalization as coined in proposition 7.1 amounts for the variance and covariance to

$$\sigma^2(z) \sim |z|^{2H} \tag{7.11}$$

and

$$\mathcal{C}\{B_H(z_1), B_H(z_2)\}(z_1, z_2) \sim |z_1|^{2H} + |z_2|^{2H} - |z_1 - z_2|^{2H} \tag{7.12}$$

both obviously being *non-stationary*. As for ordinary Brownian motion the increments behave stationary allowing for a sensible definition of the structure function

$$D(\sigma) \sim \sigma^{2H} \tag{7.13}$$

and the associated power spectrum

$$S(k) \sim \frac{1}{k^{\beta}} \qquad 1 < \beta = 2H + 1 < 3. \tag{7.14}$$

7.4 Stochastic self-similarity, homogeneous scaling and some general properties

One of the main assets of Brownian motion and its generalization, fractional Brownian motion, is that they adhere to the self-affine scaling relationship

$$B_H(\sigma z) \stackrel{\mathrm{d}}{=} \sigma^H B_H(z), \tag{7.15}$$

which can easily be understood for the Brownian case, $H = \frac{1}{2}$, by reconciliation of the renormalization property for the sum of random numbers as being given by the central limit theorem, see section 7.5. The above scaling relation is a clear example of stochastic self-similarity[6] and is formally defined as

Definition 7.1: Self-similarity (Samarodnitsky and Taqqu, 1994)
The real valued process $\{X(z),\ z \in \mathbb{R}\}$ *is self-similar with index* $H > 0$ *if for all* $\sigma > 0$, *the finite-dimensional distributions of* $\{X(\sigma z),\ z \in \mathbb{R}\}$ *are identical to the finite dimensional distributions of* $\{\sigma^H X(z),\ z \in \mathbb{R}\}$; *i.e. if for any* $d \geq 1$, $z_1, z_2 \ldots, z_d$ *and any* $\sigma > 0$,

$$(X(\sigma z_1), X(\sigma z_2), \ldots, X(\sigma z_d)) \stackrel{\mathrm{d}}{=} (\sigma^H X(z_1), \sigma^H X(z_2), \ldots, \sigma^H X(z_d)). \tag{7.16}$$

Notation:

$$\{X(\sigma z),\ z \in \mathbb{R}\} \stackrel{\mathrm{d}}{=} \{\sigma^H X(z),\ z \in \mathbb{R}\} \quad or \quad X(\sigma z) \stackrel{\mathrm{d}}{=} \sigma^H X(z) \quad \forall z \in \mathbb{R}. \tag{7.17}$$

The different statistical moments of a homogeneous scaling process adhere to the scaling relation

$$\langle |B_H(\sigma z)|^q | \rangle \stackrel{\mathrm{d}}{=} \sigma^{qH} \langle |B_H(z)|^q \rangle, \tag{7.18}$$

where the angular brackets have to be interpreted as taking the stochastic expectation.

In order to conclude the discussion on the processes Brownian motion and its extension to fBm I will briefly summarize their most important properties.

- Records of Brownian motion and fBm are *singular* "functions" in the sense that they are not *strictly* differentiable. They are, however, differentiable in a weak sense, in the sense of distributions. To be more specific they belong to the Hölder continuous class of functions C^H, $0 < H < 1$. On the other hand Brownian motion and fBm are *non-integrable* a notion which corresponds to the stochastic *non-stationarity* or *non-conservation* of the mean, see Schertzer and Lovejoy (1987a). This non-integrability corresponds to the infrared divergence for the spectrum and can be regularized by consideration of the first increments rather than the process itself. So in fact the Hurst or Hölder exponent H plays a kind of double role. On the one hand it describes the differentiability, index-H Brownian motion is "H" times differentiable whereas on the other hand the H rules the divergence of the mean, $\langle B_H(z) \sim z^H \rangle$.

[6] The terms self-similarity and self-affinity are both used in literature. It is worthwhile to mention that self-affine scaling is related to functions where the scaling of the dependent variable is *deviant*.

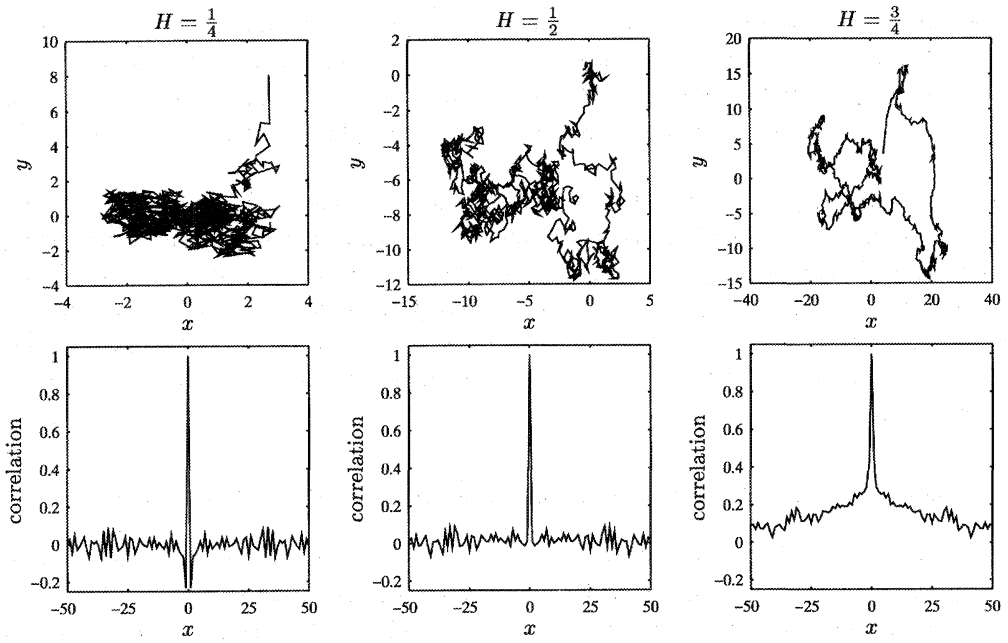

Figure 7.1 *Examples of simulations of ordinary Brownian motion (middle column, top row) and its extensions to antipersistent fBm ($H < \frac{1}{2}$) for $H = \frac{1}{4}$ (first column, top row) and persistent fBm ($H > \frac{1}{2}$) for $H = \frac{3}{4}$ (last column, top row). I also included the covariance function of the increments (bottom row).*

- For $H \downarrow 0$ the process becomes increasingly antipersistent and the increments (also known as fractional Gaussian noise (fGn)) become increasingly *anticorrelated* and the "function"[7] will be discontinuous at every abscissa in the limit, i.e. a so-called flicker noise $S(k) \sim k^{-1}$. But in the limit for H going to zero the process becomes stationary! See figure 7.1 top row for an example of a simulated record of fBm with the Hurst exponent H set to $H = \frac{1}{4}$ and the corresponding covariance for the stationary increments. Notice the anticorrelation being manifested in the negative first lag terms[8].

- For $H = \frac{1}{2}$ the process becomes Brownian and the increments, independent uncorrelated[9] Gaussian white noise. See the middle column of figure 7.1 where I depicted a simulated record of Brownian motion and the corresponding delta function shaped covariance function.

[7] It is not really a function in the limit but a so-called generalized function, see section 5.8.

[8] This covariance function is strictly speaking a tempered distribution, a fractional derivative of the Dirac delta functional.

[9] That is in fact how Gaussian white noise is defined.

- For $H \uparrow 1$ the process becomes increasingly persistent and the increments correlated. In the limit the process will be smooth (one times differentiable), i.e. $B_H(z) \in C^1$ $H \uparrow 1$. See figure 7.1 right column where I depicted a simulation for $H = \frac{3}{4}$. It is clear that the covariance functions displays a long term correlation.

Finally I would like to conclude this rather informal discussion on Brownian Motion and fBm by stating that white Gaussian noise can also be seen as a homogeneous (one scaling exponent) self-affine scaling functional but now with a value for H equal to $H = -\frac{1}{2}$ and similarly this white noise process can be extended to fractional Gaussian noise (the increments of fBm) with $-1 < H < 0$. This process is, however, not *continuous* and not *differentiable*, it is a singular signed measure, and can only be given a meaning by means of a regularization.

7.5 Digression on α_l Levy stable distributions, universality under addition

Now that the most essential tools for the description of a random variable have briefly been introduced the question arises: How to characterize the joint behaviour of random variables? In that respect the study of the limiting behaviour for sums of *independently identically distributed* (i.i.d.) random variables becomes of great importance, especially since it allows for a natural introduction of the Gaussian distribution. This study also serves as a point of departure for the extension of the central limit theorem, with its associated Gaussian distribution, to a more general version of the central limit theorem yielding a comprehensive class of probability functions of which the Gaussian distribution is just a special case.

One has to go as far back as the 1930's when Paul Levy and Aleksander Khinchine developed the theory of *stable distributions* (Feller, 1950; Samarodnitsky and Taqqu, 1994). This theory deals with the interesting and important question whether a universal renormalization group[10] property exists, the so-called stable fixed points, for the sum of n, $n \geq 2$ i.i.d. random variables, i.e. the $\sum_{i=1,n} X_i$ with $X_i \stackrel{\mathrm{d}}{=} X_1$, $\stackrel{\mathrm{d}}{=}$: identical in distribution. So the question is whether the real constants a_n and b_n, $b_n > 0$ exist for which the following linear renormalization equation,

$$\sum_{i=1,n} X_i \stackrel{\mathrm{d}}{=} b_n X_1 + a_n, \tag{7.19}$$

holds for the sum of the n random variables?

Indeed the well-known *central limit theorem* states that the sum for random variables with arbitrary probability functions, but with finite and equal first and second moment,

[10]The reader is referred to (Wilson, 1983) who developed the framework for "multiple-scale-of-length" approach (Nottale, 1995) in the field of statistical physics. Note, however, that the properties of the stable distributions were derived before renormalization group theory.

$\langle X_i \rangle < \infty$, $\langle X_i^2 \rangle < \infty \ \wedge \langle X_i \rangle = \langle X_1 \rangle$, $\langle X_i^2 \rangle = \langle X_1^2 \rangle \ \forall \{i \in 1, \ldots, n\}$ with $\langle X_1 \rangle$, $\langle X_1^2 \rangle$ being the preset values for the cumulants, approaches a limiting form for its probabilistic behaviour. This limiting behaviour involves the property that the probability function of X_1,

$$X_1 \stackrel{\mathrm{d}}{=} \lim_{n \to \infty} [\sum_{i=1,n} X_i - a_n] \frac{1}{b_n}, \tag{7.20}$$

has the same form as that of the renormalized sum, in case the renormalization factors equal,

$$\begin{aligned} b_n &= n^{\frac{1}{2}}, \\ a_n &= (n - \sqrt{n}) \langle X_1 \rangle. \end{aligned} \tag{7.21}$$

Such an observation clearly marks the insensitivity of the sum, on the right hand side of equation (7.20), for the details in the probability functions of the constituents and it is this property which led to the well-known Gaussian probability distribution (Feller, 1950). This Gaussian probability measure is defined by

$$\mathrm{d}F(x) \triangleq \frac{1}{\sqrt{2\pi}} \exp[-\frac{(x - \langle X \rangle)^2}{2\sigma^2}] \mathrm{d}x, \tag{7.22}$$

with its second characteristic function given by

$$K(q) = \frac{\sigma^2}{2} q^2. \tag{7.23}$$

and $\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2$. From this kind of second characteristic it can easily be deduced that the Gaussian is completely defined by its first two cumulants because only the first two derivatives of equation (7.23) differ from zero.

Indeed the above renormalization property, the property that the random variables are stable fixed points under the renormalized addition, equation (7.20), can be interpreted as if the Gaussian distribution is acting as an *universal attractor* for the sums of i.i.d. random numbers. More formally,

Definition 7.2: Stable distribution (Samarodnitsky and Taqqu, 1994)
A random variable X is said to have a stable distribution if for any $n \geq 2$, there is a positive number b_n and a real number a_n (recentering term) such that

$$X_1 + X_2 + \cdots + X_n \stackrel{\mathrm{d}}{=} b_n X + a_n, \tag{7.24}$$

where $X_1, X_2, \ldots, X_n$ are independent copies of X.

The stability refers to the notion that the first characteristic function for the sum of say two random variables – in this example the strict stable situation is assumed leading to the omission of the recentering term,

$$bX = b_1 X_1 + b_2 X_2, \tag{7.25}$$

where the constant b is related to the constants b_1 and b_2 – can be written as

$$\langle e^{iqbX} \rangle = \langle e^{iqb_1X_1} \rangle \langle e^{iqb_2X_1} \rangle = \langle e^{iq\{b_1X_1+b_2X_2\}} \rangle. \tag{7.26}$$

This enables one to write for the second characteristic function (West, 1990)

$$K(bq) = K(b_1q) + K(b_2q). \tag{7.27}$$

Notice that the multiplicative property for the first characteristic function, see equation (7.26), points to an invariance of a *stable* probability function to the operation of *convolution*; since multiplication in the Laplace/Fourier domain corresponds to convolution in the other domain. Consequently the behaviour for the sum of n random numbers tends to the following limit for the second characteristic function

$$K(q) = \lim_{n\to\infty} \sum_{i=1}^{n} K_i(b_nq) + a_n. \tag{7.28}$$

At this point Levy asked the following question: What is the most general form for a probability function which obeys a similar renormalization property as the Gaussian? He showed that by relaxing the condition on the finiteness of the mean and variance the renormalization property of equation (7.21) can be generalized to

$$\begin{array}{rcll} b_n & = & n^{\frac{1}{\alpha_l}} & 0 < \alpha_l \leq 2 \\ a_n & = & n^{\frac{1}{\alpha_l}} \langle X_1 \rangle - \langle X \rangle & 1 < \alpha_l \leq 2, \end{array} \tag{7.29}$$

where the divergence, in case $0 < \alpha_l \leq 1$, withstands a definition of the a_n over this interval.

These generalized renormalization factors lead to the definition of an extended class of stable distributions, indexed by the Levy index α_l, by means of the following theorem and definition for the first characteristic function.

Theorem 7.1: Index of stability α_l (Samarodnitsky and Taqqu, 1994)
For any stable random variable X, there is a number $\alpha_l \in (0, 2]$ such that the number C satisfies

$$C^{\alpha_l} = \sum_{\forall i} X_i^{\alpha_l}. \tag{7.30}$$

The number α_l is called the index of stability or the characteristic exponent. See Feller (1950), section VI.1 for the proof.

Definition 7.3: First characteristic function (Samarodnitsky and Taqqu, 1994)
A random variable X is said to have a stable distribution if there are parameters

$0 < \alpha_l \leq 2$, $-1 \leq \beta \leq 1$, *and* μ *real such that its first characteristic function has the following form*

$$\langle e^{iqX} \rangle \triangleq \begin{cases} \exp\left[-\sigma|q|^{\alpha_l}(1 - i\beta \text{sign}(q) \tan\frac{\pi\alpha_l}{2}) + i\mu q\right] & \alpha_l \neq 1, \\ \exp\left[-\sigma|q|(1 + i\beta \text{sign}(q)\frac{2}{\pi}\ln|q|) + i\mu q\right] & \alpha_l = 1. \end{cases} \tag{7.31}$$

The parameter α_l *is the index of stability and* sign *the* sign *function. The parameters* α_l, β *and* μ *are unique.*

With this definition the second Fourier characteristic is found to be of the form

$$K(q) = \log\langle e^{iqX} \rangle = -\sigma(iq)^{\alpha_l}. \tag{7.32}$$

It can be shown (Feller, 1950; Schertzer and Lovejoy, 1987a; West, 1990) that this power law form – for the admissible range of the Levy index $0 < \alpha_l \leq 2$ – constitutes the only functional form supporting both the renormalization behaviour as depicted in equation (7.28) as well as the technical conditions to be imposed on the inverse Laplace/Fourier transform of the first characteristic function, the probability function.

Only for the special cases of $\alpha_l = \frac{1}{2}$, $\alpha_l = 1$ and $\alpha_l = 2$ closed form expressions for probability function can be found. In the latter case the inverse Laplace/Fourier transform yields the Gaussian which is extremely regular since all the moments $q > 2$ exist. For the other situations, $0 < \alpha_l < 2 \wedge \alpha_l \neq \frac{1}{2}, 1$, one is limited to information on the asymptotic behaviour of the distribution solely (Montroll and Schlessinger, unknown; Feller, 1950; Schertzer and Lovejoy, 1987a; West, 1990),

$$\mathrm{d}F(x) \sim \frac{1}{x^{\alpha_l+1}}\mathrm{d}x \qquad x \to \infty, \quad 0 < \alpha_l < 2. \tag{7.33}$$

Random variables displaying this kind of behaviour are referred to as *hyperbolic* or *fat-tailed* random variables due to the algebraic fall off of the probability function. A direct consequence of this algebraic tail is that the Levy index α_l introduces a divergence for the statistical moments exceeding the exponent, $q > \alpha_l$. In case $1 < \alpha_l < 2$ the variance will diverge and this corresponds, in practical situations, to an absence of convergence for the *sample* variance as the sample size becomes increasingly larger. In matter of fact this "divergence" refers to the way in which the variance approaches infinity and should not be interpreted literally in practical circumstances because then one could argue that this type of behaviour is irrelevant since all *sample* estimates are always finite! For the situation where $0 < \alpha_l < 1$ the mean diverges and that can be linked to sample data records with (a) huge outlier(s). These outliers are connected to the long fat tail and completely dominate the sample mean. For that reason the sample mean will be fully governed by this sole outlier, irrespective of the other samples, and one can therefore not longer talk in sensible way about the mean and the variance in these cases because these

quantities do not give any information. Finally notice that this particular behaviour is quite opposite to the behaviour of the Gaussian where the mass is mainly concentrated around the mean.

7.6 Fractional Levy motion

In a similar fashion as for the Brownian motion, Levy motion (Levy flight) can be obtained by replacing the Gaussian forcing term in equation (7.1) by a α_l indexed stable Levy process. There is, however, a substantial difference. The self-similarity exponent H (scaling exponent) of the Levy flight equals $\frac{1}{\alpha_l}$ rather than $\frac{1}{2}$ for Brownian motion. This difference is a direct consequence of the generalization of the central limit theorem as coined by Paul Levy. The following self-affine scaling relation – being directly related to the generalized renormalization group properties for the sum of stable random variables – applies to a record of Levy motion

$$L_{\alpha_l}(\sigma z) \stackrel{\mathrm{d}}{=} \sigma^{\frac{1}{\alpha_l}} L_{\alpha_l}(z), \tag{7.34}$$

where $L_{\alpha_l}(z)$ denotes the Levy flight. Clearly, the deviant behaviour of the Levy flight restricts the possible degree of fractional integration/differentiation to

$$\begin{aligned} \alpha_l < 1 \quad &\Longrightarrow \quad 0 < H < \frac{1}{\alpha_l} \\ \alpha_l \geq 1 \quad &\Longrightarrow \quad 0 < H \leq 1, \end{aligned} \tag{7.35}$$

yielding a similar generalization of the Levy flight as the generalization of Brownian motion to fractional Brownian motion. The formal definition of fractional Levy motion $L_{\alpha_l,H}$ reads,

Definition 7.4: Fractional Levy noise (Samarodnitsky and Taqqu, 1994)
If H complies to the ranges as indicated in equation (7.35) then standard fractional Levy motion $\{L_{\alpha_l,H}(z)$ has the integral representation

$$L_{\alpha_l,H}(z) = \frac{1}{C(H)} \int_{-\infty}^{+\infty} K(z,z') \mathrm{d}L_{\alpha_l}(z'), \quad z \in \mathbb{R} \tag{7.36}$$

where the Schwartz's kernel is given by

$$K(z,z') = [((z-z')_+)^{H-\frac{1}{\alpha_l}} - ((-z')_+)^{H-\frac{1}{\alpha_l}}]$$

and where

$$C(H) = \left\{ \int_0^{\infty} ((1+z')^{H-\frac{1}{2}} - (z')^{H-\frac{1}{2}})^2 \mathrm{d}z' \right\}^{\frac{1}{2}} + \frac{1}{2H}$$

.

Chapter 8

Multiscale analysis by the continuous wavelet transform

8.1 Introduction

The recent advent of the continuous wavelet transform in the field of multiscale analysis (Morlet et al., 1982a,b; Holschneider, 1987, 1995; Jaffard, 1989; Mallat and Hwang, 1992; Mallat and Zhong, 1992; Daubechies, 1992; Ghez and Vaienti, 1992; Bacry et al., 1993; Farge et al., 1993) is not surprising since this scale transformation allows for a localized decomposition of mathematical objects – e.g functionals containing algebraic singularities or fractals being singular everywhere – as well as physical objects – e.g. well-log measurements[1] – into their multiscale constituents. From these multiscale constituents, living in the space-scale space, it is possible to regain the original function by applying the inverse wavelet transform.

Consequently the multiscale decomposition by means of the wavelet transform not only constitutes the perfect tool for conducting a scaling analysis focussing on certain local features of the object but it also allows for an unraveling of scaling behaviour in a global sense. These two different procedures boil down to either inspecting the local behaviour of certain relevant features in the space-scale space – via inspection of the wavelet coefficients – or by defining a multiscale partition function unraveling the scaling complexity of the total object.

By way of its construction the wavelet transform is very closely related to ideas behind distribution theory, see chapter 5, because the sequences of test functions appearing in that theory can be related to the family of analyzing wavelets, of decreasing supports, whereas the object under investigation can be identified with possibly singular distributions. In fact wavelets constitute a double indexed family of functions that are obtained by the invocation of the shift and dilatation operation on an initial wavelet function.

[1] See chapter 2.

The fact that the wavelets are constructed in this way makes them suitable to analyze, by decomposition, scaling objects such as multifractals that contain a whole hierarchy of different singularities that can be characterized by varying scaling exponents. Therefore the wavelet transform is the proper choice for conducting the multiscale analysis. These scaling exponents refer to linear estimates for the decay/growth rate displayed by the logarithm of wavelet coefficients versus the logarithm of the scale indicator. These estimates can either refer to the local scaling behaviour induced by one specific event or to, say, the global average scaling behaviour to be interpreted in a more statistical sense, in the sense that one speaks on the relative number of occurrences of a certain scaling exponent.

By conducting the local analysis one is able to obtain information on the local degree of differentiability by appointing a scaling exponent, also known as the Hölder exponent, to the local scaling behaviour of the wavelet coefficients (Holschneider, 1987; Jaffard, 1989; Mallat and Hwang, 1992; Mallat and Zhong, 1992). Information on the global differentiability, on the other hand, can be obtained by inspecting the endpoints of the singularity spectrum. This singularity spectrum constitutes an unraveling of the emerging scaling exponents intrinsic in the multifractal structure (Mandelbrot, 1974; Parisi and Frisch, 1985; Collet, 1986; Collet et al., 87; Schertzer and Lovejoy, 1987b; Mandelbrot, 1989; Siebesma, 1989; Lichtenberg and Lieberman, 1992; Bacry et al., 1993) and is directly related to the generalized or Renyi dimensions (Renyi, 1970) that are readily obtained from the multifractal partition function defined in terms of the wavelet transform (Arneodo et al., 1988; Bacry et al., 1993).

The setup of this chapter is as follows. First of all I will start to define the wavelet transform and provide a resume of some of its properties. Then I will show how the wavelet transform can be used to reveal the scaling behaviour and to measure the degree of differentiability. Since the space-scale space is highly redundant – it is the result of a decomposition of a one-dimensional object into function living in a two-dimensional space-scale space – it is beneficial to introduce an efficient partitioning for this space. This partitioning has to act as a guide for the analysis delineating the essential features in the space-scale plane. I will show that the wavelet transform modulus maxima formalism as coined by Stephane Mallat provides such a mechanism (Mallat and Hwang, 1992; Mallat and Zhong, 1992). He demonstrates that this optimal partitioning does not only allow one to designate a local Hölder exponent to a singularity, via inspection of the space-scale plane along the wavelet transform modulus maxima lines, but it also serves as the optimal covering required to define a partition function required to analyze the multifractal scaling characteristics. Then I will continue this chapter by going into detail on the relation between the properties of analyzing wavelet and detectable range of scaling exponents. Finally notice that this chapter serves as the theoretical basis for the application of the multiscale ideas presented in chapter 2.

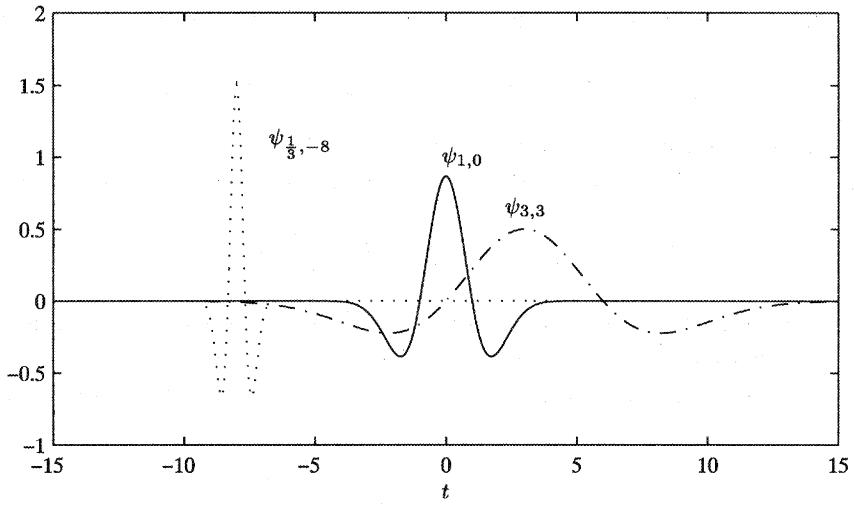

Figure 8.1 *The Mexican hat for various values of scale parameter σ and translation parameter x'. If $|\sigma| < 1$ the wavelet is contracted, if $|\sigma| > 1$ the wavelet is expanded. If $x' < 0$ the wavelet is shifted to the left, if $x' > 0$ the wavelet is shifted to the right, assuming $\sigma > 0$.*

8.2 The continuous wavelet transform

Given the fact that scaling plays a crucial role in this thesis it is a natural step to use the continuous wavelet transform as a vehicle to conduct the multiscale *analysis*, i.e. the multiscale decomposition, and the multiscale *synthesis*, i.e. the multiscale composition.

Let $\psi(x)$ be a complex valued function and define its dilatations as

$$\psi_\sigma(x) = |\sigma|^{-p}\psi(\frac{x}{\sigma}) \qquad \sigma \neq 0 \ \wedge \ p \geq 0, \tag{8.1}$$

where p is a suitable chosen normalization constant and σ is the *scale indicator*. The scale indicator rules the dilatations: the contractions, in case $\sigma < 1$, and stretchings, in case $\sigma > 1$, of the spatial variable x, see figure 8.1. Similarly the factor $|\sigma|^{-p}$ rules the dilatation in the vertical direction, the scaling of the dependent variable[2]. Choosing the exponent p is arbitrary and depends on which norm one wants to fix under the scale transformations. For example, by setting $p = \frac{1}{2}$ the $L^2(\mathbb{R})$ norm[3], $||\psi_\sigma||_2 = ||\psi||_2$, the "energy" of the wavelet, will be unaffected by the dilatations. A double-indexed family

[2]Notice the similarity with the scaling displayed by the homogeneous functions of order p as defined in chapter 5.

[3]This $L^2(\mathbb{R})$ norm will predominantly be used in this thesis since it expresses the energy.

of wavelets is formed via a combination of the affine scale transformations, as defined in equation (8.1), with translations along the spatial coordinate, $x \mapsto x - x'$, yielding

$$\psi_{\sigma,x'}(x) \triangleq |\sigma|^{-p}\psi(\frac{x-x'}{\sigma}) \qquad (x,x') \in \mathbb{R},\ \sigma \neq 0. \tag{8.2}$$

With this double-indexed family of wavelets one is able to obtain a decomposition of a functional[4] $f(x) \in L^2(\mathbb{R})$ onto its multiscale constituents via the transformation

$$\check{f}(\sigma,x') \triangleq \mathcal{W}\{f,\psi\}(\sigma,x') \triangleq \int_{-\infty}^{+\infty} f(x)\,\frac{1}{|\sigma|^p}\,\psi^*(\frac{x-x'}{\sigma})\,\mathrm{d}x. \tag{8.3}$$

This scale transformation was first introduced by Grossmann and Morlet (1985) and is known as the continuous wavelet transform. This scale transformation has, in order to allow for a reconstruction, to adhere to the following condition for the Fourier transform of the wavelet,

$$0 < C_\psi = \int_0^\infty \frac{|\hat{\psi}(k)|^2}{k}\mathrm{d}k = \int_{-\infty}^0 \frac{|\hat{\psi}(k)|^2}{|k|}\mathrm{d}k < \infty, \tag{8.4}$$

where k is the spatial frequency and $\hat{\psi}$ the Fourier transform of ψ. This condition is known as the so-called admissibility condition. For the class of wavelets being used in this thesis – namely those wavelets for which $x\psi(x) \in L^2(\mathbb{R})$, i.e. the wavelet ψ decays faster than x^{-1} for $x \to \pm\infty$, yielding continuity for $\hat{\psi}(k)$ – the admissibility condition boils down to at least demanding that the mean of the wavelet has to be zero, $\int_{-\infty}^{+\infty}\psi(x)dx = 0$, or equivalently its DC component has to vanish, i.e. $\hat{\psi}(0) = 0$. Remark that the zero mean requirement implies that the sign of the wavelet has to change at least once along the real line and it has been proclaimed that it is this property which coined the name wavelet referring to constructs that live in a Hilbert space but that still display wave-like behaviour within a limited spatial range.

By choosing the proper norm for the wavelet family, setting $p = \frac{1}{2}$, one can reformulate the continuous wavelet transform in terms of the inner product, defined as

$$\langle f,g\rangle \triangleq \int f(x)g^*(x)\mathrm{d}x, \tag{8.5}$$

acting on a Hilbert space, of f and $\psi_{\sigma,x'}$, hence

$$\mathcal{W}\{f,\psi\}(\sigma,x') \triangleq \langle f,\psi_{\sigma,x'}\rangle. \tag{8.6}$$

When a symmetric and real wavelet is used one can recognize the wavelet transform as a spatial convolution of the function f with the analyzing wavelet ψ_σ,

$$\mathcal{W}\{f,\psi\}(\sigma,x) \triangleq (f * \psi_\sigma)(x), \tag{8.7}$$

[4]By way of its construction the wavelet transform is able to host a larger class of objects than the conventional functions, the so-called generalized functions or distributions, see chapter 5 and the references therein.

where the $*$ refers to the convolution product.

It is clear that the scale indicator σ governs the effective support of the wavelet and as such allows for observations of the functional $f(x)$ at the different scales σ. The different wavelets act as a mathematical microscope zoomed in at scales σ. In figure 8.1 I included an example of shifted and dilated members of the double-indexed continuous wavelet family, in this case the second derivative of the Gaussian, the so-called Mexican hat.

By way of its construction the wavelet transform can easily be extended to host a larger class of objects, the so-called tempered distributions, and the reader is referred to chapter 5 and the references therein. What does this extension imply for the wavelet? Following Mallat and Hwang (1992), for example, one can demonstrate that when $f(x)$ is a tempered distribution of order n then one just has to invoke the condition of n times differentiability on the wavelet in order to regularize[5] the singular nature of the tempered distribution and equations (8.3), (8.6) and (8.7) will make sense. For the δ-distribution, a zero order tempered distribution, see section 5.8 of chapter 5, this procedure implies that one solely has to impose the condition of *continuity* on the analyzing wavelet. It is this property of the wavelet transform that allows one to replace ordinary functions by functionals as subject of investigation for the subsequent analysis.

At this point it is worthwhile to mention that the original functional $f(x)$ can be recovered from its space-scale representation via the inverse wavelet transform. This can be accomplished by resolving the identity, see (Daubechies, 1992) for the proof, which states that for every $f, g \in L^2(\mathbb{R})$, the following identity has to hold

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \check{f}(\sigma, x)\check{g}(\sigma, x)^* \sigma^{2p-3} \mathrm{d}\sigma \mathrm{d}x = C_\psi \langle f, g \rangle. \tag{8.8}$$

Here C_ψ is defined as in equation (8.4). This explains why the admissibility was required because in that case the resolution of identity will only makes sense. Define the inverse wavelet transform as follows

$$f(x) = \mathcal{W}^{-1}\{\check{f}, \psi\}(x) \triangleq \frac{1}{C_\psi} \int_0^{\infty} \int_{-\infty}^{+\infty} \check{f}(\sigma, x')\psi(\frac{x - x'}{\sigma})\sigma^{p-3} \mathrm{d}x' \mathrm{d}\sigma. \tag{8.9}$$

In the integral only the positive scales have to be considered because of the choice of using real wavelets only. Finally notice that the equation (8.9) only holds in the weak sense.

8.2.1 Properties of the continuous wavelet transform

In this section I will review the properties of the continuous wavelet transform. I commence by enumerating the basic properties of the continuous wavelet transform followed

[5] See chapter 5 for a theorem on regularization by a test function.

by a discussion on its localization capabilities yielding a decomposition of the space-spatial frequency plane. From now on I will, unless stated otherwise, work with the $L^2(\mathbb{R})$-norm for the wavelets, i.e. setting $p = \frac{1}{2}$.

Basic properties

The continuous wavelet transform satisfies the following basic properties:

- *Linearity*

$$\mathcal{W}\{\alpha f + \beta g, \psi\}(\sigma, x') = \alpha \mathcal{W}\{f, \psi\}(\sigma, x') + \beta \mathcal{W}\{g, \psi\}(\sigma, x'), \qquad \alpha, \beta \in C. \tag{8.10}$$

- *Shift-invariance*

$$\text{If } f(x) \to f(x - x'_0) \text{ then } \check{f}(\sigma, x') \to \check{f}(\sigma, x' - x'_0). \tag{8.11}$$

- *Dilation-invariance*

$$\text{If } f(x) \to \frac{1}{\sqrt{\sigma_0}} f(\frac{x}{\sigma_0}) \text{ then } \check{f}(\sigma, x') \to \check{f}(\frac{\sigma}{\sigma_0}, \frac{x'}{\sigma_0}), \qquad \sigma_0 > 0. \tag{8.12}$$

- *Conservation of energy*

$$\langle f, f \rangle = \int_{-\infty}^{+\infty} |f(x)|^2 \mathrm{d}x = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |\check{f}(\sigma, x')|^2 \frac{\mathrm{d}x' \mathrm{d}\sigma}{\sigma^2}. \tag{8.13}$$

This equation is analogous to Plancherel's equation.

The reproducing kernel

There is a danger hidden in the way the continuous wavelet transform has been set up. This danger lies in the fact that the continuous wavelets are not mutually orthogonal, a notion that becomes clear when one takes into consideration that the wavelet transform constitutes a map of a one-dimensional function to the two-dimensional space-scale plane. Therefore the space-scale plane is intrinsically redundant and the wavelet transform of every function $f(x) \in L^2(\mathbb{R})$ will only constitute a subset of all finite energy functions living in the space-scale plane. Consequently one can not ensure that the following map

$$\check{f}(\sigma, x') \xrightarrow{\mathcal{W}^{-1}} f(x) \xrightarrow{\mathcal{W}} \check{g}(\sigma, x') \tag{8.14}$$

will yield the identity $\check{g}(\sigma, x') = \check{f}(\sigma, x')$. Remark that this identity will generally only hold in case the wavelets used are orthogonal. On the other hand if one would start off

with an arbitrary function $f(x) \in L^2(\mathbb{R})$ then the use of equations (8.6) and (8.9) will always give rise to

$$f(x) \xrightarrow{\mathcal{W}} \check{f}(\sigma, x') \xrightarrow{\mathcal{W}^{-1}} f(x). \tag{8.15}$$

Now can one come up with a condition that guarantees whether a function in the space-scale plane constitutes a wavelet transform? To accomplish this one has to define the area of the space-scale plane that is inhabited by the wavelet transform of every $f(x) \in L^2(\mathbb{R})$. Grossmann et al. (1989) proves that this subset is exactly delineated by the reproducing kernel

$$\mathrm{K}_\psi(\sigma, x') = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \psi(x) \frac{1}{\sqrt{\sigma}} \psi^*(\frac{x - x'}{\sigma})\, \mathrm{d}x. \tag{8.16}$$

Given this subset one can come up with the required condition, stating that a function $\check{f}(\sigma, x')$ corresponds only to a continuous wavelet transform of f and with respect to ψ *if and only if* it satisfies for all points (σ_0, x_0') the equality

$$\check{f}(\sigma_0, x_0') = \int_0^{+\infty} \int_{-\infty}^{+\infty} \mathrm{K}_\psi(\frac{\sigma_0}{\sigma}, \frac{x_0' - x'}{\sigma})\, \check{f}(\sigma, x') \frac{\mathrm{d}x' \mathrm{d}\sigma}{\sigma^2}. \tag{8.17}$$

Localization properties in the space-frequency plane

Despite the non-orthogonality of the family of continuous wavelets it is still possible to acquire a satisfactory – within the limits imposed by the uncertainty relation for space and spatial frequency – decomposition of the space-spatial frequency plane. This decomposition is inspired by the fact that the reproducing kernel in equation (8.17) is only significantly different from zero when the point (σ, x') lies in the direct neighbourhood of (σ_0, x_0'). The spatial extent of this neighbourhood is, judged by the example depicted in figure 8.2, dependent on the value of the scale indicator σ. This observation is substantiated when one examines the continuous wavelet transform of the Dirac delta distribution as displayed in figure 8.3. Now let me go into more detail how the wavelet transform decomposes the space-spatial frequency plane.

First I need to define the central locations and widths for the wavelet, ψ, in the two domains. Define

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \langle x \rangle)^2\, |\psi(x)|^2 \mathrm{d}x, \tag{8.18}$$

expressing the spatial width and where $\langle x \rangle = x_0$ is the mean location given by

$$\langle x \rangle = \int_{-\infty}^{\infty} x\, |\psi(x)|^2 \mathrm{d}x. \tag{8.19}$$

I assumed the following normalization $\|\psi(x)\|_2 = \|\hat{\psi}(k)\|_2 = 1$.

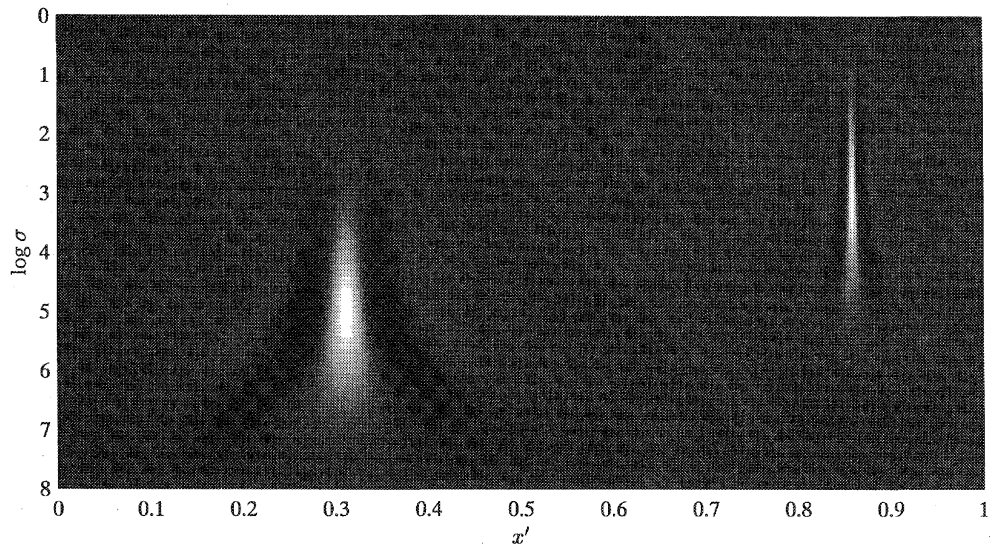

Figure 8.2 *The reproducing kernel* $\mathrm{K}_\psi(\sigma_0/\sigma, (x_0' - x')/\sigma)$ *for two different values of* (σ_0, x_0'). *Note the logarithmic axis for* σ.

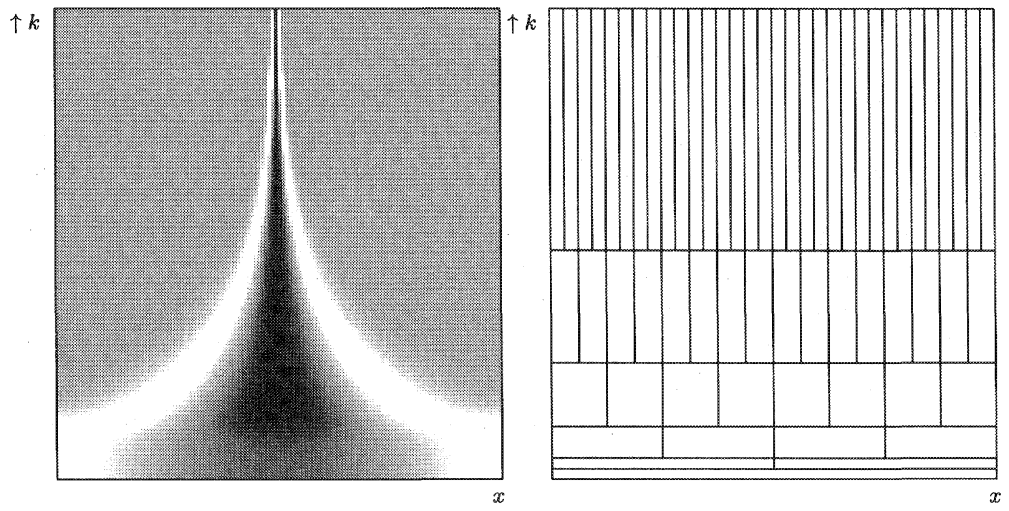

Figure 8.3 *(a) The increasing width of the analyzing functions of the continuous wavelet transform: at high spatial frequencies, the analyzing functions have small supports in the spatial domain and at the low frequencies the analyzing functions almost completely cover the spatial domain. (b) Space-spatial frequency decomposition via dyadic scale indicator increments, i.e.* $\sigma = \sigma_0^j \quad j \in \mathbb{Z} \wedge \sigma_0 = 2$. *The coupling between the width in the space and spatial frequency domain is evident.*

Similarly, the bandwidth in the spatial Fourier domain is given by

$$\sigma_k^2 = \int_{-\infty}^{\infty} (k - \langle k \rangle)^2 \, |\hat{\psi}(k)|^2 \mathrm{d}k, \tag{8.20}$$

where $\langle k \rangle = k_0$ is the mean location defined by

$$\langle k \rangle = \int_{-\infty}^{\infty} k \, |\hat{\psi}(k)|^2 \mathrm{d}k. \tag{8.21}$$

Not surprisingly these definitions are similar to the definitions for the averaged quantities, the mean location and duration and the mean frequency and bandwidth, as presented in chapter 3. Given these expressions one can come up with an estimate for the obtainable localization in both domains via

$$\sigma_x \sigma_k \geq \frac{1}{2}, \tag{8.22}$$

known as the uncertainty principle for space and spatial frequency as coined by Gabor (1946) for the time-frequency ambiguity. Notice that this condition implies that the area of the rectangles depicted in figure 8.3 is everywhere the same.

Using the above definitions the spatial localization corresponds to the interval

$$(-\frac{1}{2}\sigma_x, +\frac{1}{2}\sigma_x), \tag{8.23}$$

for a wavelet located at the origin while the localization in the spatial frequency domain extends over the interval

$$(k_0 - \frac{1}{2}\sigma_k, k_0 + \frac{1}{2}\sigma_k). \tag{8.24}$$

What happens with the localization if one dilates or translates the wavelet $\psi(x)$ to compose the double-indexed family of analyzing wavelets? The translation is rather simple to deal with and will be skipped. The dilation $\psi(x) \rightarrow \sigma_x^{-1/2}\psi(x/\sigma)$ will be dealt with here. In that case the localization in the space domain becomes

$$(-\frac{1}{2}\sigma\sigma_x, \frac{1}{2}\sigma\sigma_x) \tag{8.25}$$

whereas the localization in the spatial frequency domain reduces to

$$(\frac{k_0}{\sigma} - \frac{1}{2}\frac{\sigma_k}{\sigma}, \frac{k_0}{\sigma} + \frac{1}{2}\frac{\sigma_k}{\sigma}). \tag{8.26}$$

Equation (8.25) and (8.26) describe the space-spatial frequency behavior of the wavelet transform. One may now ask the following question. Which domain of the space-scale plane is influenced by the value of the signal $f(x)$ at an arbitrary spatial location $x = x_0$? Using equation (8.25) one sees that this domain is given by the cone

$$|x_0 - x| < \frac{1}{2}\sigma\sigma_x. \tag{8.27}$$

This cone is referred to as the cone of influence. Let me continue to dwell upon a number of interesting conclusions that can be drawn from these equations (8.25) - (8.27).

1. *The product of the widths of the wavelets σ_x and σ_k remains constant*, which is to be expected. A smaller localization in the space-domain corresponds to a larger localization in the spatial frequency domain. This illustrates exactly the zooming property of the wavelet transform: high frequency components can be analyzed at a better space resolution than lower frequency components.

2. *The translation parameter should be adjusted to the scale σ.* An observation of the cone of influence in the space domain is given in Figure 8.4. A small scale, i.e. a small value of σ, requires a fine sampling, whereas a coarse scale, i.e. a large value of σ, can deal with a coarser sampling.

3. *Constant Q-factor.* Viewing the wavelet transform as the output of a filterbank with bandfilters with a central frequency k_0/σ and bandwidth σ_k^2/σ, one sees that the Q(uality) factor of the filters is constant.

$$Q = \frac{\sigma_k/\sigma}{k_0/\sigma} = \frac{\sigma_k}{k_0} = C. \tag{8.28}$$

The constant behavior of the analyzing wavelets suggests a logarithmic behavior of the scale-axis, as in Figure 8.4 (see section 8.4). In room acoustics the choice $\sigma = 2^j$ corresponds to octave filtering.

Types of wavelets

Two characteristics are important for any family of wavelets either continuous or discrete. These properties refer to the number of vanishing moments possessed by the wavelet and the regularity of the wavelet. The first property is defined as:

Definition 8.1: Number of vanishing moments (Mallat and Hwang, 1992)
A wavelet is said to have M vanishing moments, if and only if for all positive integer $m < M$, it satisfies

$$\int_{-\infty}^{+\infty} x^m \psi(x) \mathrm{d}x = 0. \tag{8.29}$$

Definition 8.1 shows to which order of a polynomial the wavelets are orthogonal. I showed that the admissibility condition requires at least one vanishing moment a notion being reflected in the fact that the wavelet contains at least one zero-crossing. The number of zero-crossings increases for increasing number of vanishing moments M. The number of vanishing moments is directly linked to the regularity of the Fourier transform of the wavelet at the origin. Speaking about regularity, the second important property of

the wavelet concerns the differentiability and this corresponds to the decay rate of the Fourier coefficients as the frequency is being increased, i.e. as $k \to \infty$. As will be shown later the choice of the proper wavelet, choosing the right number of vanishing moments, constitutes a dominant factor in the process of setting up a proper multiscale analysis. In section 8.6 I will illustrate, by means of a number of examples, what can go wrong when erroneous wavelets are selected.

8.3 Local regularity analysis by the wavelet transform

The main advantage of using the continuous wavelet transform to measure the regularity of a functional lies in its localized focussing capability. This capability is in sharp contrast with the Fourier transform which is limited to offer information on the regularity of a functional, as showed in section 5.7.2 of chapter 5, in a delocalized way, i.e. the Fourier transform delocalizes, withstanding a regularity analysis of the local features. The continuous wavelet transform, on the other hand, is able to provide local regularity information, since the value of the wavelet coefficients depends on contributions from a confined region, from the support of the wavelet. In case of analysis via the Fourier transform, however, the Fourier coefficients depend on contributions from the whole real line, resulting in a decay of these coefficients which is dominated by the strongest singularity in the signal on the line.

In the pertaining sections an attempt is made to give the reader a concise overview of the main mathematical tools for the detection and measurements of singularities. At first a relationship will be established between the decay/growth rate of the wavelet coefficients and the *Hölder exponents* expressing the *degree of regularity.* Then the *wavelet transform modulus maxima* formalism will be introduced providing an optimal partitioning of the space-scale plane.

8.3.1 Regularity and the wavelet transform

In chapter 5 I already mentioned the existence of an intricate relationship between the *smoothness* of a functional and the growth/decay rate of its Fourier transform. Given this relationship it is possible to come up with estimates for the degree of (ir)regularity by taking the superior bound of all exponents α that satisfy the condition

$$\int_{-\infty}^{+\infty} |\hat{f}(k)|(1+|k|)^{\alpha} \mathrm{d}k < \infty. \tag{8.30}$$

In this way one is able to attribute a *global* – because the Fourier transform delocalizes – regularity estimate in terms of the Hölder or Lipschitz exponent α as introduced in section 5.9 of chapter 5. Take, for example for f the δ-distribution[6] then one finds that

[6]The reader is referred to chapter 5 for a detailed discussion on the distribution theory.

equation (8.30) holds for Hölder exponents $\alpha < -1$. By taking the superior bound of all these exponents one finds $\alpha = -1$ for the δ-distribution. Besides the disadvantage of delocalization, allowing for a global estimate only, another disadvantage becomes apparent in the sense that the condition of equation (8.30) only provides a sufficient but not a necessary condition.

Following Mallat and Hwang (1992) it is possible to come with an alternative condition for the wavelet coefficients that is sufficient as well as necessary(Jaffard, 1989; Holschneider and Tchamitchian, 1990; Daubechies, 1992). This is possible because the wavelet transform provides local information, while zooming in, $\sigma \downarrow 0$, from which one can infer information on the local regularity. To show this let me first restate a theorem that comes up with a global estimate followed by a theorem enabling the local assessment of the regularity.

Theorem 8.1: Wavelet transform and global regularity (Jaffard, 1989; Holschneider and Tchamitchian, 1990; Daubechies, 1992; Mallat and Hwang, 1992)
Let $f(x) \in L^2(\mathbb{R})$ and $[a,b]$ be an interval of $\mathbb{R}$. Let the wavelet ψ adhere to the admissibility condition. Let $0 < \alpha < 1$. For any $\epsilon > 0$, $f(x)$, is uniform/global Hölder α over $(a+\epsilon, b-\epsilon)$, if and only if there exists a constant C such that for $x \in (a+\epsilon, b-\epsilon)$ and for $\sigma > 0$

$$|\mathcal{W}\{f,\psi\}(\sigma,x)| \leq C\sigma^{\alpha+1/2}. \tag{8.31}$$

The conditions laid down in equations (8.30) and (8.31) both provide a way to access the global regularity. For the Fourier transform the sufficient condition implies that $|\hat{f}(k)|$ decays 'faster' than $k^{-\alpha}$ as $|k| \to \infty$ (Mallat and Hwang, 1992), whereas the condition in equation (8.31) acts in a similar fashion. This is understood when the scale is locally considered as the reciprocal of the frequency, i.e. $k \sim \sigma^{-1}$. The condition is now sufficient and necessary. Besides this advantage the estimates are localized on intervals $(a+\epsilon, b-\epsilon)$ rather than over the whole real line. This latter property can be attributed to the fact that the wavelet coefficients depend on the evaluation over a finite interval depending on the scale indicator. To this interval one can associate a cone whose support diminishes when the wavelet zooms in. In this way Jaffard (1989) was able to come up with a proof of a theorem that allows one to come up with local estimates for the regularity at a point x. Following Mallat and Hwang (1992) I will use this theorem without proving it.

Theorem 8.2: Wavelet transform and local Hölder exponents (Jaffard, 1989; Mallat and Hwang, 1992)
Suppose $\psi(x)$ has M vanishing moments and is M times differentiable. Let $f(x) \in L^2(\mathbb{R})$ and let $\alpha \leq M$. If $f(x)$ is Hölder α at x_0, then there exists a constant A, such that for all x in a neighbourhood of x_0 and for any scale σ,

$$|\mathcal{W}\{f,\psi\}(x,\sigma)| \leq A\sigma^{\frac{1}{2}}(\sigma^{\alpha} + |x - x_0|^{\alpha}). \tag{8.32}$$

Conversely, let $\alpha < M$ be a non-integer value. The functional $f(x)$ is Hölder α at x_0, if the following two conditions hold:

- *there exists some $\epsilon > 0$ and a constant A, such that for all points x in a neighbourhood of x_0, and any scale σ,*

$$|\mathcal{W}\{f,\psi\}(x,\sigma)| \leq A\sigma^{\epsilon+\frac{1}{2}}. \tag{8.33}$$

- *there exists a constant B such that for all points x in a neighbourhood of x_0 and any scale σ*

$$|\mathcal{W}\{f,\psi\}(x,\sigma)| \leq B\sigma^{p}\left(\sigma^{\alpha} + \frac{|x-x_0|}{|\log|x-x_0||}\right). \tag{8.34}$$

In this theorem the admissible range of positive Hölder exponents α has been expanded from $0 < \alpha < 1$ to $0 < \alpha < M$ by imposing an additional constraint on the number of vanishing moments, see definition 8.1, for the analyzing wavelet.

Given this definition together with definition 5.2, see chapter 5, for the regularity, it is possible to extend the range of detectable positive Hölder exponents by taking into consideration that a wavelet with M vanishing moments can be written as the $(M-1)^{\text{th}}$-order derivative of a wavelet with one vanishing moment, $\psi(x) = \psi^{(1)}(x)$. This latter wavelet is defined as the first derivative of a smoothing function $\phi(x)$, according to $\psi^{(1)}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\phi(x)$. So

$$\psi^{(M)}(x) = \frac{\mathrm{d}^{M-1}}{\mathrm{d}^{M-1}x}\psi^{(1)}(x) \tag{8.35}$$

and therefore

$$\mathcal{W}\{f,\psi^{(M)}\}(\sigma,x) = (-\sigma)^{(M-1)}\mathcal{W}\{\frac{\mathrm{d}^{M-1}}{\mathrm{d}x^{M-1}}f,\psi(1)\}(\sigma,x). \tag{8.36}$$

This means that wavelet transforming a function $f(x)$ with respect to a wavelet $\psi(x)$ with M vanishing moments corresponds to taking the wavelet transform of the $(M-1)^{\text{th}}$-order derivative of f with respect to the wavelet $\psi^{(1)}$, or to taking the *smoothed* version of the M^{th}-order derivative of f, i.e.

$$\mathcal{W}\{f,\psi\}(\sigma,x) = (-\sigma)^{(M)}(\frac{\mathrm{d}^{M}}{\mathrm{d}x^{M}}f * \phi)(\sigma,x). \tag{8.37}$$

Here $\psi(x)$ denotes a wavelet with M vanishing moments and from here on the superscript M will be dropped when no confusion arises. Take $0 < \alpha - (M-1) < 1$ and use definition 5.3 of chapter 5 which stated that the function f is global Hölder α on the open interval (a,b) if and only if $\frac{\mathrm{d}^{M-1}}{\mathrm{d}x^{M-1}}f(x)$ is uniform Hölder $\alpha-(M-1)$ on the same interval. Then one finds, by using theorem 8.1, that

$$|\mathcal{W}\{\frac{\mathrm{d}^{M-1}}{\mathrm{d}x^{M-1}}f,\psi^{1}\}(\sigma,x)| \leq C\sigma^{\alpha-(M-1)+\frac{1}{2}}, \tag{8.38}$$

where C is an arbitrary constant and where $\psi^{(1)}$ represents the wavelet with $M = 1$. Substitution of equation (8.36) into equation (8.38) and choosing $\psi(x)$ as the $(M-1)^{\text{th}}$-order derivative of $\psi^1(x)$ one finds

$$|\mathcal{W}\{f,\psi\}(\sigma,x)| \leq C\sigma^{\alpha+\frac{1}{2}}. \tag{8.39}$$

This equation extends the effective range of permissible Hölder exponents from $0 < \alpha < 1$ in equations (8.31) and (8.33) to $0 < \alpha < M$. In this way one can come up with proper regularity estimates referring to the $(M-1)^{\text{th}}$-order derivatives of f.

Another way of explaining the condition on the number of vanishing moments is the apparent relationship between this number and the regularity of the Fourier transform at the origin. It is easy to show that the condition for a wavelet to have M vanishing moments is equivalent to demanding a $(M-1)$-fold zero or $(M-1)$-order regularity at the origin, i.e. $\{\partial_k^{M-1}\hat{\psi}(k)\}_{k=0}$. In that way the Fourier transform of the function, f, is allowed to contain a $\mathcal{O}(k^{-M})$ singularity at $k = 0$ which is then regularized, in the sense of Hadamard's finite part, see section 5.4.1 of chapter 5 by the $\mathcal{O}(k^M)$ regularity of the wavelet.

The main asset of theorem 8.2 is that it allows for the determination of the local Hölder exponent at one isolated point. Notice, however, that this only goes at the expense of trading off the sufficient or necessary condition for a necessary and sufficient condition as can be seen from theorem 8.1. Moreover, from theorem 8.2 and equation (8.33) one deduces that f is uniform, global Hölder ϵ in some neighbourhood of x_0 with ϵ being arbitrary small. To interpret the other two conditions depicted in theorem 8.2 and given by equations (8.33) and (8.34) one has distinguish between regions of the scale-space plane that lie in or outside the cone

$$|x - x_0| \leq \frac{1}{2}\sigma_x\sigma. \tag{8.40}$$

For the part of the scale-space plane lying inside the cone the $|\mathcal{W}\{f,\psi\}(\sigma,x)|$ behave for $\sigma \downarrow 0$ as $\mathcal{O}(\sigma^{\alpha+\frac{1}{2}})$, see equations (8.32) and (8.34). Outside this cone the necessary and sufficient conditions have different bounds as $\sigma \downarrow 0$ and in that situation the value for $|\mathcal{W}\{f,\psi\}(\sigma,x)|$ is determined by the distance of x with respect to x_0 and has to be treated separately for equations (8.32) and (8.34).

To summarize, theorem 8.1 supplemented with the condition on the number of vanishing moments allows one to issue statements on the global regularity of the $(M-1)^{\text{th}}$-order derivative of f via inspection of the asymptotic decay of the wavelet coefficients $\mathcal{W}\{f,\psi\}(\sigma,x)$. However, it does not allow one to draw conclusions on the regularity of the M^{th} derivative of f. The same can be said about the capability of theorem 8.2 but now it provides a local estimate. Now what happens when a infinitely differentiable functions is being considered? Take for example $f(x) = \cos(x)$. If ψ has exactly M vanishing moments, it can be shown that $\mathcal{W}\{\cos,\psi\}(\sigma,x)$ decays as $\sigma^{M+\frac{1}{2}}$ on any interval

(a,b). But be careful this does not necessarily imply that $\cos(x)$ is of Hölder $\alpha = M$ on any interval (a,b). Since $\cos(x)$ is infinitely continuously differentiable, $\cos(x) \in C^{\infty}$, on any interval (a,b), it has a global Hölder exponent $\alpha = \infty$ on any interval (a,b). When taking the wavelet transform of the function f with respect to a wavelet with M vanishing moments results in an estimate for the Hölder exponent equal to the number of vanishing moments M then one may safely conclude that the function is at least Hölder $\alpha = M$.

It was shown that by increasing the number of vanishing moments it is possible to extend the range of detectable algebraic singularities in the direction of the positive Hölder exponents. Can this argument be "reversed" towards constructs with negative Hölder exponents, i.e. $\alpha < 0$? Yes, this can be done by choosing the wavelet smooth enough, a notion corresponding to demanding a fast enough decay for the Fourier coefficients as $k \to \infty$, and by making again use of definition 5.3 of chapter 5. It follows that in case the function $f(x)$ is global Hölder $-k-1 < \alpha < -k$, the wavelet must be at least be k times differentiable, i.e. C^k function.

Remark that for a singular distribution, a generalized function with only one isolated negative singularity, e.g the δ-distribution, the wavelet only requires one vanishing moment because of the admissibility condition[7].

Finally, theorems 8.1 and 8.2 prove that the wavelet transform is a useful tool to perform (localized) regularity estimations. From the numerical point of view, however, both theorems cause difficulties since they require, in case of isolated singularities, the evaluation of a two-dimensional area defined by the cone $|x - x_0| \leq \frac{1}{2}\sigma_x \sigma$ "surrounding" x_0. In the next section an alternative approach will be reviewed. This alternative is based on an effective partitioning of the scale-space plane in terms of the extrema for the modulus of the continuous wavelet transform and thereby reducing the number of calculations, which are required to estimate the local regularity.

8.4 Measuring the Hölder regularity with the wavelet transform

In chapter 5 I already mentioned that the information in a measured signal or a mathematical object, such as a distribution, is carried by its *singularities*. These singularities are also known as *essential* points or as the regions of rapid variation. The observation that this information is confined to these localized "areas" naturally finds its way into the wavelet framework where the singular behaviour is directly reflected into the growth/decay rate for the wavelet coefficients. Why is this the case? The answer lies in the theory presented in chapter 5 where the class of algebraic singularities, singular distributions containing an isolated algebraic singularity lying in a smooth environment,

[7]Notice that for assessing the regularity for negative Hölder exponents one does not require vanishing moments, i.e. it suffices to compute $\int_{-\infty}^{+\infty} f(x)\phi(\frac{x-x'}{\sigma})\mathrm{d}x$, where ϕ stands for a smoothing kernel. However, if one wants to reconstruct the function the admissibility condition must be supplemented.

was shown to display an unboundedness for derivatives exceeding a certain finite order.

Judging by the fact that the wavelet transform, conducted with respect to a wavelet with M vanishing moments, approximates the M^{th}-order derivative, it can readily be understood that the wavelet transform triggers on the singularities. With other words the wavelet transform can be compared by taking the derivative in the sense of distributions and that is the reason why the non-differentiable singularities show up as *maxima* for the *modulus* of the wavelet transform. It is exactly this perception that probably inspired Mallat and Hwang (1992) to introduce the wavelet transform modulus maxima formalism. This formalism makes explicitly use of the property that a local irregularity maps to the extrema for the modulus of the wavelet transform across the different scales. In this way the extrema provide an optimal multiscale partitioning of the two-dimensional scale-space plane.

In this way an effective covering of the signal's complexity is accomplished and this is done by the so-called *wavelet transform modulus maxima lines*, the WTMML's. These lines interconnect the maxima for the modulus of the wavelet transform within the cone of influence and across the scale-space plane. They emanate from the abscissa where the singularities are located and proceed into the scale-space plane, in the direction of the coarser scales, until they possibly arrive at a bifurcation. This bifurcation indicates the point in the scale-space where the cones of influence of two WTMML's start to overlap.

Given the effective partitioning the WTMML's provide an efficient measurement procedure – via the inspection of the amplitudes along the locations delineated by the WTMML's – of the local Hölder exponents to be associated with the singularities detected by the wavelet transform.

To summarize, Mallat and Hwang (1992) have shown that the WTMML representation is adequate to estimate the local regularity by evaluating the asymptotic behaviour for the amplitudes along the WTMML's. Moreover they demonstrated that these WTMML's provide sufficient information to non-uniquely[8] reconstruct the original function $f(x)$ from the WTMML partitioning[9]. The treatise is commenced by providing a theorem that assigns a WTMML to a non-oscillatory singularity. Then a theorem is reviewed enabling the measurement of Hölder exponent estimates for *isolated* non-oscillatory singularities followed by a theorem paving the way for the detection and subsequent measurement of non-isolated non-oscillatory singularities.

8.4.1 Partitioning by the WTMML

By definition a modulus maximum of the continuous wavelet transform is a strict local maximum of the modulus at a fixed scale σ_0 and either on the left- or right-hand side of

[8] See Meyer (1990).

[9] See the master's thesis of Hoekstra (1996) who beautifully shows that the WTMML's can be used to reconstruct one- and two-dimensional objects.

x_0. More formally, following Mallat and Hwang (1992):

Definition 8.2: Wavelet transform modulus maxima (Mallat and Hwang, 1992)
Let $\mathcal{W}\{f,\psi\}(\sigma,x)$ be the wavelet transform of a function $f(x)$.

- *A local extremum is any point (σ_0,x_0) for which $\partial_x\mathcal{W}\{f,\psi\}(\sigma,x)$ has a zero-crossing at $x=x_0$, when x varies.*
- *Call a wavelet transform modulus maximum, a* WTMM, *any point (σ_0,x_0) such that $|\mathcal{W}\{f,\psi\}(\sigma_0,x)|<|\mathcal{W}\{f,\psi\}(\sigma_0,x_0)|$ when x belongs to either the right or the left neighbourhood of x_0, and $|\mathcal{W}\{f,\psi\}(\sigma_0,x)|\leq|\mathcal{W}\{f,\psi\}(\sigma_0,x_0)|$ when x belongs to the other side of the neighbourhood of x_0.*
- *Call a wavelet transform modulus maxima line, a* WTMML, *any connected curve in the scale space (σ,x) along which all points are modulus maxima.*

Let me now present the first in a series of three theorems presented by Mallat and Hwang (1992) and enabling the characterization of algebraic singularities. For the proof the reader is referred to their work.

Theorem 8.3: Singularity detection (Mallat and Hwang, 1992)
Let M be a strict positive integer. Let ψ_σ be a wavelet of compact support, M vanishing moments and M times continuously differentiable. Let $f(x)\in L^1([a,b])$.

- *If there exists a scale $\sigma_0>0$ such that for all scales $\sigma<\sigma_0$ and $x\in(a,b)$, $|\mathcal{W}\{f,\psi\}(\sigma,x)|$ has no local maxima, then for any $\epsilon>0$ and $\alpha<M$, $f(x)$ is uniformly Hölder α in $(a+\epsilon,b-\epsilon)$.*
- *If $\psi(x)$ is the M^{th} derivative of a smoothing function, then $f(x)$ is uniformly Hölder M on any such interval $(a+\epsilon,b-\epsilon)$.*

The crux of this theorem lies in the fact that it proves that a function is global Hölder α with $\alpha<M$ in the closed interval (a,b) when there is no modulus maximum emerging in the wavelet transform as the scale indicator is decreased, $\sigma\downarrow 0$. Reversing the argument: a WTMML points to a singularity, is not necessary true in cases where the functional contains oscillatory regions. However, when the functional is singular or singular in one of its derivatives then a WTMML emerges under the condition of using the proper analyzing wavelet. In this case the WTMML points, as the scale is decreased, to the abscissa where the functional is singular. It is this localization property – the WTMML zooms in on the singularity – that makes it possible to come up with a theorem providing estimates for the Hölder exponent to be associated with an isolated singularity. Of course the WTMML's will guide the analysis in the scale-space plane.

8.4.2 Isolated singularities

When a functional contains a single non-oscillating singularity it is possible to come up with a measurement procedure aimed at capturing the Hölder exponent. The analyzing wavelet is assumed to have M vanishing moments and to be M times differentiable.

Theorem 8.4: Measuring isolated singularities (Mallat and Hwang, 1992)
Let $f(x)$ be a tempered distribution[10] *whose wavelet transform is well defined over (a,b) and let $x_0 \in (a,b)$. Suppose there exists a scale $\sigma_0 > 0$, and a constant C, such that for $x \in (a,b)$ and $\sigma < \sigma_0$, all the modulus maxima of $\mathcal{W}\{f,\psi\}(\sigma,x)$ belong to the cone defined by*

$$|x - x_0| \leq C\sigma. \tag{8.41}$$

Then, at all points $x_1 \in (a,b)$, $x_1 \neq x_0$, $f(x)$ is uniformly Lipschitz M in a neighbourhood of x_1. Let $\alpha < M$ be a non-integer. The function $f(x)$ is Hölder α at x_0, if and only if there exists a constant A such that each modulus maxima (σ,x) in the cone defined in equation (8.41)

$$|\mathcal{W}\{f,\psi\}(\sigma,x)| \leq A\sigma^{\alpha}. \tag{8.42}$$

The proof of this theorem can be found in (Mallat and Hwang, 1992). The scaling behaviour of the WTMML as depicted in equation (8.42) amounts in the log scale log amplitude space to an expression equivalent to,

$$\log|\mathcal{W}\{\sigma,x\}| \leq \log A + \alpha \log \sigma. \tag{8.43}$$

This equation, together with the conditions laid down in theorem 8.4, permit the estimation of the singularity strength by taking the supremum for the slope obtained after conducting a linear regression.

8.4.3 Non-isolated singularities

As a final step in the local characterization of the regularity of a functional it is of interest to consider functionals containing non-isolated and non-oscillatory singularities. Why is this type of configuration important? Because the objects of interest in this thesis can be singular everywhere on their support. Think for example of the mono- and/or multifractal measures presented in chapters 6 and 7. Keep, however, in mind that this latter category of highly complicated signals make it difficult to assign a Hölder exponent to a singularity over a wide scale range. What it does is that it gives an effective estimate for the Hölder valid over a limited scale range. That is to say that it expresses the notion that the function within that scale range can not be discerned from being Hölder so and so.

[10] See chapter 5 where I define the class of tempered distributions.

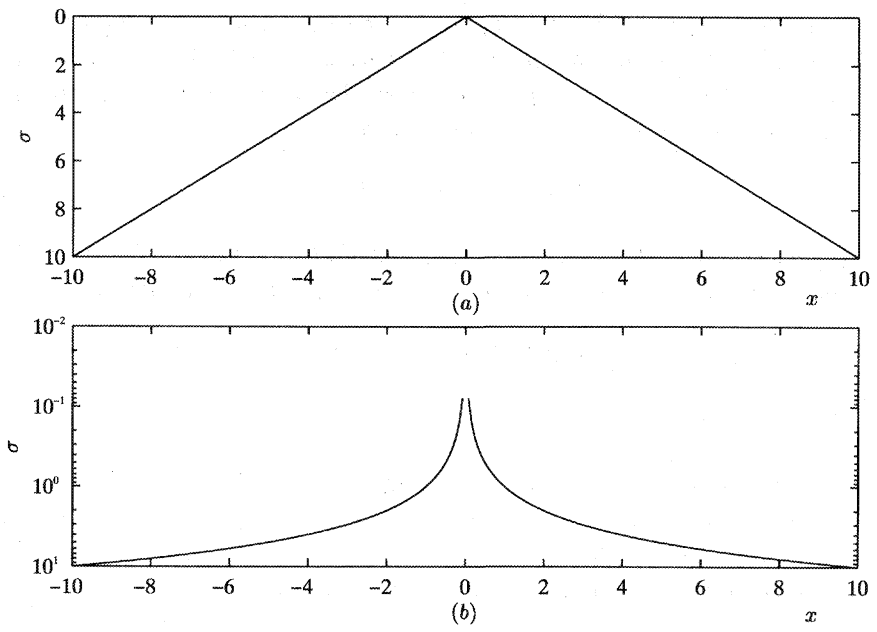

Figure 8.4 *The cone of influence for a linear σ-axis (a) and a logarithmic σ-axis (b).*

To measure the singularity strength of non-isolated singularities (Mallat and Hwang, 1992), also known as incipient singularities (Schertzer and Lovejoy, 1987a), it is necessary to introduce an additional concept, namely a *cone of influence*. Suppose a wavelet with a symmetric support on the closed interval $[-K, K]$ then the cone of influence in the scale-space plane, is the set of points (σ, x) that satisfy

$$|x - x_0| \leq K\sigma. \tag{8.44}$$

It is this cone, delineating the points in the scale-space plane where the values of the $\mathcal{W}\{f, \psi\}(\sigma, x)$ are influenced by the value of f at abscissa x_0, see figure 8.4. Now if one assumes that the functional or signal does not contain oscillatory singularities than one can characterize the regularity of f at a point x_0 by examining the behaviour of any line in the scale-space plane that belongs to a cone that is strictly smaller than the cone of influence defined in equation (8.44). Suppose that the analyzing wavelet is real, is M times continuous differentiable and has a symmetric support confined to the interval $[-K, K]$. Furthermore the wavelet is assumed to be the M^{th}-order derivative of a strict positive smoothing function $\phi(x)$ having the same support on the same closed interval $[-K, K]$.

Theorem 8.5: Measuring non-isolated singularities
Let $x_0 \in \mathbb{R}$ and let $f(x) \in L^2(\mathbb{R})$. We suppose that there exists a neighbourhood (a,b) of x_0 and a scale $\sigma_0 > 0$, such that the wavelet transform $\mathcal{W}\{f,\psi\}(\sigma,x)$ has a constant sign for $\sigma < \sigma_0$ and $x \in (a,b)$. Suppose now also that there exists a constant B and $\epsilon > 0$ such that for all points $x \in (a,b)$ and any scale σ

$$|\mathcal{W}\{f,\psi\}(\sigma,x)| \leq B\sigma^{\epsilon}. \tag{8.45}$$

Let $x = X(\sigma)$ be a curve in the scale space (σ,x) such that $|x_0 - X(\sigma)| \leq C\sigma$, with $C < K$. If there exists a constant A such that for any scale $\sigma < \sigma_0$, the wavelet transform satisfies

$$|\mathcal{W}\{f,\psi\}(\sigma,X(\sigma))| \leq A\sigma^{\gamma}, \qquad \text{with } 0 \leq \gamma \leq N, \tag{8.46}$$

then $f(x)$ is Hölder α at x_0, for any $\alpha < \gamma$.

The proof of this theorem can be found in Appendix C of Mallat and Hwang (1992). This theorem proves that the regularity of f is governed by the behaviour of the wavelet transform inside the cone of influence only if the M^{th}-order derivative does not have a oscillatory behaviour as x approaches x_0, e.g. $f(x) \neq \sin((x-x_0)^{-1})$.

As for theorem 8.4, valid for isolated singularities, theorem 8.5 can be used to find estimates for the Hölder regularity of singularities by inspection of the WTMML's. Again this is done by examining the amplitudes of $|\mathcal{W}\{f,\psi\}(\sigma,x)|$ along the WTMML's that converge to the singular "points" as the resolution is decreased, i.e. $\sigma \downarrow 0$.

For a discussion on the practical aspects, such as the selection of the proper family of analyzing wavelets, the reader is referred to chapter 2 where also attention is paid to placing the multiscale analysis and characterization techniques, presented so far, in a more general context. The actual application of this concept to well-log measurements will also be treated in this chapter.

8.4.4 Measuring procedure for the local multiscale analysis

In this section the measurement procedures for the estimation of the Hölder regularity of signals containing isolated as well as non-isolated singularities is reviewed. The prescription for the procedure is given followed by the actual application of the formalism in a number of examples.

Measuring the local Hölder exponent for an isolated singularity

The following procedure is proposed to measure the local Hölder exponent for an isolated singularity.

Procedure 8.1: Measurement and detection of isolated singularities
The procedure to detect and subsequently measure the regularity of functionals con-

taining an isolated and non-oscillatory singularity with Hölder estimates consists of the following steps:

1. *Select the proper family analyzing wavelet in accordance with the conditions summarized in the theorems dealing with the detection and measurement of isolated non-oscillatory singularities.*

2. *Compute the continuous wavelet transform of the signal f with respect to this family of analyzing wavelets, see figure 8.5 (b) middle row, using equation* (8.3).

3. *Locate the* WTMML*'s in accordance to definition 8.2. Perform the* WTMML *partitioning: that is create a set $L = \{1, 2, \ldots, l\}$ of l curves, parameterized by $\{X(\sigma)\}_{m \in L}$ on which the modulus of the continuous wavelet transform is maximum.*

4. *Study the behaviour of the function,*

$$\mathcal{W}\{f, \psi\}(\sigma, X_m(\sigma)), \tag{8.47}$$

on the curves parameterized by $X_m(\sigma)$ belonging to the set L, with an emphasis on the behaviour for σ approaching the resolution at which the measurement has been taken.

5. *Use the linear relationship in equation* (8.43) *to find the maximum slope of the straight lines that remain above the logarithm of the amplitude of the modulus maxima line, on a logarithmic scale, see figure 8.5 (c), bottom row.*

To elucidate the local multiscale measurement procedure I include a number of examples in Figure 8.5. The functional analyzed is of the following form

$$f_+^\alpha = \begin{cases} \dfrac{x_+^\alpha}{\Gamma(\alpha+1)} & \forall \alpha \neq l \in \mathbb{Z} \\ \delta^{(-l)} & \forall \alpha = l \in \mathbb{Z}, \end{cases} \tag{8.48}$$

where $\Gamma(\cdot)$ refers to the Gamma function and $\delta^{(l)}$ to the l^{th} distributional derivative of the delta distribution. This type of functionals are known as generalized functions with algebraic singularities or as singular distributions generated by Hadamard's 'finite part' of an algebraically divergent, when $\alpha < -1$, integral (Gel'fand and Shilov, 1964; Zemanian, 1965). In case α becomes integer the distribution f_+^α becomes the l^{th} primitive, for $l > 0$, or the $|l|^{\text{th}}$ derivative, $l < 0$, of the δ-distribution all taken in a distributional sense. For the non-integer values of α this distribution's action corresponds to fractional differentiators or integrators again depending on the sign of α. It is this property that explains why the differentiability and integrability conditions are so important for the analyzing wavelet. This has to be understood in the sense that the multiscale analysis is

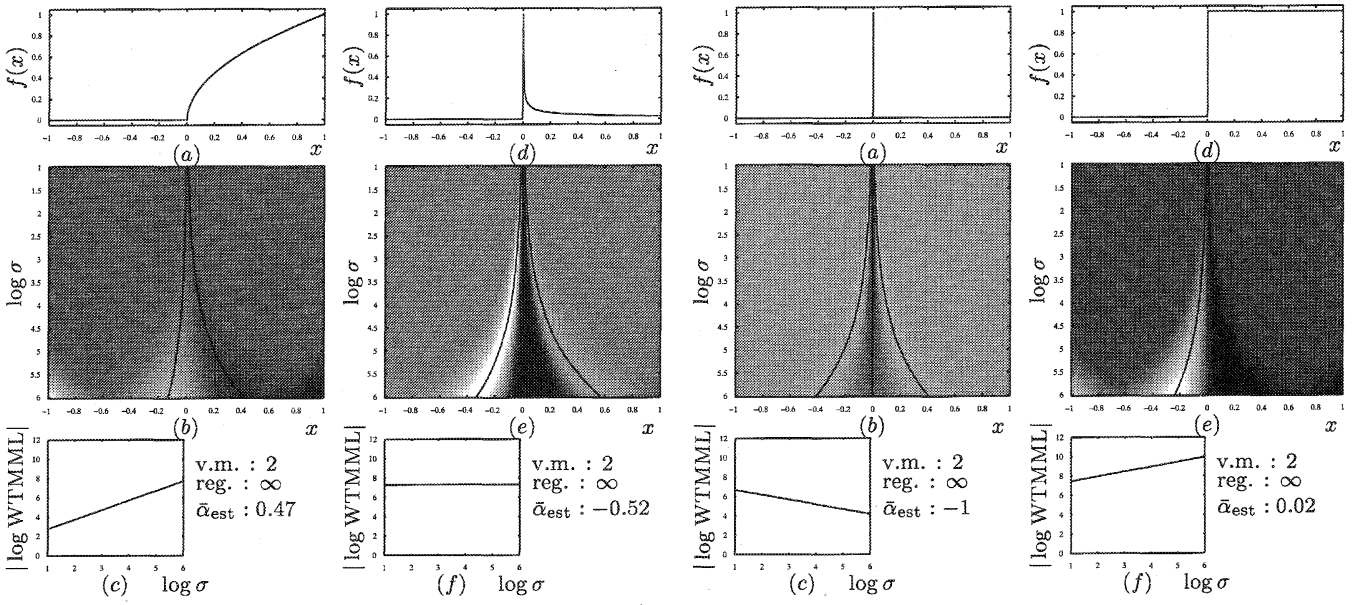

Figure 8.5 *This figure contains the regularity analysis conducted on distributions defined in terms of equation* (8.48) *with the Hölder exponent* α *set to the values* $\alpha = -0.5, 0.5, -1, 0$. *In all plots:* (b, e) *The wavelet transform of* (a, d) *viewed from above. The lines indicate the location of the modulus maxima.* (c, f) *The logarithm of the modulus maxima lines against* $\log\sigma$. *The slopes together with the abscissa of the lines for the small scales indicate that the distributions depicted in the* (a, d) *have local Hölder exponents* α *that closely match the values with which the singular distributions have been constructed.*

doomed to fail in those circumstances where the analyzing wavelet is unable to "absorb" the differential or integral action effectuated by the above distribution.

From the examples it becomes clear that the proposed method is able to capture the Hölder exponent accurately. It is also clear that this kind of singularities act as (fractional) differentiators annex integrators, a notion that becomes clear when the generalized functions with $\alpha \geq 0$ are being considered because in that case the number of WTMML decreases by one due to the (fractional) integration.

Measuring the Hölder exponent for non-isolated singularities

Now let me continue the discussion by proposing the following procedure to measure the local Hölder exponent for non-isolated singularities.

Procedure 8.2: Measurement and detection of non-isolated singularities
The procedure to detect and subsequently measure the regularity of functionals containing non-isolated and non-oscillatory singularities with Hölder estimates consists of the following steps:

1. *Select the proper family analyzing wavelet in accordance with the conditions summarized in the theorems dealing with the detection and measurement of isolated non-oscillatory singularities.*

2. *Compute the continuous wavelet transform of the signal f with respect to this family of analyzing wavelets, see figure 8.6, using equation* (8.3).

3. *Locate the* WTMML*'s in accordance to definition 8.2. Perform the* WTMML *partitioning: that is create a set $L = \{1, 2, \ldots, l\}$ of l curves, parameterized by $\{X(\sigma)\}_{m \in L}$ on which the modulus of the continuous wavelet transform is maximum, i.e. apply definition 8.2.*

4. *Check whether these* WTMML*'s lie inside the cone of influence defined in equation* (8.44) *and depicted in figure 8.4.*

5. *Study the behaviour of the function,*

$$\mathcal{W}\{f, \psi\}(\sigma, x), \tag{8.49}$$

on the curves parameterized by $X_m(\sigma)$ and belonging to the set L, with an emphasis on the behaviour for σ approaching the resolution at which the measurement has been taken.

6. *Use the linear relationship in equation* (8.43) *to find the maximum slope of the straight lines that remain above the logarithm of the amplitude of the modulus maxima line, on a logarithmic scale, see figure 8.6.*

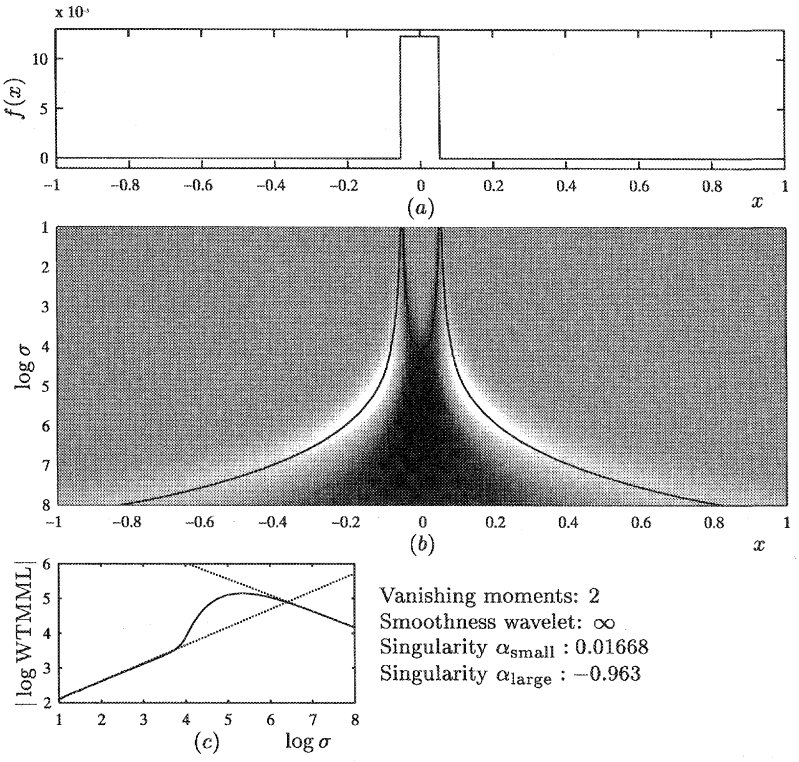

Figure 8.6 *This example illustrates the local scaling characteristics of two non-isolated singularities, discontinuities in this case. Clearly for the small scale range Hölder exponents are found consistent with the discontinuities. But as soon as the cones of influence start to overlap a bifurcation occurs having a drastic impact on the amplitude behaviour along the middle and to a lesser extent for the outer* WTMML*'s.*

For obvious reasons one is generally more interested in finding local estimates for the Hölder exponents yielded by signals containing more than one singularity. In that case there will always be a scale range within which the WTMML's start to interfere with each other. With other words the cones of influence will start to overlap and this is indicated by a bifurcation of the WTMML. Despite the apparent mutual influence by the singularities it is still possible to come up with Hölder exponent estimates. This, however, requires a rather technical theorem, and for this reason I will limit myself to reviewing the simple example of a box-car only. In Figure 8.6 I have included this box-car which can, within the seismic context, be associated to a 'thin' layer sandwiched between two homogeneous halfspaces. Inspection of the WTMML yielded by this box-car clearly shows that two scale ranges, separated by a bifurcation, can be distinguished. These two scale

ranges refer, on the one hand, to a scale range acting on scales smaller than the spatial extent of the box-car, whereas on the other hand to a scale range exceeding the support of the box-car. Inspection of the amplitude behaviour of the WTMML's over the first scale range clearly shows that the estimates for the Hölder exponent correspond to a Hölder exponent for an isolated discontinuity, i.e. $\alpha = 0$. But as soon as the WTMML's start to overlap, as soon as the bifurcation occurs, a completely different behaviour is evidenced. The Hölder exponent to be associated to the decay rate of the WTMML equals, in this scale range, $\alpha = -1$. This means that for this scale range the box-car, layer, acts as a functional that can not be discerned form a δ-distribution for the scales exceeding the width of the box-car.

The lesson to be learned from this example is that the signal, the functional, can display a different scaling behaviour for different scale ranges. However, this does not mean that the concept of parameterization by the Hölder exponents looses its meaning. It just tells one that care should be taken with respect to the scale range one is interested in.

8.5 Measuring the singularity spectrum f(α) by the continuous wavelet transform

Given the successful development and numerical implementation of the WTMML formalism to estimate the *local regularity* of signals, functionals containing (non)isolated algebraic singularities suggest the application of this formalism to the multifractal framework traditionally dominated by measure theory. This choice to unravel the complex scaling hierarchy of singularities, emanating from the multiplicative cascades[11] by means of wavelets acting as "smart boxes" opens the way to extend the theory for measuring the multifractal singularity spectrum f(α) to the realm of singular functionals rather than singular measures only. Besides this advantage – the *global* multiscale analysis procedure becomes also insensitive to *regular* perturbations of the singular measure – the WTMM formalism offers the advantage of providing an efficient multiscale partitioning of the signal by means of the WTMML's. In fact the method can be seen as the most general form of conducting a global multiscale analysis in accordance to the mathematical rules set by distribution theory, see chapter 5 and the references therein. Given this multiscale analysis technique it is possible to measure the singularity spectrum, the mass exponent function and generalized dimensions from a larger class of objects. The reader is referred to chapter 6 for an introduction on the basic concepts related to multifractality.

In the pertaining sections an overview is given of the main results presented in the work of Bacry et al. (1993). At first a partition function is introduced. This partition is used to compute the mass exponent function $\tau(q)$ and is based on a multiscale partitioning, covering, provided by the continuous wavelet transform (Holschneider, 1987; Arneodo

[11] Notice that these cascade models are widely used to model highly nonlinear phenomena such as turbulence (Mandelbrot, 1974; Parisi and Frisch, 1985; Schertzer and Lovejoy, 1987a; Schmitt et al., 1992).

et al., 1988; Bacry et al., 1993; Muzy et al., 1993; Davis et al., 1994; Holschneider, 1995). Then I proceed with the introduction of a multiscale partitioning/covering by means of the WTMML followed by extending the theory to apply to singular signals, functionals rather than to measures (Bacry et al., 1993; Muzy et al., 1993). Finally a theorem is presented that puts a handle on how to deal with singular functionals composed of a singular measure that is perturbed by a C^∞ oscillatory contribution. Such a perturbation is not unimaginable in the context of this thesis and will result in *phase transition*, a non-analyticity in the mass exponent function resulting in a breakdown for the right-hand side of the singularity spectrum. I will conclude the discussion by testing the proposed method on the monofractal singular functionals Gaussian white noise and Brownian motion followed by a test on the binomial multifractal. For more practical considerations on the actual application I like to refer the reader to chapter 2 where I set the local/global multiscale analysis procedures to work on real well-log measurements.

8.5.1 Partitioning by the continuous wavelet transform

Let me start the discussion by defining the continuous wavelet transform for a measure rather than for a functional, i.e.

$$\mathcal{W}\{\mu', \psi\}(x, \sigma) = \int_{\text{supp}\,\mu} \psi(\frac{x - x'}{\sigma}) \mathrm{d}\mu(x'), \tag{8.50}$$

where the supp denotes the support of the measure μ and $\mu' = \frac{\mathrm{d}\mu}{\mathrm{d}x}$[12]. Concerning the wavelet itself, it is assumed to be of compact support, to be smooth enough and to have the proper number of vanishing moments M and, finally, the $\sigma^{-\frac{1}{2}}$ has been omitted so that in equation (8.51) the normalization term drops out. This is the same choice as Bacry et al. (1993) made, whose notation I will follow here.

Using the above definition for the wavelet transform of a measure $\mu(x)$ leads to the following relationship between the wavelet transform of a measure and that of a functional, i.e.

$$\mathcal{W}\{\mu', \psi\}(\sigma, x\} = -\mathcal{W}\{f, \psi'\}(\sigma, x), \tag{8.51}$$

where $f(x) = \int_0^x \mathrm{d}\mu(x)$ and where $\psi'(x)$ is the derivative of the wavelet $\psi(x)$.

Now define a partition function for a measure[13] as follows

$$K_\sigma\{\mu, \psi\}(p, q) = |\sigma|^{-p} \int_{-\infty}^{+\infty} |\mathcal{W}\{\mu', \psi\}(\sigma, x)|^{q-1} \mathrm{d}x \tag{8.52}$$

with $p \in \mathbb{R}$, and q the positive order of the "moment" or the reciprocal of the temperature

[12]This definition of μ' is purely formal to remain consistent in notation.

[13]The measure considered by Bacry et al. (1993) is a Bernoulli measure, i.e. $\mu \in \mathcal{M}$, which is obtained via iterative one-dimensional mappings known as "baker, cooky cutter" or Markov maps.

in the statistical mechanics nomenclature[14], $q \geq 1$. Given this multiscale partition function it is possible to estimate the $\tau(q)$ function (see chapter 6). For the details and proves the reader is referred to (Holschneider, 1987; Ghez and Vaienti, 1992; Bacry et al., 1993).

Theorem 8.6: Generalized fractal dimensions of a measure by the WT (Bacry et al., 1993)
Let $\mu \in \mathcal{M}$. Let $K_\sigma\{\mu,\psi\}(p,q)$ be its corresponding partition function as defined in equation (8.52) and let $D_q = \frac{\tau(q)}{q-1}$ be its generalized fractal dimensions. Then, for $q \geq 1$, $\tau(q)$ is the transition exponent such that

$$p-1 < \tau(q-1) \quad \Rightarrow \quad \lim_{\sigma \downarrow 0} K_\sigma\{\mu,\psi\}(q) = 0$$

$$p-1 > \tau(q-1) \quad \Rightarrow \quad \lim_{\sigma \downarrow 0} K_\sigma\{\mu,\psi\}(q) = \infty.$$

Inspection of this theorem clearly reveals the striking similarity between this formulation and the original definition 6.3 for the generalized dimensions in section 6.3 of chapter 6. Unfortunately this formalization is, contrary to definition 6.3, unable to allow for $q \leq 1$. This is due to the limited restrain which can be imposed on the parts of the $\mathcal{W}\{\mu,\psi\}(\sigma,x)$ being close to zero, i.e. the parts leading to a divergence for the negative moments. Ghez and Vaienti (1992) proposed the use of a strictly positive analyzing "wavelet" in order to circumvent this flaw. However, this attempt leads to a formulation very much in line with the box-counting and therefore it suffers from the same shortcoming. This shortcoming refers to the limited detection range, for positive singularities, of an analyzing function lacking vanishing moments, a notion one is all to aware off judging by the local scale analysis and the examples to be shown in section 8.6. A way out of this dilemma is to define an alternative partition function that is based on wavelets with the proper number of vanishing moments but that circumvents the above divergence problem. In that way the global multifractal analysis techniques are not longer limited to singular measures solely.

8.5.2 Partitioning by the WTMML

Under the same arguments as stated in section 8.4.1 an alternative approach is proposed which instead of using all values of the continuous wavelet transform at a specific scale, the $\mathcal{W}\{\mu',\psi\}(\sigma,x)$, $x \in \mathbb{R}$, for $\sigma = \sigma_0$, it uses a partitioning in terms of the extrema, i.e. the WTMML's (Mallat and Hwang, 1992; Mallat and Zhong, 1992). In this way the partition function becomes

$$Z_\sigma\{\mu,\psi\}(p,q) = |\sigma|^{-p} \sum_{m \in M} [\sup\{\text{WTMML}_\sigma(m)\}]^q, \quad q \in \mathbb{R}, \tag{8.53}$$

[14]See chapter 6 where I make some remarks on the relation between the phase diagrams of statistical mechanics and the $\tau(q)$ or $f(\alpha)$ spectrum.

where WTMML$_\sigma(m)$ refers to a maximum, a WTMM, at scale σ yielded by the m^{th} modulus maxima line of the set L of curves parameterized by $X_m(\sigma)$, $m \in L$, on which the modulus of the continuous wavelet transform is maximum, i.e. apply definition 8.2. The supremum, sup, denotes the supremum of the WTMML yielding an extension of the permissible q range, $q \in \mathbb{R}$. It is this supremum necessary to circumvent the possible divergence for negative powers q in those cases where the values of the modulus maxima become too small (Bacry et al., 1993; Muzy et al., 1993). Remark that the partitioning by the WTMML's is very close in line with the optimal multiscale covering of the singular support of μ maximizing, for $q \geq 0$, or minimizing, for $q < 0$, the partition function. Given the partitioning defined in equation (8.53) Bacry et al. (1993) presents the following theorem:

Theorem 8.7: Generalized fractal dimensions of a measure by the WTMML (Bacry et al., 1993)
Let $\mu \in \mathcal{M}$. Let $Z_\sigma\{\mu,\psi\}(p,q)$ be its corresponding partition function as defined in equation (8.53) and let $D_q = \frac{\tau(q)}{q-1}$ be its generalized fractal dimensions. Then, for $\forall q \in \mathbb{R}$, $\tau(q)$ is the transition exponent such that

$$p < \tau(q) \quad \Rightarrow \quad \lim_{\sigma \downarrow 0} Z_\sigma\{\mu,\psi\}(p,q) = 0$$

$$p > \tau(q) \quad \Rightarrow \quad \lim_{\sigma \downarrow 0} Z_\sigma\{\mu,\psi\}(p,q) = \infty.$$

This theorem allows for the estimation of the singularity spectrum, by conducting a Legendre transform on the $\tau(q)$ function, for a singular measure, but now by means of the WTMML's and for all $q \in \mathbb{R}$.

As a final remark I would like, for clarity, to mention again that there is a fundamental parallel hidden between the required optimal covering in the definition of the Hausdorff and generalized dimensions, see chapter 6 and the optimal covering obtained by the WTMML method. In this comparison the wavelet can be seen as an optimal σ-size "box" whereas the WTMML's act as navigators to optimally position the wavelets on the measure. The additional supremum in equation (8.53) can be interpreted as a preconditioner controlling the possible divergences, in case of $q < 0$, for the small values of $|\mathcal{W}\{\mu,\psi\}(\sigma,x)|$.

8.5.3 Singularity spectrum of a function

Consider the class of functionals being defined by

$$f(x) \triangleq \int_0^x \mathrm{d}\mu(x') + P_n(x), \tag{8.54}$$

where $\mu(x) \in \mathcal{M}$ and $P_n(x)$ a polynomial of arbitrary but finite order n. This polynomial perturbation will later be generalized to a C^∞ function, i.e. $n \to \infty$. Let me first define

the singularity spectrum f(α) for a functional[15] as

Definition 8.3: Singularity spectrum of a function (Bacry et al., 1993; Holschneider, 1995)
A singularity spectrum of a function $f(x)$ is the function f(α), $\alpha \in H_f$, *the set of finite Hölder exponents of f, such that*

$$\mathrm{f}(\alpha) = \dim_H \{x_0 \in \mathbb{R} | \alpha(x_0) = \alpha\}, \tag{8.55}$$

where $\dim_H$ *denotes the Hausdorff dimension.*

This definition is based on a slightly different definition for the local Hölder exponent as used so far.

Definition 8.4: Local scaling exponent of a function (Bacry et al., 1993; Mallat and Hwang, 1992; Holschneider, 1995)
A function f is said to be of local Hölder exponent α at the point $x_0 \in \mathbb{R}$, if and only if $\alpha(x_0)$ is the largest exponent α such that there exists a constant $A > 0$ and a polynomial $P_n(x)$ of order n such that for all x in the neighbourhood of x_0

$$|f(x) - P_n(x - x_0)| \leq A|x - x_0|^{\alpha}. \tag{8.56}$$

If $f \in C^{\infty}$, then $\alpha(x_0) = \infty$ for all $x_0 \in \mathbb{R}$.

Let me now restate without giving the proof the fundamental result by Bacry et al. (1993) allowing for the estimation of the singularity spectrum for singular functions[16].

Theorem 8.8: Generalized fractal dimensions of a function by the WTMML (Bacry et al., 1993)
Let $\mu \in \mathcal{M}$. Let $P_n(x)$ be a polynomial of the order n and $f(x) = \int_0^x \mathrm{d}\mu + P_n(x)$. Let ψ be an analyzing wavelet with $M > n$ vanishing moments, i.e. $\forall k, 0 \leq k \leq n$, $\int x^k \psi(x)\mathrm{d}x = 0$. Let $Z_\sigma\{\mu, \psi\}(p, q)$ be the corresponding partition function as defined in equation (8.53)

$$Z_\sigma\{f, \psi\}(p, q) = |\sigma|^{-p} \sum_{m \in M} [\sup \{\mathrm{WTMML}_\sigma(m)\}]^q, \quad q \in \mathbb{R}. \tag{8.57}$$

Then, for all $q \in \mathbb{R}$, $\tau(q)$ is the transition exponent such that

$$p < \tau(q) \quad \Rightarrow \quad \lim_{\sigma \downarrow 0} Z_\sigma\{f, \psi\}(p, q) = 0$$

$$p > \tau(q) \quad \Rightarrow \quad \lim_{\sigma \downarrow 0} Z_\sigma\{f, \psi\}(p, q) = \infty.$$

[15] I am aware of the deviant notation I opted to use for the singularity spectrum of a function. Bacry et al. (1993) and others use $D(h)$ and h for the singularity spectrum and the Hölder exponent while I like to stick to the notation used for the corresponding quantities used for the multifractal measures.

[16] Note that I am discussing here the partition function of a measure perturbed by a polynomial where I before discussed the partition function of a measure solely

By way of the formulation coined in theorem 8.8 it is clear that the computation of the $\tau(q)$ function for a singular function is very similar to that of a singular measure. There is, however, a profound difference. Namely when, for example, a constant is added to the measure, i.e. a zero order polynomial, then the multifractal can dramatically be altered by this regular perturbation in cases where a conventional coarse graining by means of a smoothing kernel with a non-vanishing mean, say the box-car, is used. This major difficulty can be attributed to the fact that the constant is completely dominating the behaviour of the partition function. This partition function is defined for a conservative measure, i.e. $\int_x \mathrm{d}\mu_\sigma(x) < \infty$ and independent of σ whereas clearly $\mu + C \notin \mathcal{M}$ leads to an apparent non-integrability killing the behaviour of the partition function. However, when WTMML method is used, one can overcome this flaw by choosing the appropriate number of vanishing moments, M, yielding an analyzing wavelet that is *orthogonal* with respect to polynomials $P_n(x)$ with $n \leq M$.

As to conclude the discussion on obtaining the singularity spectrum from a singular function I would like to review the case where the perturbation is $P(x) \in C^\infty$, i.e. perturbing the measure with an infinitely differentiable function, e.g. $\sin(x)$. Such a situation could be relevant in cases where, for example, a periodic disturbance, say a seasonal contribution in the sedimentation process, is superimposed on the multifractal measure.

Again suppose a sufficiently smooth wavelet with M vanishing moments. Because of the C^∞ perturbation one has now to do with two distinct groups of WTMML's: the first group can be associated with the singular measure μ and the second with the C^∞ perturbation. From previous experience one can say that the amplitudes along the WTMML's for the second category will all be of $\mathcal{O}(\sigma^M)$, a property already observed in the local regularity analysis. Moreover it is known that the q acts a selector for the different singularities and it will therefore not be surprising that there will be a negative[17] critical q delineating a transition in the behaviour of the mass exponent function $\tau(q)$ due to the emergence of the WTMML's from the C^∞ perturbation. This non-analyticity in the $\tau(q)$ function can be associated with a *phase transition* marking a break in the multifractal scaling induced by the perturbation.

Without going into all details I would like to restate the following theorem which allows for the determination of the critical exponent and the singularity spectrum for singular functions being perturbed by a C^∞ functions.

Theorem 8.9: Critical exponent (Bacry et al., 1993)
There exists a $q_{crit} < 0$ *such that*

1. *If* $q > q_{crit}$, *then* $\tau(q)$ *is the transition exponent such that*

$$p < \tau(q) \implies \lim_{\sigma \downarrow 0} Z_\sigma\{f, \psi\}(p, q) = 0,$$
$$p > \tau(q) \implies \lim_{\sigma \downarrow 0} Z_\sigma\{f, \psi\}(p, q) = +\infty.$$

[17] Because the negative q's hunt for the positive singularities.

2. *If* $q < q_{crit}$, *then*

$$p < qM \implies \lim_{\sigma \downarrow 0} Z_\sigma\{f,\psi\}(p,q) = 0,$$
$$p > qM \implies \lim_{\sigma \downarrow 0} Z_\sigma\{f,\psi\}(p,q) = +\infty.$$

The proof of this theorem can be found in Bacry et al. (1993). The critical moment q_{crit} marks two distinct domains for which either the $\tau(q)$ function can be recovered, in the case $q > q_{\text{crit}}$, or where no information can be inferred. The sudden change in behaviour of the mass exponent function $\tau(q)$ at $q = q_{\text{crit}}$ can be associated with a first order phase transition. This phase transition – where the linear curve switches to a nonlinear one – marks a break of the heterogeneous scaling and manifests itself by a singularity in its derivative, $\partial_q \tau(q)$ at $q = q_{\text{crit}}$. With other words, below the critical exponent q_{crit} the "system" is in its *regular* phase whereas for $q > q_{\text{crit}}$ the system enters the singular multifractal phase.

8.5.4 Measuring procedure for global multiscale analysis

Now that all essential tools necessary to recover the singularity spectrum have been introduced I would like to pay attention to actually determining the singularity spectrum. The binomial multifractal of chapter 6 will serve as a test case since its f(α) is explicitly known. In chapter 2 I will set this technique to work on a real data set, a well-log measurement.

Procedure 8.3: Measurement of the singularity spectrum f(α)
The procedure to measure the mass exponent function $\tau(q)$ *and singularity spectrum* f(α) *from singular measures and functions consists of the following steps:*

1. *Select the proper family analyzing wavelets in accordance with the conditions summarized in the theorems dealing with the estimation of the singularity spectrum for singular measures and functions. This choice has a drastic effect on the detectable range of singularities as already stressed in the local multiscale analysis. In section 8.6 I will pay ample attention to this important issue.*

2. *Compute the continuous wavelet transform of the signal* f *with respect to this family of analyzing wavelets, see figure 8.8, using equation* (8.3).

3. *Locate the* WTMML*'s in accordance to definition 8.2. Perform the* WTMML *partitioning: that is create a set* L *of curves, parameterized by* $X_m(\sigma)$,$m \in L$ *on which the modulus of the continuous wavelet transform is maximum.*

4. *Study the behaviour of the partition function,*

$$Z_\sigma(q) = Z_\sigma\{f,\psi\}(p,q)|_{p=\frac{1}{2}} \tag{8.58}$$

as defined in equation (8.57) and with the emphasis on the behaviour for σ approaching the resolution at which the measurement has been taken. This corresponds to stacking the q^{th} power of the WTMM*'s along the spatial coordinate.*

5. *Compute the $\tau(q)$ function using the property*

$$Z_\sigma(q) \propto \sigma^{\tau(q)}, \quad \sigma \downarrow 0 \tag{8.59}$$

i.e.

$$\tau(q) = \lim_{\sigma \downarrow 0} \frac{\log Z_\sigma(q)}{\log \sigma}, \tag{8.60}$$

which corresponds to a linear behaviour in the log-log *space:*

$$\log Z_\sigma(q) \propto \tau(q) \log \sigma. \tag{8.61}$$

The $\tau(q)$ function is obtained by conducting, for each q-value, a linear regression on the partition function.

6. *Compute the* f(α) *spectrum*[18] *by means of performing a Legendre transform on the $\tau(q)$ function, i.e.*

$$\begin{aligned} \tau(q) &= \min_\alpha \{q\alpha - \mathrm{f}(\alpha)\} \\ \mathrm{f}(\alpha) &= \min_q \{\alpha q - \tau(q)\}. \end{aligned} \tag{8.62}$$

Given this procedure it is now time to test it on a number of examples, the outcome of which is known beforehand, therefore acting as a check of confidence for the proposed method.

Singularity spectrum of distributions with one singularity

Consider the distribution given by

$$g_1(x) = \begin{cases} x^{-\frac{1}{2}} & \text{if } x > 0, \\ (-x)^{-\frac{1}{2}} & \text{if } x < 0. \end{cases} \tag{8.63}$$

This distribution has only one singularity of strength $-\frac{1}{2}$, so it is expected that its singularity spectrum $f(\alpha)$ is zero everywhere except at $\alpha = -\frac{1}{2}$, where it is negligible small.

Figure 8.7 (a) on the left shows $g_1(t)$. In figure 8.7 (b) the locations of the modulus maxima lines are depicted. The wavelet used is the first derivative of the Gaussian.

[18] Notice that I use a deviant notation for the singularities of a function, α instead of h.

Figure 8.7 (c) gives the logarithm of $Z_\sigma(q)$ against $\log\sigma$ for various values of q. It can be shown that $\tau(q)$ must be

$$\tau(q) = -\frac{q}{2}. \tag{8.64}$$

Figure 8.7 (d) shows the computed $\tau(q)$, which is in accordance with equation (8.64). Finally, figure 8.7 (e) shows the singularity spectrum f(α), which is indeed zero except at $\alpha = -\frac{1}{2}$, where it is very small. The above procedure is repeated for a positive singularity of strength $\alpha = +\frac{1}{2}$,

$$g_2(x) = \begin{cases} x^{+\frac{1}{2}} & \text{if } x > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{8.65}$$

It is of course expected that the same singularity spectrum is found as in the previous example, but now shifted by $+1$. So f(α) is zero everywhere except at $\alpha = +\frac{1}{2}$ where it is negligible.

Figure 8.7 (a) on the right shows g_2 and figure 8.7 (b) shows the location of the modulus maxima line. For this singularity it can be shown that

$$\tau(q) = +\frac{q}{2}, \tag{8.66}$$

which is confirmed by figure 8.7 (d). In figure 8.7 (e) the singularity spectrum is shown. Indeed it is zero everywhere, except at $\alpha = +\frac{1}{2}$, where it is very small.

Singularity spectrum of white noise and Brownian motion

Gaussian white noise is a stochastic process, which is characterized by a probability density function given by

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{x-\mu}{\sigma})^2}. \tag{8.67}$$

Let the process $\varepsilon(x)$ symbolize a process which has a Gaussian probability density as in equation (8.67) with a mean μ and a variance σ^2.

It can be shown that $\varepsilon(x)$ is singular everywhere and that its singularity strength $\alpha = -\frac{1}{2}$ everywhere. This means that its singularity spectrum f(α) is zero everywhere except at $\alpha = -\frac{1}{2}$ where it is 1. Since $\varepsilon(t)$ has singularities of strength $-\frac{1}{2}$ everywhere, the modulus maxima lines behave as σ^0, see equation (8.46), i.e. they behave in exactly the same manner as in the first example of the previous section. However, there is one difference in the sense that white noise is singular everywhere a notion being reflected in a somewhat different behaviour for the partition function which now reads (Muzy et al., 1993)

$$Z_\sigma(q) \propto \sigma^{-\frac{1}{2}q-1}, \tag{8.68}$$

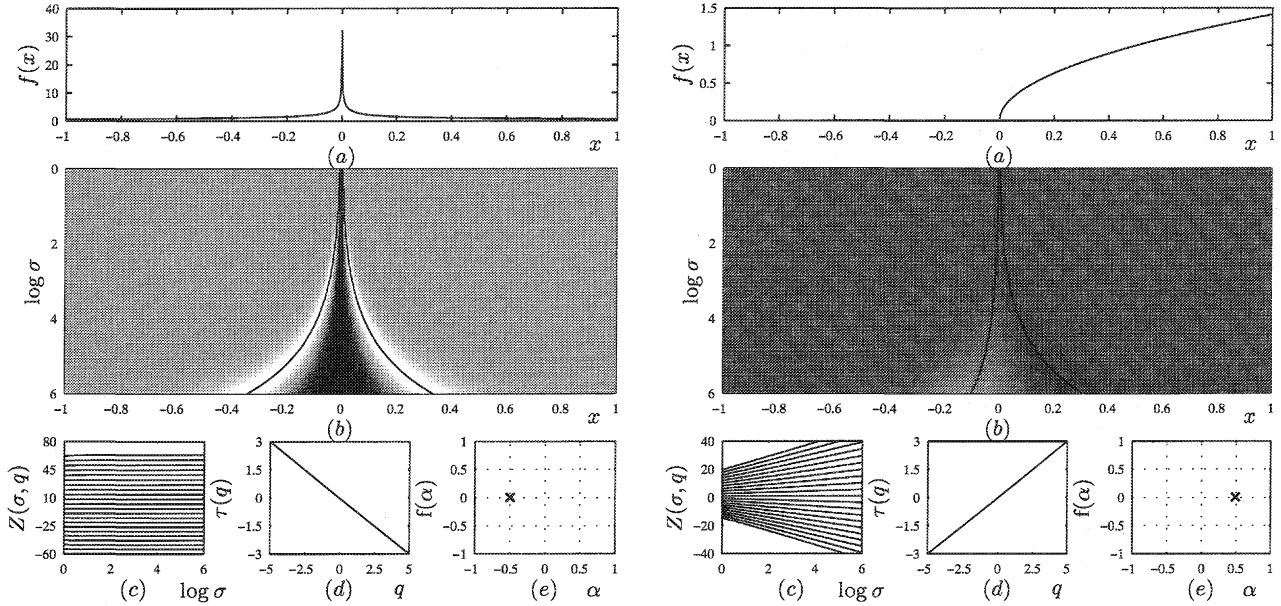

Figure 8.7 *(a) The distributions of equations* (8.63) *and* (8.65) *are reviewed in this figure. These distributions have singularities of strength* $\alpha = \mp\frac{1}{2}$. *(b) The location of the modulus maxima lines of the wavelet transform of (a). The wavelet used is the second derivative of the Gaussian, so it has two vanishing moments. (c) The partition functions* $Z_\sigma(q)$ *computed from the modulus maxima lines. (d) The* $\tau(q)$ *functions. The predicted function (dots) is plotted and the computed function is plotted (solid). (e) The* f(α) *functions belonging to these distributions. They are zero everywhere except at* $\alpha = \mp\frac{1}{2}$ *where they are very small. This is expected since the dimension of a point is zero.*

yielding an offset by a one with respect to the behaviour of the partition function computed for an isolated singularity. This offset bears the notion that the white noise is singular everywhere which becomes apparent by the fact that the associated dimension equals unity, i.e. $f(-\frac{1}{2}) = 1$. This is illustrated in figure 8.8 on the left. Figure 8.8 (a) gives a realization of $\varepsilon(x)$. Figure 8.8 (b) shows the location of the modulus maxima lines, which start at every abscissa in accordance with the fact that $\varepsilon(x)$ is singular everywhere. Figure 8.8 (c) shows $Z_\sigma(q)$. Figure 8.8 (d) gives $\tau(q)$ which equals $-q/2$. The singularity spectrum $f(\alpha)$ is depicted in figure 8.8 (e) and is indeed 1 for $\alpha = -\frac{1}{2}$ and zero everywhere. Notice the differences and similarities with the first example of the previous section, see figure 8.7. The main difference is the number of modulus maxima lines, which determine the value of $f(\alpha)$ at $\alpha = -\frac{1}{2}$. As a second example of a process which has singularities of the same strength everywhere, consider the process $B(x)$ given by

$$\frac{\mathrm{d}}{\mathrm{d}x}B(x) = \varepsilon(x), \quad B(0) = 0, \tag{8.69}$$

or

$$B(x) = \int_0^x \varepsilon(x')\mathrm{d}x'. \tag{8.70}$$

This process is referred to as Brownian motion, see chapter 7 and the references therein, and it describes the random movement of a particle through a liquid or a gas. The quantity $B(x)$ is the velocity or momentum of a particle of unit mass and $\varepsilon(x)$ is the random force acting upon the particle.

With the knowledge of the previous examples it is not hard to predict that $B(x)$ has a singularity spectrum $f(\alpha)$ which equals 1 for $\alpha = \frac{1}{2}$ and is zero elsewhere. Figure 8.8 on the right confirms this prediction, showing $B(x)$, the WTMML's, the partition function $Z_\sigma(q)$ and mass exponent $\tau(q)$, and finally the singularity spectrum $f(\alpha)$. The spectrum is indeed 1 for $\alpha = \frac{1}{2}$ and zero elsewhere while the mass exponent function equals $\tau(q) = \frac{1}{2}q - 1$.

White noise and its primitive, in the sense of distributions, Brownian motion are two specific members of the class of fractional Gaussian noises and fractional Brownian motions (Mandelbrot and Wallis, 1969; Mandelbrot, 1982; Montroll and Schlessinger, unknown; Schertzer and Lovejoy, 1993; Samarodnitsky and Taqqu, 1994; Klafter et al., 1996). The fractional Brownian motion processes are indexed by the exponent H. This exponent expresses the degree of fractional integration/differentiation with respect to Brownian motion and determines the homogeneous/monofractal scaling. Moreover it provides estimates for the powerlaw behaviour displayed by the stochastic quantities that can be assimilated to these processes. For instance, the powerlaw decay rate for the energy spectrum of fractional Brownian motion is given by

$$S(k) \propto \frac{1}{k^\beta} \tag{8.71}$$

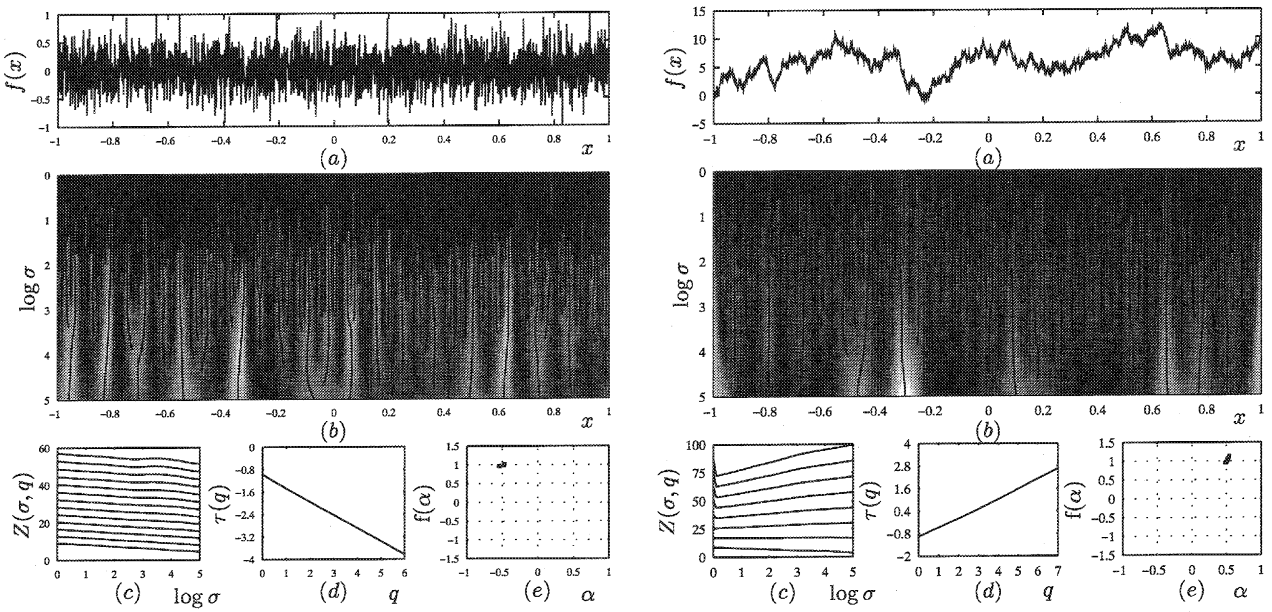

Figure 8.8 *(a) Realizations of Gaussian white noise (left) and Brownian motion (right). These signals have singularities of strength* $\alpha = \mp\frac{1}{2}$ *everywhere. (b) The location of the modulus maxima lines of the wavelet transform of (a). The wavelet used is the first derivative of the Gaussian, so it has one vanishing moment. These plots show that on the finest scale the modulus maxima lines start everywhere, which is a reflection of the fact that the signals are singular everywhere. (c) The partition functions* $Z_\sigma(q)$*, computed from the wavelet modulus maxima lines. (d) The* $\tau(q)$ *functions, computed from the partition functions. Both the theoretical expected functions (dots) as the computed functions (solid) are depicted. (e) The* $f(\alpha)$ *spectra, estimated from the computed* $\tau(q)$ *of (d). It is zero everywhere, except at* $\alpha = \mp\frac{1}{2}$*, where it is one. This is because the signals have singularities of equal strength everywhere, i.e. they are monofractals.*

with $1 < \beta < 3$ with $\beta = 2H + 1$ and $0 < H < 1$. The partition function for this type of process is given by

$$Z_\sigma(q) \propto \sigma^{q\alpha - 1} \tag{8.72}$$

with $\alpha = H$ being the single singularity strength occurring. As I defined in chapter 2 the exponent H, for fractional Brownian motions, is related to the mass exponent function according to

$$H \triangleq \{\zeta(q)\}_{q=1} = \{\tau(q)\}_{q=1} + 1, \tag{8.73}$$

where $\zeta(q)$ refers to the scaling displayed by the generalized structure function[19], i.e. $\langle |B_h(x+\sigma) - B_H(x)|^q \rangle \propto \sigma^{\zeta(q)}$ with $B_H(x)$ being the index-H Brownian motion (Schmitt, 1993; Muzy et al., 1993; Davis et al., 1994). Application of this definition, see figure 8.8, to white noise yields $H = -\frac{1}{2}$ for white noise and $H = \frac{1}{2}$ for Brownian motion itself. The fact that there is a difference of one between these two can be attributed to the fact that white noise is the derivative of Brownian motion to be understood in the sense of distributions.

The singularity spectra considered so far are concentrated at one point. In the previous section they were even nearly zero at that point. This is a reflection of the fact that the signals of the previous sections are extremely sparse fractals, i.e. the dimension of their support is zero. The examples reviewed in this section on the other hand have singularity spectra equalizing 1 which corresponds to the notion that the dimension of the singular support is one, i.e. these distributions are singular everywhere. Notice, however, that all these f(α)'s share the property that they are concentrated at one point which means that the signals are *monofractal*, they display a scale-invariance. In more intricate cases the $\tau(q)$ becomes nonlinear corresponding to a convex singularity spectrum which indicates that the signal contains singularities with all kinds of strengths. In this latter case the signal becomes multifractal and not longer displays a trivial type of scale-invariance.

Singularity spectrum of the binomial multifractal

In order to validate the procedure introduced in the previous section it is tested on a simple multifractal of which the singularity spectrum is known. For that reason a binomial multifractal measure introduced in chapter 6 will be examined.

In figure 8.9 (a) on the left the generated binomial multifractal measure is depicted together with its partitioning by the WTMML's in figure 8.9 (b). In the bottom row of figure 8.9 the partition function $Z_\sigma(q)$ is depicted, see figure 8.9 (c), computed from the WTMML partitioning of the measure. The mass exponent function $\tau(q)$, figure 8.9 (d) together with its expected values obtained from the equation $\tau(q) = \log_2[p_1^q + p_2^q]$ (see

[19]Notice that this structure functions is only applicable for processes with stationary first increments an observation limiting the observation range for the singularities to $\alpha \in (0, 1)$ (Muzy et al., 1993).

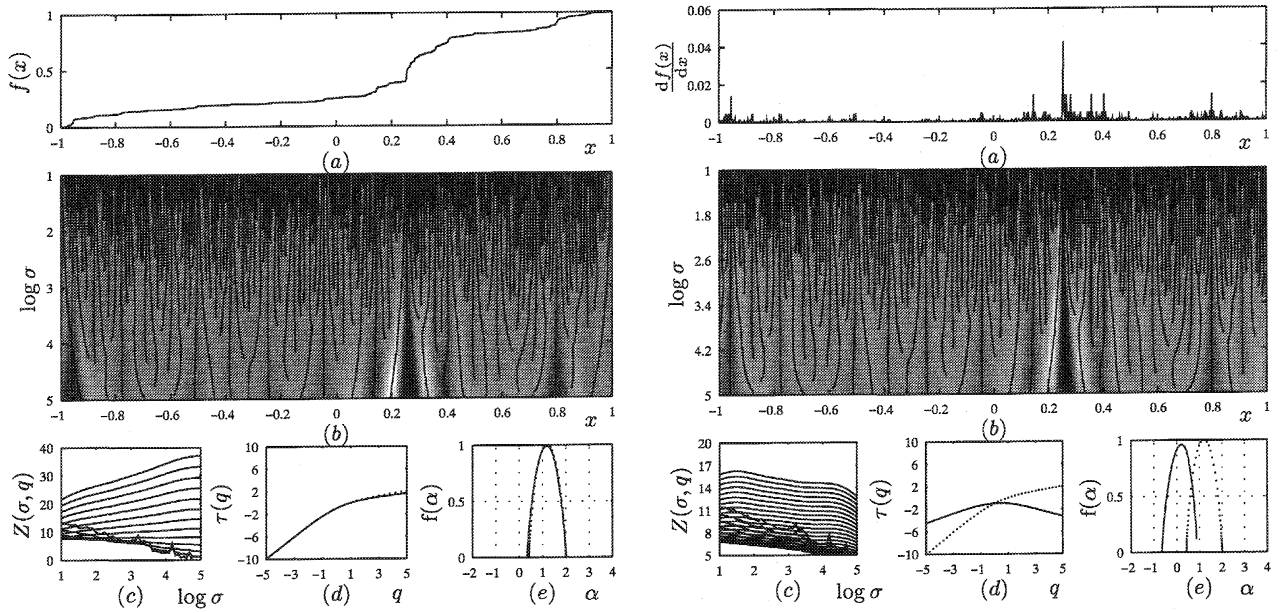

Figure 8.9 *(a) In this figure the binomial multifractal measure (left) and its density are reviewed (right). These two fractals contain a hierarchy of singularities and are singular everywhere. With the choices made, $p_1 = 0.25$ and $p_2 = 0.75$ the predicted values for the endpoints of the singularity spectrum are $\alpha_{\max} = 2$ and $\alpha_{\min} = 0.4$ for the measure while the spectrum and its endpoints is shifted one to the left for the density. (b) The location of the wavelet transform modulus maxima lines. The wavelet used has three vanishing moments. Since the second derivative of this signal is singular everywhere, the modulus maxima lines start on the finest scale everywhere. (c) The $Z_\sigma(q)$ function, computed from the modulus maxima lines of (b). The predicted (dots) and the computed (solid) $\tau(q)$ function. (e) The computed $f(\alpha)$ spectrum (solid), which matches fairly good with the predicted (dots) spectrum for the measure. The maximum of $f(\alpha)$ is found to be one, in accordance with the fact that the second derivative of this distribution is singular everywhere.*

Siebesma (1989)) with $p_1 = 0.25$ and $p_2 = 0.75$ and finally, in figure 8.9 (e) the singularity spectrum $f(\alpha)$ itself together with the expected values yielded by the equation $f(\alpha) = -c(\alpha)\log_2(\alpha) - (1 - c(\alpha))\log_2(1 - c(\alpha))$ with $c(\alpha) = (\alpha - \alpha_{\min})/(\alpha_{\max} - \alpha_{\min})$ and $\alpha_{\min} = \log_2 p_1$ and $\alpha_{\max} = \log_2 p_2$.

Inspection of both the mass exponent and the singularity spectrum shows that the proposed analyzing technique accurately captures multifractal characteristics. The estimated values do not differ much from the theoretical curves and both asymptotes $(q \to +\infty)$ and $(q \to -\infty)$, delineating $\alpha_{\min}$ and $\alpha_{\max}$ are also accurately recovered. The maximum is located at $q = 0$ and is approximately equal to one. This latter observation is consistent with the fact that the binomial multifractal is singular everywhere.

The example displayed in figure 8.9 on the right hand side illustrates the notion that the singularity spectrum is shifted by one to the left when the multifractal measure is differentiated[20]. Note, however, that these results only hold when the proper analyzing wavelet is used. In the next section I will pay attention to the impact of choosing erroneous analyzing wavelets.

8.6 The significance of choosing the proper wavelet

In this section I will discuss by means of a number of examples the implication of selecting an improper analyzing wavelet to conduct the global or local multiscale analysis and characterization. It is shown that an erroneous choice limits the detectable range of singularities. The two examples to be shown comprise

- the implication of using an insufficient number of vanishing moments in the context of conducting a global multiscale analysis on an integrated binomial multifractal.
- the implication of using a wavelet that is not smooth enough in the context of the local multiscale analysis conducted on a singular distribution containing one isolated singularity.

Notice that the observations to be made refer to the multiscale analysis as a whole. This means that when something goes wrong in the local multiscale analysis it will also go wrong in the global multiscale analysis and *vice versa*.

Erroneous number of vanishing moments

First order integration of a multifractal yields a shift of magnitude one to the right for the singularity spectrum. However, this can only be observed when the analyzing wavelet used to conduct the coarse-graining possesses an adequate number of vanishing moments.

[20] Again to be interpreted in the distributional sense which corresponds to the fact that the derivative is approximated at a scale corresponding to the scale towards which the multifractal is generated. The derivative is taken at the inner scale of the coarse-grained binomial multifractal.

Otherwise the analysis will be blind for the singularities the Hölder exponents of which exceed the number of vanishing moments M. To illustrate this notion I included figure 8.10 where on the left a wavelet with the proper number for M was used while on the right the number of vanishing moments was taken to be insufficiently high, namely $M = 2$ rather than the required $M = 3$. This wrong doing manifests itself in a breakdown of the singularity spectrum towards the $q \to \infty$ asymptote, towards the $\alpha_{\max}$ endpoint. Notice also that the singularity spectrum "saturates" for approximately $\alpha = 2$, a value that, as expected, exactly corresponds to the inadequate number of vanishing moments used, $M = 2$. When the number of moments is increased by one the spectrum is completely restored although it is shifted by one to the right as anticipated.

Insufficient regularity

Let me now examine the implication of selecting an analyzing wavelet with insufficient regularity to conduct a proper local multiscale analysis on an isolated singularity. The singularity opted for is the distributional derivative of the delta distribution, i.e. $f(x) = \delta^{(1)}$. The analyzing wavelet used for the multiscale analysis is the Haar wavelet

$$\psi(x) = \begin{cases} 1 & -\frac{1}{2} < x < \frac{1}{2}, \\ 0 & \text{otherwise} \end{cases} \tag{8.74}$$

and is depicted in figure 8.11. Using equation (8.6) the wavelet coefficients are found via the inner product

$$\langle \delta^{(1)}, \psi(\frac{x - x'}{\sigma}) \rangle = -\langle \delta, \psi^{(1)}(\frac{x - x'}{\sigma}) \rangle \tag{8.75}$$

which is *not* defined for the wavelet of equation (8.74). To see this the reader is referred to chapter 5 where I give a resume of distribution theory which formalizes the notion that extreme care has to be taken when computing the inner product of two singular distributions. Now how does this observation relate to approximate numerical experiments where the notion of non-differentiability seems to be irrelevant in the light of the inherent coarse-graining? To put matters to the test I conducted the local multiscale analysis with either the Haar wavelet which is clearly discontinuous, i.e. $\psi \in C^0$ or with the first derivative of the Gaussian being C^∞. What happens is that the Hölder exponent estimates differ for both cases, see figure 8.11. For the Haar wavelet the outcome equals $\alpha = -1$ while for the smooth, C^∞, first derivative of the Gaussian one finds -1.985, which is close to -2. Clearly the analysis with the Haar wavelet came up with the wrong answer, $\alpha = -1$, because the $\delta^{(1)}$ is Hölder $\alpha = -2$, while the analysis with the smooth wavelet gave the correct answer.

What happened is that the analysis by the Haar wavelet was completely dominated by the regularity of the analyzing wavelet itself rather than by the regularity of the functional under consideration.

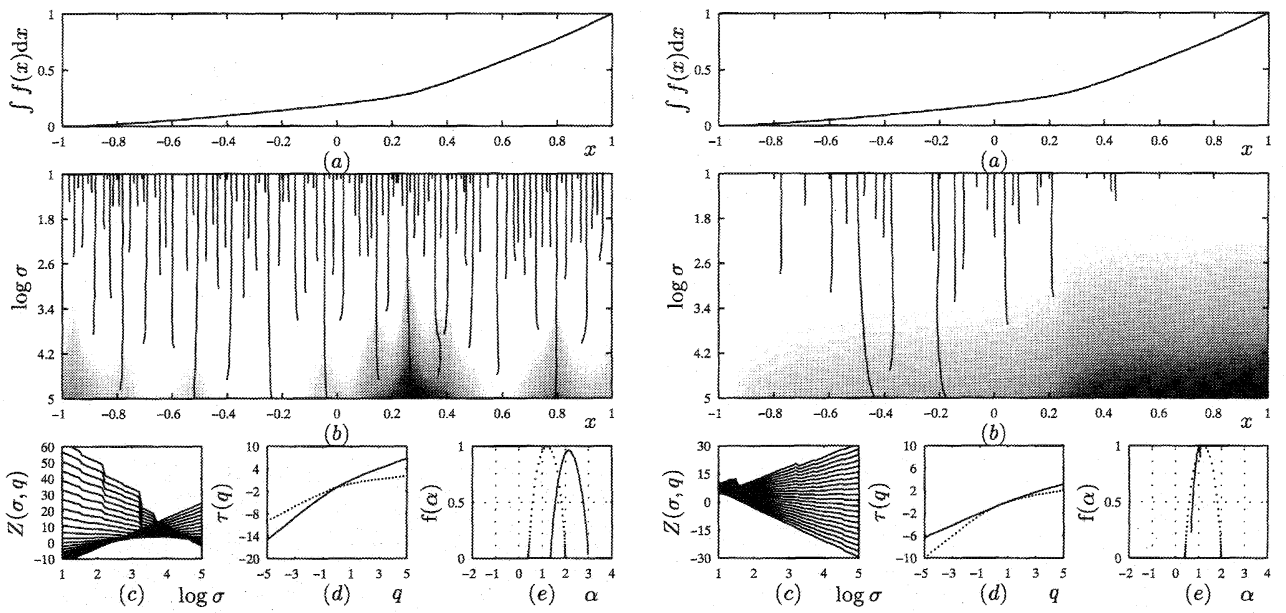

Figure 8.10 *(a) In this figure the primitive of the binomial multifractal measure of figure 8.9 is submitted to the multifractal analysis. On the left I depicted the* WTMML *partitioning yielded by a coarse-graining effectuated by a wavelet with $M = 3$ while on the right a similar procedure is displayed but now with $M = 2$. On the left one can see that the spectrum is captured and shifted by one to right as expected by the integration. The experiment on the right, where $M = 2$, shows a complete different picture. Here the number of* WTMML*'s decreases drastically because the functional is not singular everywhere in its second derivative and this leads to erroneous results for the $\tau(q)$ function and the singularity spectrum* f(α)*. For the left and right hand sides. (a) the primitive of the binomial multifractal measure. (b) The location of* WTMML*'s. (c) The $Z_\sigma(q)$ function, computed from the* WTMML*'s of (b). The predicted (dots) and the computed (solid) $\tau(q)$ function. (e) The computed* f(α) *spectrum (solid).*

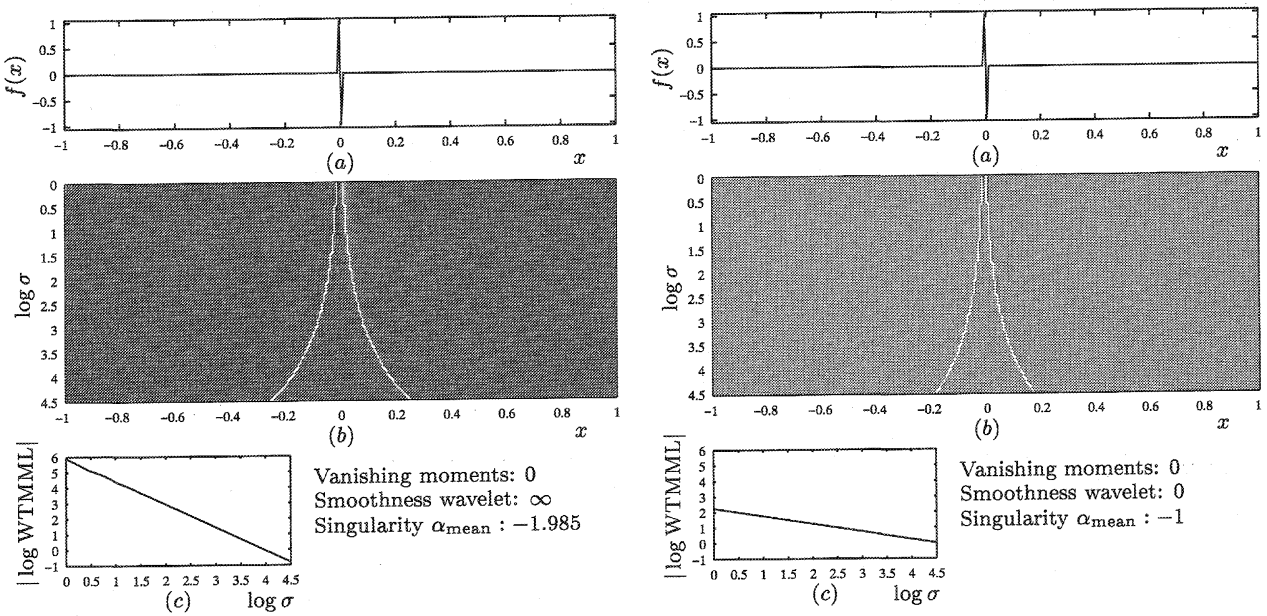

Figure 8.11 *This figure contains the regularity analysis conducted on distributions defined in terms of equation* (8.48) *with the Hölder exponent* α *set to the values* $\alpha = -0.5, 0.5, -1, 0$. *In all plots:* (b, e) *The wavelet transform of* (a, d) *viewed from above. The lines indicate the location of the modulus maxima.* (c, f) *The logarithm of the modulus maxima lines against* $\log\sigma$. *The slopes together with the abscissa of the lines for the small scales indicate that the distributions depicted in the* (a, d) *have local Hölder exponents* α *that closely match the values with which the singular distributions have been constructed.*

8.7 Concluding remarks

In this chapter the main tools required to conduct a proper local and global multiscale analysis have been presented. The validity of these two approaches have been verified by a number of elucidating examples. The continuous wavelet transform supplemented with the WTMML formalism certainly proved its added value in the sense that a larger class of signals can be examined and the WTMML allowed one to keep better track of what happens in space-scale plane. Given these multiscale analysis techniques one has a comprehensive vehicle at hand to assess the differentiability, integrability and the hierarchy of scaling exponents from sampled data sets as is demonstrated in chapter 2. In that chapter the proposed method is set to work on actual well data where I also discuss the more fundamental issues related to the estimates for the differentiability, integrability and the implication of the selection of the type of analyzing wavelet. The latter question on the proper selection of the wavelet has been addressed, by means of a series of examples, in this chapter as well. These examples unmistakably demonstrate that the theoretical considerations on the wavelet's regularity and orthogonality with respect to polynomials, i.e. its number of vanishing moments, find a one to one correspondence in the actual numerical implementation. That is to say that, for instance, the question of insufficient regularity for the wavelet leads to a complete domination of the multiscale measuring process by the erroneous instrument, analyzing wavelet, while strictly speaking the notion of differentiability is a non issue when coarse grained sampled data are concerned. One may postulate, however, that the mathematical concepts, such as differentiability, maintain their validity despite the fact that from the observational point of view no explicit conclusions can be drawn based on a "strict" mathematical interpretation of mathematics. What matters is that one has to interpret these concepts within the physical context and that is where the challenge lies. This line of thought leads, for instance, to issuing statements like *'a sample data set can not be discerned from being non-differentiable in a certain scale range'* and as the reader may have noticed such observations are highly relevant when mathematical operations on sampled data are concerned.

Part III

Epilogus

Epilogue

Specular reflections on acoustic wave motion in scaling media

E.1 Introduction

In this epilogue, I present my ideas on the implications of scaling with respect to wave theory. These ideas constitute the basis for the work I am going to continue at Stanford University, an opportunity I am offered by N.W.O.[1] in the form of a Talent Stipendium Fellowship.

Although this epilogue is not an official part of my thesis, it forms the main motivation for my research. This motivation is built upon a few observations. Firstly, the complexity of many (geo)physical phenomena are, at this very moment, not very well understood. This complexity is revealed when observing physical phenomena at a range of scales, a notion accounted for by the scaling medium representation introduced in chapter 2. Secondly, since in the derivation of the wave equation a *separation of scales* is assumed, no explicit reference is made towards the scale and operators invoking scale changes. There are indications that partial differential equations, which are derived at an infinitesimal scale, can not be straightforwardly scaled up towards the physical scale range of interest in cases where a *separation of scales* is meaningless. For that reason, the evaluation of the potential role of scale and scale dynamics in relation to the wave motion deserves the benefit of the doubt.

Given the above motivations, the purpose of this epilogue is two-fold. First of all I demonstrate the potential use of the multiscale characterization within the current formulation of acoustic wave motion. Secondly, I like to convey a possible road towards an alternative formulation for the wave motion which takes the scaling medium representation as a starting point. In this way I hope to better understand how the complexity of the medium is transferred to the wavefield.

[1] The Netherlands Organization for Scientific Research.

I will commence the discussion by stating my working hypothesis, which results in the prime motivation for the attempts towards a new formulation. Then I will proceed by bringing the findings of the multiscale characterization in relation to the main presuppositions surrounding solution techniques for acoustic wave motion.

After this overview I will go back to the constitutive relations – ruling acoustic wave motion – and re-evaluate them in the perspective of the scaling medium representation. This re-evaluation leads to the conjecture that one has to intervene at the level of the constitutive relations by bringing in the scale derivative together with an explicit reference to scale, both from *first principles*. The reason that I expect a scale derivative in the wave equation is that this operator describes the rate of *dilatation*. So far, only time and space derivatives have been considered, which describe the rate of *translation* in time and space. This time-space translation describes, via the velocity, how time and space are related. When a scale derivative comes in, one expects to be able to describe how the dilatation of time-space takes place or how translations may invoke dilatations (Nottale, 1996).

During the discussion I will point out that the homogenization theory (Auriault, 1991; Schoenberg and Muir, 1989; Gilbert and Backus, 1966) is based on the same first principles as those underlying the wave equation itself. Therefore, one may better understand the dilemma of the non-existence of a dynamic homogenization theory by observing that spatial scale derivatives are absent.

In the literature there are other discussions on the possible unjustified neglection of scale in certain problems in physics. Again, the wave equation is maybe one of these and this raises the question how to generalize the wave theory and how to formulate a proper "correspondence" principle[2]. This correspondence principle expresses a transition back to the conventional equation in case there is no scale dependence.

I am convinced that the discussion in this epilogue will shed some new and original light on wave interaction in complex media.

E.2 Working hypothesis

Recently the demands on the quality standards of the migration results have been increased. This is because of a change of interest in the direction of a more production oriented approach rather than focusing on the exploration aspects solely. This shift of interest also sparked an increase in interest on acquiring more petrophysical information via an inversion. When dealing with such a demand for information, the more dynamical aspects of wave interaction come into play. That means that one has to invoke more "physics" into the problem and so far the results have, to the author's opinion, not been

[2]In theoretical physics such a principle expresses the transition from quantum physics to classical physics.

overwhelmingly successful to come up with good predictions for the wave motion.

Let me now, without being all inclusive, categorize some points that are of interest in this discussion. From a practical point of view:

- the necessity of a sufficient[3] redundancy for the spatial frequency content of the medium fluctuations. Generally this requirement is attributed to the apparent non-linear way in which the coefficients occur in the wave equation. Unfortunately this requirement withstands a practical formulation for the inverse problem.
- the apparent lack of a full understanding on why the current modelling techniques do not in all cases come up with accurate predictions. These predictions comprise:
  - the predictions for the first break traveltimes in a VSP setting (Gretener, 1961; Kjartansson, 1981; Hsu et al., 1992; Burridge et al., 1993)
  - the predictions for the apparent anisotropy (Esmersoy et al., 1989)
  - the predictions for the apparent dispersion (Kjartansson, 1981; Hsu et al., 1992; Folstad and Schoenberg, 1993)
  - and possibly the predictions for the specular reflectivity (Folstad and Schoenberg, 1993).

From a wave theoretical point of view there is an apparent lack of a full understanding of the mechanism responsible for the specular reflections, i.e. the local scattering. In my opinion this observation is directly related to the question on the velocity experienced by the wave packet. In cases the medium contains singularities, it is difficult to define the velocity because the velocity becomes dependent on scale. Exactly in this situation specular reflections are expected to occur and in case of isolated singularities progress has been made on the local nature of the scattering process by supplementing boundary conditions. For more complicated situations there is still a lot of progress to be made in understanding wave interactions in media that strongly fluctuate on a scale range that corresponds to the dominant wavelength[4].

Maybe an even more important aspect is formed by the necessity of invoking a separation of scales. This separation of scales is required within localization theory to exclude "non-physical" chaotic waves. It is also required within homogenization theory to define equivalent media and it is tacitly assumed in setting up the constitutive relations. The recent advent of chaos and the accompanying fractal theory (Mandelbrot and Wallis, 1969; Mandelbrot, 1974, 1982; Hentschel and Procaccia, 1983; Parisi and Frisch, 1985; Schertzer and Lovejoy, 1987b, 1993; le Méhauté, 1991; Bacry et al., 1993; Muzy et al., 1993) and results coming from it at least deserve a discussion on some physical aspects

[3] As far as I know, there exist no specific estimates delineating the required amount of spatial frequency content for the medium properties.

[4] Whatever that may be, see the discussion below.

of wave motion that occur in media where the assumption of a separation of scales might be useless. Examples are media that are generated by some kind of chaotic process such as the heterogeneities in the atmosphere yielded by hydrodynamic turbulence (Schertzer and Lovejoy, 1987b; West, 1990; Tartarskii, 1971) or the subsurface's heterogeneities induced by a sedimentation process (Walden and Hosken, 1985; Leary, 1991; Saucier and Muller, 1993; Collier, 1993; Painter, 1995).

Given the above observations I think it is interesting to explore the possibilities to extend the current wave theory with the notion of scale and scale dynamics. Any such effort requires a problem and an experiment. Therefore I assume as a working hypothesis that the current formulation of the wave equation is unable to come up with good predictions in case its coefficients do not exhibit a separation of scales. For that reason I opt for an inclusion of the notion of scale and the decisive experiment will be the accomplishment of a better integration of well-data and seismic data, a problem of scales. It is my conviction that the data sets at disposal are unique in the sense that there is an excellent ratio between the scale range where the *in situ* well-log measurements are taken and the scale range inhabited by the seismic waves.

E.3 Multiscale characterization in relation to scattering and localization theory

To investigate the added value of multiscale analysis, let me reiterate the main findings which came up while conducting a local/global multiscale analysis on well data by means of the continuous wavelet transform. It appeared that the multiscale analysis expresses information on

- the local regularity in terms of local Hölder exponents, see figure E.1 (c), (d), (e) and (f).
- the global scaling displayed by the partition function, see figure E.1 (g).
- the global regularity or the lack of it in terms of the α-axis of the singularity spectrum $f(\alpha)$, see figure E.1 (i).
- the Hausdorff dimensions[5] that can be assigned to the subset that scales according to the scaling exponents ranging from α to $\alpha + d\alpha$. In this way the definition of the Hausdorff measure becomes similar to the Lebesgue measure in case the Hausdorff dimension becomes integer valued (le Méhauté, 1991).

Now how can the above information be of assistance within the fields of scattering and localization theory? To answer this, let me briefly categorize the assumptions underlying these theories. Without being all inclusive I refer to

[5]The reader is referred to chapter 6 for an introduction to the concept of multifractals while in chapter 2 and chapter 8 I paid more attention to the actual application of this concept.

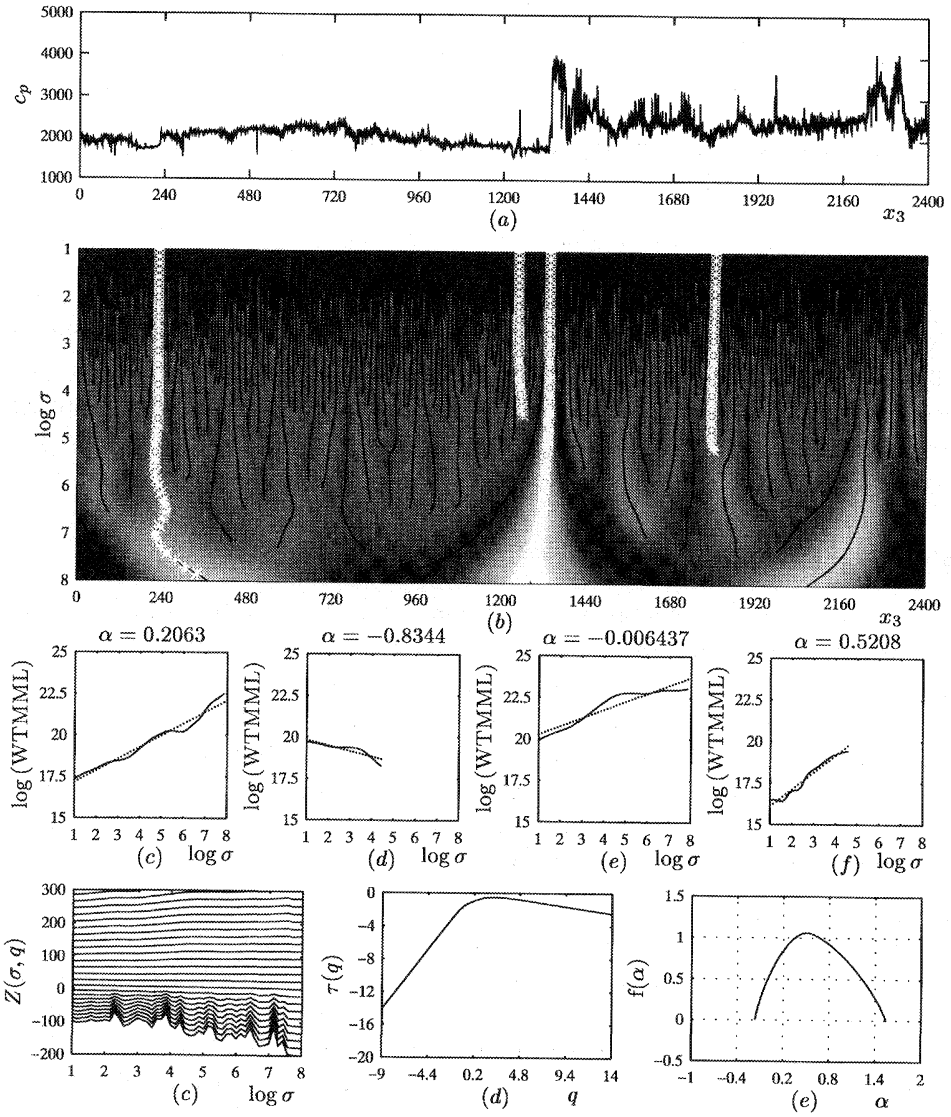

Figure E.1 *This plot illustrates the local/global multiscale analysis conducted on compressional wavespeeds. (a) the well-log. (b) the multiscale partitioning by the* WTMML*'s. (c)-(f) example of a local scale analysis by inspection of the* log *of the amplitudes along the annotated* WTMML*'s and versus the* log *of the scale, i.e.* $\text{WTMML}_\sigma \propto \sigma^\alpha$. *(g)-(i) example of a global multifractal analysis. (g) the partition function, i.e.* $Z_\sigma(q) = \sum_{m \in M} [\sup\{\text{WTMM}_\sigma(m)\}]^q \propto \sigma^{\tau(q)}$. *(h) the mass exponent* $\tau(q)$ *vs.* q *(i) the singularity spectrum* $\text{f}(\alpha)$.

- *a separation of scales.* This generally refers to a finite correlation length[6]. This assumption plays an important role in localization theory pertaining to random one-dimensional Schrödinger operators (Souillard, 1986; Carmona and Lacroix, 1990; Pastur and Figotin, 1991; Faris, 1995).
- *a sufficient regularity.* This refers to differentiability conditions while, say, defining the $\mathcal{H}_2$ operator, see chapter 4.
- *a boundedness.* This assumption plays a role in setting up a scattering, spectral and localization theory[7].
- *an integrability.* This condition refers to the notion that the medium profiles have to be locally or globally integrable.

Given this brief summary I am now able to issue in some more detail the previous assumptions in relation to the empirical findings of the multiscale analysis. In this discussion I will not come up with any definitive statements concerning the applicability of certain theories. This is not to be expected, since it is virtually impossible to disprove a theory with itself.

E.3.1 Separation of scales

The absence or presence of a break in the scaling, e.g. a break in the powerlaw behaviour of the partition function, forms as far as I know the most important prerequisite when dealing with wave phenomena. In the context of localization theory the separation of scales generally refers to a finiteness of the correlation length. The most commonly used random process displaying such a behaviour is the first order Markov process or the random telegraph model, see chapter 2, both of which share the property of having an exponentially decaying covariance function despite the fact that the way in which they obtain their randomness is quite different.

Judged by the findings of the global multiscale analysis it is difficult to reconcile the displayed behaviour of the partition function with an exponential decay of the covariance function. The evidenced powerlaw behaviour displayed by the partition function points more in the direction of a highly *intermittent* and *stochastically non-stationary* process. More specifically this means that the partition function does ***not*** behave like $Z\{f,\psi\}(\zeta,2) \sim (1-e^{-\frac{\zeta}{\zeta_c}})$, i.e. as a stationary process with an exponentially decaying covariance function and with a correlation length $\zeta_c \neq \infty$. Quite the contrary, the partition function displays a powerlaw type of behaviour and it does this not only for $q=2$ but for a whole range of q values. In figure E.1 I included a summarizing plot of the local and

[6] The correlation function is assumed to be exponentially decreasing, the decay rate being characterized by the correlation length.

[7] Note that Wilcox (1984) remarks in his book that the boundedness condition can be removed, without showing this rigorously.

global multiscale analysis conducted on a compressional wavespeed profile assembled by a well-logging tool. One can see from the partition function that there is at least for the specified scale range no direct indication for an exponential behaviour and there is no apparent break of the scaling evidenced over this interval.

It appears that the assumption of a separation of scales is of profound importance when pursuing a better comprehension of wave dynamics. As I will show in the sequel this does not only apply to localization theory, but also to homogenization theory and other types of asymptotic approaches (Asch et al., 1991, 1990; Papanicolaou et al., 1990; Burridge and Chang, 1989; Burridge et al., 1992; Pastur and Figotin, 1991; Lawecki et al., 1994; Lawecki and Papanicolaou, 1994) based on a small parameter that expresses a possible separation of scales. As a matter of fact, the required separation of scales applies to a condition on the degree of stochastic stationarity. It happens to be that the degree of non-stationarity, H, constitutes one of the ruling parameters in the multifractal parameterization, the value of which was empirically found to be $H \simeq 0.21$. From this empirical finding one may conclude that

- the medium's sample statistics are not translationally invariant (Pastur, 1994) in the mean, i.e. the mean is not a conserved quantity.
- the correlation does not vanish at infinity (Pastur, 1994).

In case of long tailed correlation – as is evidenced during multiscale analysis – the existence of the singular continuous part of the spectrum can not be excluded (Pastur, 1994; Pastur and Figotin, 1991). These observations do not strive well with the supplemented conditions used to set up the theory for random Schrödinger or wave operators. Apparently, one has in my case to do with a complexity in the medium properties on which very little is known. Indeed there are already some precursors on the advent of a more generalized theory amongst which the generalized Laplacian I treated in chapter 4 (le Méhauté, 1991). But still little is known about this type of operator (Carmona and Lacroix, 1990).

E.3.2 Regularity and boundedness

Differentiability and the congruent notion of *rectifiability*, see chapters 2 and 8, constitute important preconditions for deriving laws of physics (le Méhauté, 1991, 1995; Nottale, 1992, 1995, 1996), see also section E.6. On first sight the notion of differentiability and continuity seem to be void of a physical meaning. By this I mean that these conditions apply to the mathematical construct of a function that strict locally assigns a number to a point. This notion is difficult to reconcile with a physical interaction that can not occur on an infinitely small scale. However, when interpreted properly within a physical context, these concepts certainly have an important meaning.

Within the realm of the current wave theory itself it is difficult and maybe even impossible to put one's finger on the exact spot where a possible regularity condition is jeopardized. To put it simple, I believe one is already too late when worrying on the differentiability within the representation of the wave equation itself. With this I mean that this type of question can be better asked at the level of the constitutive relations, rather than at a subsequent level of equations representing the wave motion.

As I will show later, the singularity spectrum contains more information but for the time being let me assume that I have to do with one isolated singularity, so $\mathrm{f}(\alpha) = 0$, since the dimension of the singular support is zero. The strength of this singularity is α. Now what can one do when the wave dynamics are of concern in a medium containing one such singularity, i.e. the medium is defined in terms of a homogeneous distribution? For $\alpha = 0$, a jump discontinuity, one can find a solution by imposing boundary conditions. These boundary conditions refer to an imposed continuity of the wavefield constituents and give rise to the definition of *local specular reflection* and *transmission* coefficients. Recently a similar road of attack has been followed by Wapenaar (1996a) who derives expressions for the local reflection and transmission coefficients at a compressibility profile given by a homogeneous distribution with $\alpha \neq 0$, thereby assuming that boundary conditions can be used in a similar manner when dealing with a jump discontinuity.

Another aspect which comes forward during the theory is the boundedness. Therefore let me review the potential input delivered by the global multifractal characterization when applied to the notion of boundedness that comes with the theoretical treatments of scattering, spectral theory and localization theory pertaining to the $\mathcal{H}_2$-operator. This boundedness condition refers to the boundedness conditions of the medium properties, the density of mass and compressional wavespeed in this case (Reed and Simon, 1978, 1979; Wilcox, 1984; Carmona and Lacroix, 1990; Dautray and Lions, 1992), i.e.

$$0 < \rho_0 \leq \rho(x_3) \leq \rho_1 < \infty, \quad 0 < c_0 \leq c(x_3) \leq c_1 < \infty, \quad c, \rho \in L^\infty(\mathbb{R}).$$

Note that these "functions" $\rho(x_3)$ and $c(x_3)$ are obtained by coarse graining (see the discussion around equation (2.54)). On first sight these conditions seem not to be too demanding but when having second thoughts one may come to the conclusion that the multifractal behaviour with its generalized Hausdorff dimension exceeding zero, may jeopardize this presupposition in the sense that the set exceeding a threshold lies *dense* in the multifractal situation. This notion is illustrated in figure 2.7. Theoretically this notion lies in the fact that the probability of exceeding the threshold σ^α scales as, see chapter 2 and equation (2.29) also,

$$\Pr(f(\sigma, \cdot) \geq \sigma^\alpha) \sim \sigma^{D-\mathrm{f}(\alpha)},$$

where $f(\sigma, \cdot)$ denotes a coarse-grained quantity, $\mathrm{f}(\alpha)$ the Hausdorff dimension and D the embedding dimension.

To put it simply, one may draw the conclusion that apparently the maximum value for say the compressional wavespeeds depends on the scale and irrespectively of the value for the threshold σ^{α}, this threshold will be surpassed infinitely many times. The set transgressing the threshold lies *dense* and has a co-dimension that is non-zero, yielding $f(\alpha) > 0$. As a matter of fact the coefficients are singular with respect to the Lebesgue measure, as the monofractal Cantor set is, see chapter 6 and Reed and Simon (1980), since the Hausdorff measure does not equal the Lebesgue measure in case the dimension is non-integer.

Another way of looking at the boundedness condition is to take into consideration that the partition function and its scaling provide information on the L^{∞}-norm. To see this remember that the partition function is defined along the lines

$$Z\{f,\psi\}(\sigma,q) = \int |f(\sigma,x_3)|^q \mathrm{d}x_3 \sim \sigma^{\tau(q)},$$

whereas the L^q-norm for f is defined as

$$\|f\|_q = \left(\int |f(x_3)|^q \mathrm{d}x_3 \right)^{\frac{1}{q}}.$$

Clearly this suggests that a possible divergence for the partition function, depending on the behaviour of $\frac{\tau(q)}{q}$ as $q \to \infty$ and $\sigma \to \infty$ or $\sigma \downarrow 0$, may imply a lack of boundedness.

To summarize, it is my opinion that it is fair to say that it is difficult to reconcile the boundedness conditions with fractal scaling. That is to say that for example the maximum value for the medium coefficient profiles, defined in terms of well data, depend on the scale at which they are observed. Of course at the finest grained scale level, yielded by the measurement, one can always assign a maximum value but that value potentially changes when the well-logging tool would have acquired the data at a finer scale level. For that reason it is my conjecture that these empirical observations, always relating to finite scale ranges, already have a meaning. It is hard to prove this conjecture but I am convinced it is a worthwhile observation in the discussion, certainly given the experiences I had when conducting a multiscale analysis with erroneous analyzing wavelets, see chapter 6 and 8.

Let me conclude this discussion by posing the question on what happens in case the boundedness violation is committed by a set that does not lie dense, say, an isolated singularity of the algebraic type. That is exactly the case for which Wapenaar (1996a) derived closed form expressions for the induced local reflectivity and transmittivity. Clearly the dimension of the singular support is zero in this case. Consequently one can argue that this notion offers a road of escape because the set violating the boundedness conditions is of measure zero likewise the delta distribution. Now what happens when one has to do with more than one singularity? Say two. In that case one has – as with the singularities that lie dense – to do with mutual interferences of the singularities. This

interference causes a drastic transition in the scaling behaviour as soon as the cones of influence pertaining to the singularities start to overlap. In case of more singularities the same observations can be made.

E.3.3 Intermezzo on the use of mathematics in relation to physics

What strikes me first, is that the knowledge and comprehension on waves in heterogeneous media is still relatively limited, see for instance Faris (1995). Besides this observation there is another observation which addresses the unavoidable controversy that comes with issuing more mathematically oriented statements based on empirical findings, which are obtained from real measurements. This controversy is due to the inherent ambiguity surrounding the measurement process withstanding unambiguous statements on say the data's differentiability. At this point one enters the delicate issues physicists have to deal with when selecting or deriving mathematical representations to be appointed to describe "real" physical phenomena. While doing so, it is my opinion that the selection procedure is preferably not void of an empirically driven input, hopefully yielding a good match between the representation and the data's behaviour within the scale range of interest. Eventually, when this choice is made then for mathematical consistence, one has, in my opinion, to stick to the representation in the sense that the mathematics takes it from there on, and one has to accept the possible consequences.

The multiscale analysis and subsequent characterization suffer from the same problem as above. This is because of the finiteness of the scale range over which information about the medium properties is available. On the other hand, the data suggest that the medium properties display a highly irregular behaviour over a wide scale range being captured elegantly by the singularity spectrum. This information can be put to one's advantage in two ways. First of all, it gives a handle to explore whether the current formulation for the wave motion is equipped to deal with the medium's heterogeneity. Secondly, the observation of scaling and its implied scaling representation sparks a discussion on the justifiability of the assumptions being made while deriving the constitutive relations and using homogenizability.

E.4 The constitutive relations again

In chapter 4 of the Capita Prima I introduced the *constitutive relations* without paying ample attention to the assumptions under which they have been derived. The purpose of this section lies in shedding some new light on this interesting but delicate issue. I have used the recent book by de Hoop (1995) as a primary source.

Let me first restate the two ruling constitutive relations that are essential for setting up the equations for acoustic wave motion. The first constitutive relationship expresses the

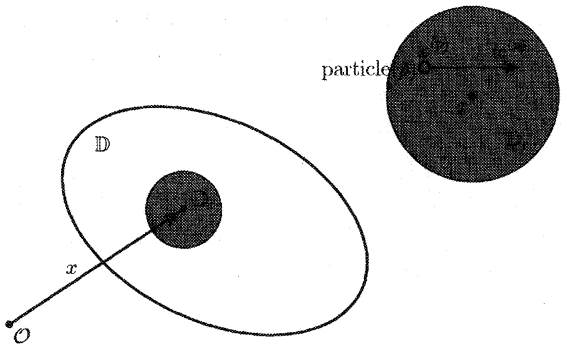

Figure E.2 *A bounded domain $\mathbb{D}$ containing a collection of particles or heterogeneities. $\mathbb{D}_\varepsilon$ is a time and translation invariant representative elementary volume that contains many particles/small enough size heterogeneities.*

mass flow density rate, in subscript notation,

$$\dot{\Phi}_k(\mathbf{x},t) = \rho(\mathbf{x}) D_t v_k(\mathbf{x},t), \tag{E.1}$$

whereas the second

$$\dot{\Theta}^i(\mathbf{x},t) = -\kappa(\mathbf{x}) D_t p(\mathbf{x},t) \tag{E.2}$$

delineates the *induced cubic dilatation rate*. The $D_t = \partial_t + v_l \partial_l$ denotes the co-moving time derivative with v_l being the macro averaged drift velocity. In the subsequent discussion *I will limit myself to the constitutive relation for the mass flow density rate* and I will assume the fluid to be linear, time invariant, isotropic[8] and relaxation free.

Now what are the main steps and basic assumptions that led to the *local* result depicted in equation (E.1)? To answer this question I am obliged to go through some of the steps I perceive as being important and that admitted the establishment of the constitutive relationship for the *mass flow density rate.*

What strikes me most in establishing the main results is the definition of a time and shift invariant and co-moving *representative elementary domain* $\mathbb{D}_\varepsilon(\mathbf{x},t)$, see figure E.2, of size ε, with its center located at $\mathbf{x}$ and with a volume equal to $V_\varepsilon = \int_{\mathbb{D}_\varepsilon} \mathrm{d}V$. This *representative elementary domain* is assumed to be small compared to the characteristic *macroscale* but large enough to contain many *microscale* heterogeneities/particles[9]. The notion of

[8] I assume the density fluctuations to vary stochastically isotropically in all coordinate directions.

[9] I see the heterogeneities as frozen turbulence like clusters of different sizes made of conglomerates of particles. These clusters may on their turn, when small enough, be considered as particles as compared to the size of the *representative elementary domain.*

being small compared to the macroscale not only makes reference to the scale range corresponding to the geometrical dimensions of the macro system being studied, but also to the scale range at which the *macroscopic transients show spatial variations* (de Hoop, 1995)! The second presupposition – the *representative elementary domain* must be large compared to the characteristic size of the heterogeneities in order to host very many of them – makes reference to the fact that one is able to designate the appropriate averages. These averages run over the characteristic size of the *representative elementary domain* and yield the appropriate macroscopic quantities that are *stationary*[10] in their *mean* on the macroscale and that do not depend on:

- *shifts* over the characteristic size of the *representative elementary domain.*
- changes in the *resolution/scale*, i.e. changes in the characteristic size of the *representative elementary domain.*

Subsequently it is assumed that the macroscopic quantities vary *piecewise continuously* with position and this is called the *continuum hypothesis*, see de Hoop (1995). As to live up to the continuum hypothesis there must be a distinct *separation of scales*, see also Auriault (1991), in order to guarantee the required stationarity in the mean over the macroscopic scale range. This latter notion constitutes *the main prerequisite for the continuum hypothesis.* With this remark I am not implying that one has to drop the notion of continuity. I only want to say that the continuum hypothesis of above may not longer be universally applicable without making specific reference to the scale, for those cases where a separation of scales is infeasible.

For the time being let me suppose that the required separation of scales is *feasible*, then one can show that the following crucial identity holds (de Hoop, 1995)

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbf{x}\in\mathbb{D}(t)} n(\mathbf{x},t)\Psi(\mathbf{x},t)\mathrm{d}V = \int_{\mathbf{x}\in\mathbb{D}(t)} n(\mathbf{x},t)D_t\Psi(\mathbf{x},t)\mathrm{d}V, \tag{E.3}$$

which represents an instance of *Reynolds' transport theorem* (Pierce, 1981).

In this global equality the $\mathbb{D}(t)$ denotes a bounded domain consisting of *representative elementary domains*, $\mathbb{D}_\varepsilon$, $n(\mathbf{x},t)$ is the particle density and $\Psi(\mathbf{x},t)$ represents an arbitrary macroscopic associated quantity whose value is obtained via averaging over the elementary volumes $\mathbb{D}_\varepsilon$. By defining the bounded domain, de Hoop (1995) is able to compute the total number of particles

$$N(t) = \int_{\mathbf{x}\in\mathbb{D}(t)} n(\mathbf{x},t)\mathrm{d}V, \tag{E.4}$$

[10]Statements with respect to the stationarity in this context do not refer to stationarity in the temporal sense. They refer to stationarity for the spatial fluctuations in the statistical sense.

where the integral runs over the different representative elementary subdomains $\mathbb{D}_\varepsilon$, that are contained within the bounded domain $\mathbb{D}(t)$. On their turn these representative elementary subdomains define the macroscopic number density $n(\mathbf{x},t)$ via

$$n(\mathbf{x},t) \triangleq \frac{N_\varepsilon(\mathbf{x},t)}{V_\varepsilon} \tag{E.5}$$

with $N_\varepsilon(\mathbf{x},t)$ being the number of particles within a individual elementary representative domain. Because of the elementary representative domains one is able to define the macro quantity $n(\mathbf{x},t)$ as a piecewise continuous function of $\mathbf{x}$. Notice that de Hoop (1995) does not make specific reference to the scale ratio between $\mathbb{D}_\varepsilon(\mathbf{x},t)$ and $\mathbb{D}(t)$. This observation applies also to the derivations supplied by Pierce (1981) and Wapenaar and Berkhout (1989), who follow a Eulerian rather than a Lagrangian approach, i.e. they keep the elementary domain at a fixed location. Finally remark that de Hoop (1995) is able to refer to macroscopic quantities that are either being obtained via averaging over $\mathbb{D}_\varepsilon(\mathbf{x},t)$ or $\mathbb{D}(t)$ because of the assumed *separation of scales* underlying the *continuum hypothesis*. That is the reason why the subscript ε has been dropped!

The importance of equation (E.3) lies in the notion that it relates the time rate of change at time instant t of the macroscopic/averaged quantity on the left to an expression where the time rate of change is found via the action of the co-moving time derivative on the associated macro quantity $\Psi(\mathbf{x},t)$ solely. So the local time change in the number density does enter into the formulation. Now what are the consequences of the identity depicted in equation (E.3)? Suppose the associated quantity represents the mass of the particles[11] then one finds

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbf{x}\in\mathbb{D}(t)} \rho(\mathbf{x},t)\mathrm{d}V = 0 \tag{E.6}$$

where $\rho(\mathbf{x},t) = n(\mathbf{x},t)m$ with m the mass of the particles. This identity is known as the *conservation law of mass* (de Hoop, 1995) since $\Phi_k(\mathbf{x},t) \triangleq \rho(\mathbf{x},t)v_k(\mathbf{x},t)$, the *mass flow density*, and

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbf{x}\in\mathbb{D}(t)} n(\mathbf{x},t)\Psi(\mathbf{x},t)\mathrm{d}V = &\int_{\mathbf{x}\in\mathbb{D}(t)} \partial_t[n(\mathbf{x},t)\Psi(\mathbf{x},t)]\mathrm{d}V \\ &+ \int_{\mathbf{x}\in\delta\mathbb{D}(t)} n(\mathbf{x},t)v_k(\mathbf{x},t)\Psi(\mathbf{x},t)\mathrm{d}A_k.\end{aligned} \tag{E.7}$$

Here the associated quantity is taken to be equal to the mass of the particles, while $\delta\mathbb{D}(t)$ and $\mathrm{d}A_k$ refer to the boundary and a surface element of $\mathbb{D}(t)$ respectively. Equation (E.6) is valid in the absence of annihilation and creation of particles and the *conservation law of mass* becomes manifest

$$\int_{\mathbf{x}\in\mathbb{D}(t)} \partial_t\rho(\mathbf{x},t)\mathrm{d}V + \int_{\mathbf{x}\in\delta\mathbb{D}(t)} \Phi_k(\mathbf{x},t)\mathrm{d}A_k = 0 \tag{E.8}$$

[11] For now assume there is only one type of particle.

which, after applying Gauss' divergence theorem, can be defined in its local form as

$$\partial_t \rho + \partial_k \Phi_k = 0 \tag{E.9}$$

and is known as the *continuity equation of mass flow.*

How did one arrive at this result? Or more importantly what was additionally required, besides the absence of creation and annihilation, to arrive at this conservation law? The answer to this lies in the *continuity equation for particle flow* which is required to derive the Reynolds' transport theorem for the particle density[12]. This continuity equation reads

$$\partial_t n + \partial_k (n v_k) = 0 \tag{E.10}$$

and expresses the *conservation of particles,*

$$\int_{\mathbf{x}\in\mathbb{D}(t)} [\partial_t n + \partial_k (n v_k)] \mathrm{d}V = 0 \tag{E.11}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbf{x}\in\mathbb{D}(t)} n(\mathbf{x}, t) \mathrm{d}V = 0 \tag{E.12}$$

with (nv_k) being spatially continuously differentiable in $\mathbb{D}(t)$. Equation (E.11) is known as the *conservation law of particle flow.*

E.4.1 The "weak" spot?

Besides the smoothness condition, which is required by Gauss' divergence theorem and potentially may be removed, what else is assumed to arrive at equations (E.10) and (E.9)? Or in other words, at which instance a crucial step is made which on second thought might be the "weak" spot despite the fact that the derivation of the above conservation laws has been perfectly compliant with the continuum hypothesis assumed at the beginning. I suspect the approximate identity (de Hoop, 1995)

$$\int_{\mathbf{x}\in\mathbb{D}(t+\Delta t)} \partial_t n(\mathbf{x}, t) \mathrm{d}V = \int_{\mathbf{x}\in\mathbb{D}(t)} \partial_t n(\mathbf{x}, t) \mathrm{d}V + o(1) \tag{E.13}$$

as $\Delta t \to 0$ to be possibly the crucial step. It refers namely to the notion that the temporal rate of change for the particle density is *insensitive* to an infinitesimal time step Δt made by the bounded domain $\mathbb{D}(t)$. With other words the time rate of change in the number density is independent of spatial changes in the bounded domain $\mathbb{D}(t)$ that may occur during the time interval Δt, hence the temporal and spatial changes are *decoupled* at this level.

[12] Notice that the particle density can be replaced by the mass density.

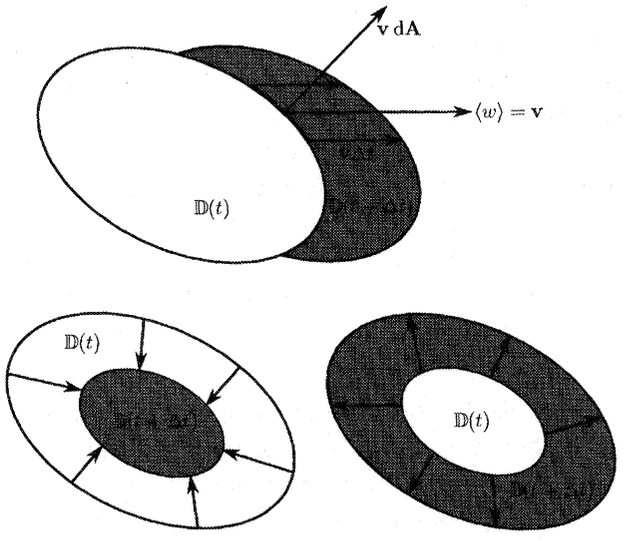

Figure E.3 *During a time lapse Δt the domain $\mathbb{D}$ is not only allowed to translate but also to dilatate.*

This decoupling is effectuated despite the fact that the time interval Δt governs both the time increment inducing the spatial increment of the bounded domain as well as the time increment in the temporal derivative and in the end it led to the identity stated in equation (E.3). Indeed the above identity is perfectly defendable when the continuum hypothesis is adhered to by the bounded domain $\mathbb{D}(t)$. With this I mean that one can – besides a translation by an amount $\mathbf{v}\Delta t$, see figure E.3 – also associate a *dilatation* to the *bounded domain*, $\mathbb{D}$, during the time step. But the implication of this dilatation seems to be irrelevant when the continuum hypothesis is invoked at the scales that extend *at least* the scale range separating $\mathbb{D}_\varepsilon$ and $\mathbb{D}(t)$. Finally remark that a similar line of reasoning can be used for the Eulerian approach instead of the Lagrangian approach.

Now the question of course is whether the continuum hypothesis is defendable for this prescribed scale range when one has to do with a medium exhibiting heterogeneities on "all" scale ranges. Please let me perhaps superfluously mention that I do not imply to say, when referring to the continuum hypothesis, that physical phenomena are not continuous. By this I mean that I am convinced that a physical interaction must always go along with a kind of coarse-graining, invoking the necessary *regularization.* This regularization smoothes away the singular behaviour, i.e. the scale divergence, see chapter 2.

Fractals, or more generally constructs[13] displaying some type of scale divergence, are mathematical entities, that display heterogeneities on all scale ranges. But as I mentioned already elsewhere, this property does not imply that these constructs actually exist. They only capture a certain behaviour empirically evidenced within a certain scale range and are, consequently, used to mathematically represent this complexity. Ordinary smooth functions do exactly the same thing in the sense that they do not either pretend to be actually existing. They only provide an adequate representation for phenomena that display a regular behaviour on the scale range of interest. Finally remark that the implication of introducing constructs with a scale divergence is, that one is replacing functions by functionals to represent the medium's complexity, i.e. one is transgressing to the scaling medium representation.

Suppose one deals with the interesting situation where there is a strong scale dependence, i.e. there is no separation of scales. What is going to happen with the above line of reasoning? That is to say, what are the implications when the stationarity condition for the mean does not longer hold? In that case I expect specular reflections to emanate[14], an observation one is all too familiar with given the emergence of reflections at jump discontinuities occuring in the medium properties.

Specular scattering is generally understood via imposing boundary conditions. But would it not be interesting to address the question what happens during an excursion out of the continuum hypothesis? To put it simply, I believe one runs the risk of jeopardizing the founding principles of the continuum hypothesis, that allowed for the identity in equation (E.13). In case one permits an abundance of jump discontinuities and/or other type of singularities to enter the macroscale, the risk is that the separation of scales no longer holds. This may manifest itself in a scale dependence of the averaging process that yields the macroscale quantities such as the number density. It may also have its consequences on the averages over the bounded domain $\mathbb{D}(t)$ that may change in size during the time step Δt. This latter observation applies to the fact that when adding detail on the intermediate scale range, between the elementary representative volume size and the spatial extent of the bounded domain, there is a possibility that the averages over this bounded domain start to depend on its size and that withstands the validity of the identity in equation (E.13). Of course this would be the case when the supposed scale dependence exceeds $o(1)$. To summarize, detail on scales exceeding the characteristic size of the representative elementary domain may endanger the assumptions that facilitated the introduction of the scale independent averaging and its consequences.

[13] Remark that these constructs are generally generated by ordinary smooth functions in a delta convergence series type of approach. This means that the scaling behaviour depends on the scale of observation, see the discussion in chapters 2, 5 and 6.

[14] In that situation I expect the dependence on the size of the bounded domain to be dominating over the $o(1)$ in equation (E.13).

E.4.2 A scale derivative comes in?

Why is the above discussion so important? The reason is that my main interest lies in improving my comprehension of the dynamics that come with the *"local"*[15] wave interaction process that occurs on the macroscopic scale range. This interaction process, also known as a *local* scattering process, is responsible for the information transfer from the medium heterogeneity to the probing wavefield. Now what happens when the macroscopic averages start to depend on the scale in such a way that averages over the bounded domain are going to share this property? Or alternatively what happens if one adds so much detail in the intermediate scale range between $\mathbb{D}_\varepsilon$ and $\mathbb{D}(t)$ that the latter starts to depend on the scale? In this situation I like to postulate that the equality of equation (E.13) needs a re-evaluation. This is because an effective decoupling has been effectuated on what may happen during the time interval, Δt, with the temporal change rates of the macroscopic quantity, say, the number density and a possible scale change of the bounded domain $\mathbb{D}(t)$, all taking place during that same time interval and on equal footing. For example one can envisage the elementary domain to translate or dilatate. To illustrate what might happen during the dilatation let me maintain references to the size of the representative elementary domains. I assume the *macroscopic number density* as defined in equation (E.5) to explicitly depend on ε. Then one may – during the time lapse Δt, the *temporal gauge* – expect a contribution emanating from a possible dilatation effectuated by the change from $\mathbb{D}(t)$ to $\mathbb{D}(t+\Delta t)$. To demonstrate this let me shift to a notation more in line with the multiscale analysis. For simplicity I will treat the equations as if I am dealing with one coordinate direction x only. Furthermore I will use the familiar σ as the scale indicator referring to the size of the elementary interval in this case. Now, let me first rewrite equation (E.13) as

$$\langle \partial_t n(x,t), \phi_{\sigma(t+\Delta t),x} \rangle = \langle \partial_t n(x,t), \phi_{\sigma(t),x} \rangle + o(1), \tag{E.14}$$

where $\Delta t \to 0$. In this identity I attributed the "size" of the bounded domain to $\sigma(t)$, the *spatial gauge* which depends on t! The angular brackets denote an inner product by a smoothing kernel. This smoothing kernel, indexed by the scale indicator $\sigma(t) \triangleq \mathbb{D}(t)$, is an element of the affine family $\phi_{\sigma,x}(x') = \frac{1}{\sigma}\phi(\frac{x'-x}{\sigma})$ and acts on the fine-grained number density. To start with I choose the indicator function as the smoothing kernel $\phi(x')$.

The question is now, how does this equality behave as a function of the temporal gauge? To answer this question one has to know the relation between the temporal gauge Δt and the change in the spatial gauge, i.e $\Delta\sigma(t,\Delta t) \triangleq \sigma(t+\Delta t) - \sigma(t)$. For the time being let me assume that $\Delta\sigma(t,\Delta t) \to 0$ when $\Delta t \to 0$. Under this assumption one can recognize a scale derivative to enter into the formulation in a manner similar to the one I reviewed in chapter 2. There it was shown that the difference between two consecutive smoothings – obtained via two subsequent convolutions with two mutually infinitesimally dilatated

[15] What the wave experiences as local.

smoothing kernels with scale index $\sigma - \mathrm{d}\sigma$ and $\sigma -$ yielded a natural definition of the continuous wavelet transform, i.e. under certain conditions for the smoother,

$$\begin{aligned} \mathcal{W}\{f,\psi\}(\sigma,x) &= -\sigma\partial_\sigma\langle f,\phi_{\sigma,x}\rangle \\ &= \langle f,\psi_{\sigma,x}\rangle. \end{aligned} \tag{E.15}$$

In order to preserve the mean, I use an L^1-normalization for the smoothing kernel instead of an L^2-normalization. The wavelets are given by the self-affine family $\psi_{\sigma,x}(x') = \frac{1}{\sigma}\psi(\frac{x'-x}{\sigma})$ and they are related to the family of smoothing kernels via $\psi(x') = (x'\partial_{x'} + 1)\phi(x') = j\mathcal{C}_x\phi(x')$ where, due to the L^1-normalization, $\mathcal{C}_x$ has been defined differently compared to the definition in chapter 3.

Application of this definition to the approximate identity of equation (E.14) shows that

$$\langle\partial_t n(x,t),\phi_{\sigma(t+\Delta t),x}\rangle - \langle\partial_t n(x,t),\phi_{\sigma(t),x}\rangle = -\frac{\Delta\sigma(t,\Delta t)}{\sigma(t)}\langle\partial_t n(x,t),\psi_{\sigma(t),x}\rangle. \tag{E.16}$$

What is the meaning of this equality? As in chapter 2 it represents the *detail* of $\partial_t n(x,t)$ *at* the scale $\sigma(t)$ whereas the quantity $\langle\partial_t n(x,t),\phi_{\sigma,x}\rangle$ contains *all* details *up to* the scale σ.

What is the consequence of this observation for the conservation law of particle flow in case the scale derivative exceeds $o(1)$ in the limiting process for $\Delta t \to 0$? Clearly this will only happen in those situations where the wavelet coefficients, $\langle\partial_t n(x,t),\psi_{\sigma(t),x}\rangle$, dominate over the $o(1)$ term. So in my conjecture, equation (E.11), the conservation law of particle flow, becomes for this situation

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle n(x,t)\phi_{\sigma(t),x}\rangle = \lim_{\Delta t\to 0} -\frac{\Delta\sigma(t,\Delta t)}{\sigma(t)}\langle\partial_t n(x,t),\psi_{\sigma(t),x}\rangle. \tag{E.17}$$

The careful reader may have noticed that as a consequence of applying Gauss' divergence theorem one has to impose a differentiability condition on $n(x,t)v_x(x,t)$. This condition can, however, be removed by selecting a proper smoothing kernel and differentiating in the sense of distributions. More important is that in case of no scaling the right hand side of equation (E.17) vanishes, yielding equation (E.11) again.

To conclude, let me emphasize again that this is not more than a premature conjecture, but it is still worthwhile and interesting to explore its possible consequences. Before doing that, I think it is better to first link up with the ideas and findings of homogenization theory followed by postulating a candidate for the spatial gauge. Given this gauge it will be possible to reflect more on a possible dynamic homogenization which in my opinion is more or less equivalent to setting up an alternative representation for the wave dynamics.

E.5 Merits and limitations of the current formulation and its homogenization

In this section I will try to convey my comprehension on wave interactions in media that display a highly irregular behaviour in their medium properties. I am aware of the reflective nature of the ideas I am about to share with the reader. They merely serve as a starting point for a discussion on a potential alternative approach tackling the wave problem for media that fluctuate erratically over a wide scale range. If in the end this type of approach is failing then still I hope that this discussion will contribute to a possibly better comprehension of wave interactions in these type of media.

One way of expressing the question I am asking myself, refers to the underlying reason for the inexistence of a *dynamic* homogenization theory, that, *from first principles*, relates the wave dynamics to the medium's complexity. Indeed, it can be demonstrated that only a *static* homogenization procedure[16] exists that prescribes the way in which one has to average the constitutive parameters in order to describe the behaviour of the response to an excitation living on a macroscale. Unfortunately this approach can not be extended to a dynamic one while preserving the original time-space structure of the wave equation. Moreover, it is limited in its application, since it requires a separation of scales. If this separation of scales is not being adhered to then, the homogenized system yields a behaviour that is static at the first order of magnitude (Auriault, 1991). This does not imply that there are no alternative methods – for example the O'Doherty-Anstey formula (O'Doherty and Anstey, 1971; Banik et al., 1985a,b; Resnick et al., 1986; Burridge et al., 1988; Burridge and Chang, 1989; Burridge et al., 1993; de Hoop et al., 1991b,a; Shapiro and Zien, 1993; Shapiro et al., 1994) or the order $\mathcal{O}(1)$ approach by Asch et al. (1990, 1991) and others (Papanicolaou et al., 1990; Burridge and Chang, 1989; Burridge et al., 1992; Lawecki et al., 1994; Lawecki and Papanicolaou, 1994) – but these are difficult, if not impossible, to link up with the wave equation in its original form. Moreover these approaches presuppose either the medium contrasts to be small or the pulse width to be large compared to the correlation length of the medium fluctuations and small compared to the propagation distance. Needless to say these presuppositions are difficult to reconcile with the multiscale findings. However, in an approximative manner links still can be established between the O'Doherty-Anstey framework and an acoustic wave equation with an anisotropic viscous relaxation mechanism (Wapenaar et al., 1994).

The set up of this section will be as follows. I will commence by reviewing some of the notions that come with homogenization theory. The similarity between the line of reasoning within this theory and the one within the derivation of the constitutive relations themselves is explained next. Then I will make an attempt to clarify what the homogenized wave equation exactly describes. That brings me to a discussion on how I

[16] That is a formulation of the wave dynamics whose action is invoked on the same scale range as where the wave interactions are believed to take place. As a matter of fact it is my personal view that the formulation of the wave motion in itself deserves a similar interpretation, a notion I will pay ample attention to later.

interpret linear partial differential equations and that leads me back to a reflection on the possible modifications at the level of the constitutive relations in sections E.6 and E.7.

E.5.1 Basic idea of homogenization

As coined by Auriault (1991), one of the basic assumptions, when investigating a heterogeneous medium at the macroscale, refers to the existence of an equivalent medium. Such an equivalent medium corresponds to a continuous replacement medium, yielding the average behaviour of the medium when submitted to an excitation that lives on the macroscopic scale. In this, the macroscale refers to a scale where the volume contains a large number of microscale heterogeneities and the question of homogenization commits itself to seek an answer to the question how the microscopic information on the medium's heterogeneity is transferred to the macroscopic scales.

Examination of homogenization theory reveals a strong semblance, but now on a different level, with the line of reasoning behind the derivation of the constitutive relations. Again the crux lies in designating a *separation of scales* allowing for the definition of an *elementary representative volume* that is *small* compared to the macroscale but that is large enough to contain a large number of microscale heterogeneities. Obviously this implies a constraint on the characteristic size of the heterogeneities.

To make it more explicit, let l denote the characteristic length scale attributed to the elementary representative volume and L the characteristic macro length scale referring to a separation of scales for both the medium fluctuations as well for the transient phenomena under consideration, i.e. the spatial length scale induced by the excitation. The separation of scales implies that

$$\frac{l}{L} = \varepsilon \ll 1 \tag{E.18}$$

and forms a prerequisite in the construction of a homogenization process where the two length scales allow for a double scale asymptotic development. This asymptotic development refers to an examination of the unknown field quantities, say the acoustic pressure, in two different asymptotic spatial scale regimes namely

$$\lim_{\varepsilon \downarrow 0} \mathcal{D}_{\varepsilon^{-1}} p(\mathbf{x}, t), \tag{E.19}$$

delineating the macroscopic scale and

$$\lim_{\varepsilon \downarrow 0} \mathcal{D}_{\varepsilon} p(\mathbf{x}, t) \tag{E.20}$$

delineating the microscale. The operator $\mathcal{D}_\varepsilon$ denotes the dilatation operator, acting on the spatial coordinate, I introduced in chapter 3, but now with an L^1-normalization. In case a separation of scales exists, then one will find that the variations in the average over the macroscale will tend to zero as $\varepsilon \downarrow 0$. In other words, the averaged quantities

are *stationary* in the sense that their average is stationary, i.e. it is invariant under local translations of order l (Auriault, 1991). As a result of this acquired stationarity one can set up a perturbative series expansion for the unknown field quantity in terms of the small parameter ε, see for instance Asch et al. (1990) or Auriault (1991). During this process one can consequently make use of the above stationarity properties by establishing a link between the macro and micro variability. In this way one can solve the unknown field quantities asymptotically at the macroscale. At this point I will refrain from specific technicalities and I will limit myself to addressing the results of homogenization for the acoustic wave equation in a one-dimensional varying media only. In the pertaining sections I will put emphasis on the physical interpretation of the homogenized solutions and on the question of homogenizability.

E.5.2 Homogenization for the acoustic wave equation

As I mentioned before, homogenization is feasible in case the spatial scale range inhabited by the transient is well separated from that of the medium's heterogeneity. Let me designate, following Auriault (1991), the wavelength as a candidate for the macroscopic scale L and l the size of the periodic cell[17] or the correlation length. For more details on quasi static equivalent medium theory the reader is referred to Gilbert and Backus (1966), Schoenberg and Muir (1989), Asch et al. (1990) and Auriault (1991).

It appears that by setting up the homogenization procedure in the afore mentioned way, one is able to define an equivalent medium consisting of the proper averages effectuated on the micro structure. That is to say, that homogenization theory commits itself to describe how a fine-grained medium – as compared to the macroscopic scale of the transient – has to be averaged in order to describe its response to an excitation living at the macroscale. Let me consider the homogenization technique set to work on the acoustic one-dimensional wave equation when casted in the (τ-p)-domain, with p the horizontal slowness and τ the intercept traveltime. Consider the source free coupled system of partial differential equations

$$\partial_3 \mathbf{u}(x_3, t; \mathrm{p}) = -\mathbf{A}(x_3; \mathrm{p}) \partial_t \mathbf{u}(x_3, t; \mathrm{p}), \tag{E.21}$$

where the $\mathbf{u}(x_3, t; \mathrm{p})$ stands for the wave vector,

$$\mathbf{u}(x_3, t; \mathrm{p}) = \begin{pmatrix} p \\ v_3 \end{pmatrix} (x_3, t; \mathrm{p}) \tag{E.22}$$

and where $\mathbf{A}(x_3; \mathrm{p})$ represents the *system matrix* given by

$$\mathbf{A}(x_3; \mathrm{p}) = \begin{pmatrix} 0 & \rho \\ \kappa - \frac{\mathrm{p}^2}{\rho} & 0 \end{pmatrix} (x_3; \mathrm{p}). \tag{E.23}$$

[17] In case the microscale heterogeneities are taken to be periodic.

For convenience I use t rather than τ to indicate the intercept traveltime.

The homogenization procedure boils down to invoking a smoothing operation on the system matrix. From the deliberations of chapter 2 it is known that care must be taken while conducting such a coarse-graining operation, but for the time being assume that the characteristic function suffices as a proper smoothing kernel. Then coarse-graining gives rise to the following system

$$\partial_3 \mathbf{u}(x_3, t; \mathrm{p}) = -\mathbf{A}(L, x_3; \mathrm{p}) \partial_t \mathbf{u}(x_3, t; \mathrm{p}). \tag{E.24}$$

In this system the $\mathbf{A}(L, x_3; \mathrm{p})$ is obtained via

$$\mathbf{A}(L, x_3; \mathrm{p}) \triangleq \langle \mathbf{A}, \phi_{\sigma, x_3} \rangle|_{\sigma = L} \tag{E.25}$$

with ϕ_σ being the L^1-normalized indicator function. Under the condition $\frac{l}{L} = \varepsilon \ll 1$, the smoothed system matrix will vary slowly on the macroscopic scale range, i.e. it will vary slowly with respect to the wavelength $|\boldsymbol{\lambda}| = 2\pi/|\mathbf{k}|$ which is being taken as the spatial gauge. The vertical slownesses can be obtained from the coarse-grained system matrix by applying an eigenvalue decomposition in which they will occur as the eigenvalues (Wapenaar and Berkhout, 1989; Schoenberg and Muir, 1989; Folstad and Schoenberg, 1993; Kerner, 1992). It can be shown that the equivalent medium slownesses can directly be expressed in terms of the coarse-grained vertical and horizontal velocities and are given by

$$q_{eq}(L; x_3, \mathrm{p}) = \frac{1}{c_v(L; x_3)} \sqrt{1 - c_h^2(L; x_3) \mathrm{p}^2}. \tag{E.26}$$

The values for these coarse-grained velocities can be found explicitly via the inner products

$$c_v^2(L; x_3) = \frac{1}{\langle \kappa, \phi_{L, x_3} \rangle} \cdot \frac{1}{\langle \rho, \phi_{L, x_3} \rangle} \tag{E.27}$$

and

$$c_h^2(L; x_3) = \frac{1}{\langle \kappa, \phi_{L, x_3} \rangle} \cdot \langle \frac{1}{\rho}, \phi_{L, x_3} \rangle. \tag{E.28}$$

The question is now, what is gained and what is the physical interpretation to be given to this exercise? The answer to the first question is simple, namely the medium is coarse-grained to the macroscopic scale range – under the condition of a separation of scales – in a way that is consistent with the wave equation. The structure of this equation is preserved, but it went at the expense of inducing anisotropy, i.e. the medium becomes transversely isotropic – judged by the emergence of the two velocities – which in itself is a manifestation of the microstructure in a static way. Moreover the medium has been *regularized* in the sense that the coarse-grained medium inherits *regularity* from

the smoothing kernel. This depends on the regularity of the smoothing kernel and the singularity structure of the medium.

What else is gained? To answer this, one has to recognize that the equivalent slowness, as defined in equation (E.26), characterizes the group velocity as it is being *evidenced by a wave packet*[18] *with a dominant wavelength proportional to* L. The general expression for the *group velocity* in the x_3-direction is (Borowitz, 1967; Messiah, 1958)

$$c_g = \frac{\mathrm{d}\omega}{\mathrm{d}k_3} \tag{E.29}$$

or equivalently by

$$c_g = c_p - \lambda_3 \frac{\mathrm{d}c_p}{\mathrm{d}\lambda_3} \quad \text{or} \quad c_g = c_p + k_3 \frac{\mathrm{d}c_p}{\mathrm{d}k_3}. \tag{E.30}$$

The group velocity expresses the velocity with which a wave packet of non-vanishing support travels. In the case where the *dispersion relation*, equation (E.29), simplifies to a constant or in the case where the support of the wave packet vanishes, the *group* velocity becomes equal to the *phase* velocity which is given by

$$c_p = \frac{\omega}{k_3}. \tag{E.31}$$

As I mentioned, the dominant wavelength is taken as the spatial gauge yielding the following evaluation of the equation (E.29)

$$c_g(L; x_3, \mathrm{p}) = \left. \frac{\mathrm{d}\omega}{\mathrm{d}k_3} \right|_{\omega = \frac{2\pi}{Lq(x_3;\mathrm{p})}}, \tag{E.32}$$

where

$$q(x_3, \mathrm{p}) = \sqrt{\frac{1}{c^2(x_3)} - \mathrm{p}^2} \tag{E.33}$$

is the vertical slowness pertaining to the *local* fine-grained compressional wavespeed $c(x_3) = \sqrt{1/\rho(x_3)\kappa(x_3)}$.

For those situations where $\dfrac{l}{L} = \varepsilon \ll 1$, the macroscopic quantity $q_{eq}(L; x_3, \mathrm{p})$ is stationary and one is "perfectly" able to propagate the wave packet over many wavelengths, like in the geometrical optical regime, where the wavelength is much smaller than the scale at which the local velocity varies. This is all possible in the absence of specular scattering. The latter property is a direct consequence of the fact that one can not associate specular reflections to occur in a smoothly varying medium and that brings me to

[18] A wave packet delineates a group of waves whose mutual interference gives rise to a coherent structure that, contrary to ordinary plane waves, extends over a limited region in space and/or time.

the downside of the homogenization method. This downside points out the inability of homogenization theory to capture the dynamics that manifest itself in the emergence of specular reflection and dispersion. To answer the question whether there is a profound reason for this observation, I like to focus on those situations where homogenization is not feasible.

E.5.3 Homogenizability

The question of homogenizability commits itself to posing the question whether a certain system of pde's, say the system depicted in equation (E.21), can be homogenized in a proper way or not. It is shown by Auriault (1991) that homogenization is only feasible in those cases where there is a distinct separation of scales. For the acoustic wave motion this corresponds to the fact that the square of the ratio microscale versus macroscale must be of the same order as the square of the ratio microscale versus wavelength, i.e.

$$\mathcal{O}(\frac{l^2}{L^2}) = \mathcal{O}(l^2/|\boldsymbol{\lambda}|^2) = \mathcal{O}(\frac{l^2}{c_p^2 \Delta t^2}) = \mathcal{O}(\varepsilon^2)$$

where Δt is the characteristic time gauge being related to the wavelength $|\boldsymbol{\lambda}| = c_p \Delta t$. By way of the self-consistent construction of the homogenization it appears that no other choice, such as for instance $|\boldsymbol{\lambda}| = \mathcal{O}(L\sqrt{\varepsilon}) \ll L$, can be made (Auriault, 1991). Clearly this has profound consequences which can be interpreted in two different ways, namely

- in the opinion of Auriault (1991) it points to the fact that in cases where homogenization is unfeasible, one violates the basic assumption under which one sets out to probe the medium at the macroscale. This basically corresponds to a possible premature assumption on the existence of a separation of scales enabling the definition of an equivalent medium.
- another opinion where one may draw the conclusion that apparently the homogenization procedure only captures the static part of the wave problem and that one has to resort to other approaches in order to deal with the dynamics.

My opinion lies more in the line of Auriault (1991) in the sense that I experience difficulties associating induced *dispersion* and *specular reflection* to be governed by the wave equation. Before going into detail on that issue, let me first describe what homogenization does have to say even though it is being applied to cases where the homogenizability condition is not met.

What comes to mind first is that one wants to circumvent the requirement of a separation of scales. Or in other words how can one make sure that the condition $\frac{l}{L} = \varepsilon \ll 1$ always holds irrespective of a l, which is bounded from above? The answer is quite simple one sets $L \to \infty$ by setting $\Delta t \to \infty$ or equivalently $\omega \to 0$. In that infinitely long wavelength

regime the homogenization is always possible and this corresponds to replacing the inner products of equation (E.25) by a spatial average that runs over the propagation distance

$$\mathbf{A}_0(\mathrm{p}) \triangleq \frac{1}{x_3}\int_0^{x_3} \mathbf{A}(x_3', \mathrm{p})\mathrm{d}x_3'. \tag{E.34}$$

The propagation distance is measured with respect to the origin at 0 while the detector is situated at x_3. In this situation, the equivalent medium pertaining to this depth range becomes homogeneous. It can be proven that this homogeneous replacement medium describes the *centroid* for, say, the pressure wavefield detected at position x_3 and induced by a spatio-temporal Dirac distribution,

$$\tau_{cen}(x_3;\mathrm{p}) \triangleq \frac{\int_0^\infty t p(x_3;t,\mathrm{p})\mathrm{d}t}{\int_0^\infty p(x_3;t,\mathrm{p})\mathrm{d}t} \equiv \tau_{eq}(x_3;\mathrm{p}) \triangleq q_0(\mathrm{p})x_3, \tag{E.35}$$

with $q_0(\mathrm{p})$ being obtained from the system depicted in equation (E.34) and p being a solution to equation (E.21) with the fine-grained system. That is to say that the *centroid* belonging to the original system, equation (E.21), and the one obtained from the system matrix in equation (E.34) are equivalent[19] irrespective of the nature of the variability[20]. In effect, the homogenized system describes the behaviour that is static at the first order of magnitude (Auriault, 1991). Moreover, the centroid can be recognized as the *average time*. This averaged time $\langle t \rangle$ is obtained by taking the expectation with respect to the "probability" density $p(x_3, t)$, i.e.

$$\langle t \rangle \triangleq \frac{\int_0^\infty t' p(x_3;t',\mathrm{p})\mathrm{d}t'}{\int_0^\infty p(x_3;t',\mathrm{p})\mathrm{d}t'}. \tag{E.36}$$

It expresses the average time required by the wave to travel a distance x_3 and equals the traveltime yielded by the reciprocal of the group velocity evaluated at $k \to 0$. Remark that in an similar manner the spatial centroid,

$$\langle x_3 \rangle \triangleq \frac{\int_{-\infty}^{+\infty} x_3' p(x_3';t,\mathrm{p})\mathrm{d}x_3'}{\int_{-\infty}^{+\infty} p(x_3';t,\mathrm{p})\mathrm{d}x_3'}, \tag{E.37}$$

can be defined. This spatial centroid travels with a speed given by the group velocity for $\omega \to 0$.

Unfortunately the static situation is generally not of interest. So is there another physical interpretation possible? Yes there is, and to demonstrate that let me go back to the Gaussian bell-shaped wave packet of characteristic width $\sigma = \langle x_3^2 \rangle$ and location $\langle x_3 \rangle$, see figure E.4. For simplicity let me consider the normal incidence situation, $\mathrm{p} = 0$. Furthermore let me suppose that I effectuated the coarse-graining as depicted in equation (E.25)

[19] That is why I used the equivalent sign $\equiv$.

[20] Of course certain precautions concerning the smoothing kernel have to be taken if necessary, see chapter 2.

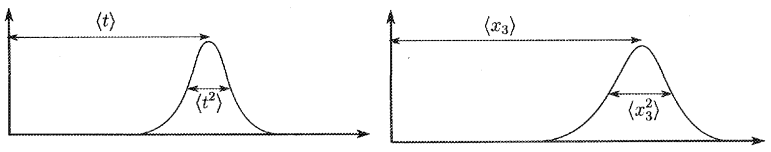

Figure E.4 *Illustration of the envelopes of the temporal and spatial wave packets with their average time and space coordinates given by the centroids. Their scales, i.e. their temporal and spatial extents, are given by their second moments.*

with $L = \sigma$. Then, in fact, what homogenization theory is saying (Asch et al., 1990) is that one is allowed to propagate the wave packet over a distance proportional to σ but ***not*** further! Under that restriction the centroid is still described correctly and this is equivalent to saying that the equivalent medium yields accurate predictions for the average time, $\langle t \rangle$, and space, $\langle x_3 \rangle$, behaviour of the wavefield up to propagation distances being proportional to the scale σ. However, not much can be said on changes in the width $\langle x_3^2 \rangle$ of the wave packet after travelling such a distance. That implies that one can not repeat the above procedure – mapping the initial spatial width of the wave packet onto the medium via the smoothing over σ – because one looses track of this spatial gauge. One does not know whether the spatial gauge changes or not. With other words, one is limited to make one single step of a size proportional to the width σ. Of course one may wonder why not use the wavelength as the gauge? But then I run into trouble because this wavelength itself depends on the equivalent velocity, i.e. the wave packet travels with a group velocity obtained via the homogenization procedure with the initial spatial gauge σ, which on its turn strongly depends on the scale. So what is the velocity after travelling a distance σ? I believe one ends up here in a catch 22[21]. This entrapment points out the notion that one has to know the spatial gauge in order to determine the velocity. But this is a rather unfortunate coincidence since on the one hand the main candidate for the spatial gauge is the wavelength, which depends itself on the velocity evidenced by the wave packet, while on the other hand the current formulation does, in my opinion, not allow for a control over σ. That is to say, there are no scale derivatives that can act as infinitesimal generators invoking spatial dilatations, see chapter 3. At this point one may of course argue that the *local*[22] compressional wavespeed defines the local wavelength $\lambda_3(x_3) = c_p(x_3)\Delta t$. But then the question is, what happens when the compressional wavespeed displays a scaling behaviour? In that case namely, the local compressional wavespeed depends heavily on the scale of observation and hence a similar problem emerges. For me the question boils down to how to define the *velocity* in a medium that scales and this is congruent to asking how to define a *spatial gauge* to mea-

[21] Webster (1988): paradoxical rule found in the novel *Catch-22* (1961) by Joseph Heller.

[22] Local in the sense of being fine-grained not with respect to the spatial gauge of the wave packet itself.

sure the velocity. Finally, notice that I expect, from physical grounds, namely linearity in time, the temporal gauge to be invariant, because it is difficult to imagine that a wave packet living at the infrared end of the temporal frequency spectrum turns into a wave packet living at ultraviolet end of the spectrum!

E.6 Towards an alternative formulation of wave dynamics?

I must say that my ideas are heavily inspired by the lines of thought of le Méhauté (1995) and Nottale (1995). In their work the role of the time and the time derivative are all important, a notion one is well aware of when dealing with ray-asymptotics, $\Delta t \rightarrow 0$, or static long wavelength equivalent medium theory, $\Delta t \rightarrow \infty$.

Therefore I will conclude the *Epilogus* with two sections providing a rough sketch for an alternative representation for wave motion. In this representation I envisage scale derivatives to emerge, that capture the scale dynamics as they did while setting up the scaling medium representation to capture the well-log's complexity. Such an alternative approach will by all means be void of a physical meaning when it does not adhere to some sort of *"correspondence principle"*. This principle guarantees the generalized formulation to match the results obtained by the conventional formulation in cases there is no longer a dominant scale dependence. There remains a possibility that the predictions of the generalized theory correspond to those yielded by the conventional formulation. In that case the alternative formulation has the advantage to be defined in terms of a scaling medium representation. This circumvents the required redundancy in spatial bandwidth. Hence, one is better set to strive for inversion.

E.6.1 Some reflections on the role of scale in physics

One of the primary scientific motivations – the more technical one would have been the integration of well and seismic data – for me to make an effort and try to initiate a discussion on the current representation for acoustic wave motion can be found in the work by Nottale (1995) and le Méhauté (1995). Therefore, I like to share two quotations from the first author which inspired me and hopefully the reader. I think the problem I feel myself confronted with has been paraphrased very nicely. The first quote drawn from Nottale (1995) reads:

> "Since the time of Newton and Leibniz, the founders of integro-differential calculus, one basic hypothesis which is put forward in our description of physical phenomena is that of *differentiability*. The strength of this hypothesis has been to allow physicists to write the equations of physics in terms of differential equations. However, there is no *a priori* principle which imposes the fundamental laws of physics to be *differentiable*.
>
> We shall make the opposite assumption: the elementary laws of physics are actually *nondifferentiable*. Under this conjecture, the successes of present differentiable physics are understood as applying to domains where the approximation of differ-

> entiability (or integrability) was good enough, i.e. at scales such that the effects of the nondifferentiability were smoothed out; but conversely, we expect its methods to fail when confronted by truly *nondifferentiable* or *nonintegrable* phenomena, namely at very small and very large length scales, and, to a smaller extent, for chaotic systems[23].
>
> The new frontier is, in our opinion, to construct a continuous but nondifferentiable physics (We stress the fact that giving up differentiability does not impose giving up continuity). Set in such terms, the project may seem to be extraordinarily difficult. Fortunately, there is a fundamental key which will be of great help in this quest, namely the concept of scale transformations".

The second one I took from a more recent overview paper (Nottale, 1996) where he writes:

> "In standard differentiable physics, it amounts to find differential equations implying the derivatives of f, namely $\partial f/\partial x$, $\partial^2 f/\partial x^2$, that describe the laws of displacement and motion. The integro-differentiable method amounts to performing such a local description, then integrating to get the global properties of the system under consideration. Such a method has often been called "reductionist", and it was indeed adapted to most classical problems where no new information appears at different scales.
>
> But the situation is completely different for systems implying fractals and nondifferentiability: very small and very large scales, but also chaotic and/or turbulent systems in physics, and probably most living systems. In these cases, new, original information exists at different scales, and the project to reduce the behaviour of a system at one scale (in general, the large one) from its description at another scale (in general, the smallest, $\delta x \to 0$) seems to lose its meaning and to be hopeless. Our suggestion consists precisely of giving up such a hope, and of introducing a new frame of thought where all scales co-exist simultaneously as different worlds, but are connected together via scale-differential equations.
>
> Indeed, in non-differentiable physics, $\partial f(x)/\partial x = \partial f(x,0)/\partial x$ does not exist any longer[24]. But the physics of the given process will be completely described if we succeed in knowing $f(x,\varepsilon)$ for all values of ε, which is differentiable when $\varepsilon \neq 0$, and can be the solution of differential equations involving $\partial f(x,\varepsilon)/\partial x$ but also $\partial f(x,\varepsilon)/\partial \ln \varepsilon$. More generally, if one seeks nonlinear laws, we expect the equations of physics to take the form of second order differential equations, which will then contain, in addition to the previous first derivatives, operators like $\partial^2/\partial x^2$ (laws of motion), $\partial^2/\partial(\ln \varepsilon)^2$ (laws of scale), but also $\partial^2 f/\partial x \partial(\ln \varepsilon)$, which corresponds to a coupling between motion laws and scale laws.

[23] Note from the author: This latter category is of relevance for my problem, since the medium heterogeneity can be seen as "frozen" turbulence. That is to say, not temporally changing during the transient wave phenomenon.

[24] Note from the author: Here $f(x,\varepsilon)$ denotes the regularized/smoothed version of a singular, showing a scale divergence, functional f, see chapter 2 and chapter 5.

> What is the meaning of the new differential $\partial f(x,\varepsilon)/\partial \ln \varepsilon$? This is nothing but the variation of the quantity f under an infinitesimal *scale transformation*, i.e. a dilatation of resolution."

Needless to say that I took this work as a source of inspiration. Let me now categorize how some of the aspects, coming forward in these quotes, found their way into the problem of waves in scaling media. In the first place Nottale's line of thinking sparked the impression at my side that the linear partial differential equations, denoting the wave equation, play a double role:

- Firstly, they constitute, in a mathematical sense, a rule capturing the space-time behaviour on an infinitesimally small scale range.
- Secondly, they prescribe the way in which one has to average in order to provide a phenomenological description on a coarser scale, the scale at which the interactions one is interested in are believed to take place.

From these observations it is quite clear that differential calculus derives its power at the infinitesimal scale range where functions are rules that assign numbers to points. In that role the pde's capture the relevant physical interactions under certain restraining assumptions. In the wave problem this assumption points to a *separation of scales*, enabling the definition of macroscopic quantities for which subsequently constitutive relations can be defined independent of the size of the elementary volume. This independence allowed for the definition of *local* forms for the conservation laws and constituted the paramount prerequisite for the derivation of the coupled system of pde's denoting the wave equation. So, in effect, the constitutive relations and the wave equation itself derive their *universality* from the basic assumptions that led to their derivation. That is why, in my opinion, homogenization theory – the formalism prescribing the way in which a pde has to be averaged – is for me so natural when one is interested in describing the wave motion at the scaled up coarser scale. The reason for this is that the prerequisite assumption for a separation of scales is shared within the framework that allowed for the derivation of the wave equation as well as for setting up homogenization theory. Hence, homogenization theory derives its applicability from the same *first principles* as the derivation of the wave equation in itself does. That means that one is free to upscale – when the separation of scales is guaranteed – the wave equation to a coarser scale level. In that process it is even possible to allow for a singular behaviour of the constitutive parameters[25] at the small scales since they will be smoothed out!

What if the homogenized system does not capture the physical wave phenomena being observed? For instance one finds that the waves disperse and that specular reflections

[25]There is a potential flaw in applying Gauss' divergence theorem, but that can be resolved by giving the integral a distributional interpretation, supplementing the proper conditions on the smoothing kernel.

occur. Clearly one has then to do with a situation where the homogenization approach fails. That raises the question whether the choice of the ruling system of pde's has been a proper one in the sense that, say, relaxation may erroneously have been ignored. On the other hand, this observation may also point in the direction that the founding principles and assumptions under which the wave equation has been derived can not withstand the touchstone of criticism within their realm of application. With this I mean the potential danger that is hidden in the upscaling operation from the infinitesimal domain to the scale range where the actual dynamics are believed to conduct their actions. This danger lies in the risk one runs in jeopardizing the founding principles under which the wave equation has been derived and this brings me to the second observation I borrowed from Nottale. That is the one which makes reference to the notion that it may be meaningless to apply the "reductionist" approach when dealing with a problem of scales – e.g. the integration of fine-grained well-log and coarse-grained seismic data – with a theory that does not make explicit reference to scale and scale dynamics.

The final theme, besides the role of the time derivative, see below, I took from the work of le Méhauté (1995) and Nottale (1995), are the important notions of *scale divergence* and its appropriate counter measure *regularization.* These are the two congruent concepts that form major ingredients motivating me to introduce a scaling medium representation in the continuous wavelet domain. This scaling medium representation is accompanied by a multiscale analysis and characterization for the medium's heterogeneities as, for instance, is being displayed by well-log measurements of the earth's sedimentary deposits. In the characterization the fractal concept emerged and it appears that the fractal structures can be seen as solutions of linearized renormalization group equations, equations that contain scale derivatives. In chapter 3, I made an effort to substantiate the main ingredient of the scale representation namely the scale derivative operator.

Indeed, the empirical findings of the multiscale analysis conducted upon real well-data demonstrate that the medium fluctuations contain structures of "all" sizes, *certainly within the scale range seismic waves are believed to interact.* That is why the scaling medium representation and its characterization, by means of scaling exponents, are heavily based on the scale derivative operator in the sense that these exponents represent eigenvalues of the scale operator. It is this observation that brings me to the question *whether it is possible to come up with a wave theory that contains a scale derivative that carries the scale information, e.g. the singularities, from the medium to the probing wavefield?*

E.6.2 Some remarks on dynamics and scaling

At this point I arrived at the crucial question whether scale derivatives enter into the formulation of the wave dynamics or not? To answer this question let me first briefly review what I believe is known within the conventional context and what the conditions

are to be imposed on a potential alternative formulation. First of all I impose that this generalized formulation has to comply with a certain "correspondence principle" guaranteeing a transition back to the "conventional" formulation in those cases where the scale derivative does not detect a change, i.e. when the regularization is insensitive to infinitesimal dilatations in the smoothing kernel. What are those situations? Certainly the static equivalent medium is a candidate, because in that case the spatial gauge tends to infinity and infinitesimal changes at infinity do not make sense. On the other hand one can think of a situation where the medium is assumed to become smooth at the small scale range. In that case the geometrical optical regime is another candidate.

What do these two extreme cases have in common? They refer to asymptotic values for the temporal gauge, Δt, namely $\Delta t \to 0$ for the geometrical optics and $\Delta t \to \infty$ for the statics. Both cases are well understood, but, unfortunately, their applicability is limited in cases where the dynamics are of interest, i.e. in the intermediate regime. Notice, however, that in the geometrical optical approach one can induce specular reflections by allowing jump discontinuities to occur at which one can solve the local scattering by supplementing appropriate boundary conditions on the wave field.

To summarize, I like to argue that in the two afore mentioned limiting cases, I expect to obtain results that become asymptotically equal for both the conventional and the extended formulation. For other situations I do not know. I do know, however, that the possible scale contribution vanishes in the two above situations. For the geometrical optical situation that is the case, because I assumed the medium to be void of singularities. When these singularities do occur, then they will be isolated because the spatial gauge goes to zero. In that case I demand the solution of the alternative formulation to converge, preferably pointwise, to the solution obtained by supplementing boundary conditions within the conventional formulation. Notice, however, that the solution of the alternative formulation will only approximate the latter solution because the spatial gauge is not allowed to go to zero! In other words, the solution obtained by using boundary conditions represents the closure for the alternative solutions as $\Delta t \to 0$. When one is not in the geometrical optical regime, then I believe that care must be taken concerning the mutual interference of the singularities that can not longer be considered as being isolated in all cases! Finally, I like to remark that the same line of reasoning probably holds for the solution of isolated singularities of the homogeneous type (Wapenaar, 1996b).

Now what else do these two situations, $\Delta t \to 0$ and $\Delta t \to \infty$, tell the reader? They both make reference on how time acts as a kind of spatial gauge. This brings me to the heart of the matter in the sense that both Nottale (1995) and le Méhauté (1995) intervene in the role played by the time when the physical dynamics at constructs displaying a scale divergence, e.g. fractals, are of concern. Their claim is that this type of singular behaviour induces *irreversibility* of the time, implying a break in the space-time symmetry. It introduces dispersion. This may link back to the notion that the solutions of the wave

equation in the absence of the pure point and possible singular continuous part of the spectrum constitute a *group* whereas in the case of irreversibility the solutions belong to a *semi-group*, see for instance Dautray and Lions (1992) or Ziedler (1991).

It is le Méhauté (1995) who argues that the *time* must be used to regularize the spatial scale divergence and hence establishing an inherent link between space and time that is not longer independent of the spatial gauge as it is in a regular environment where only one velocity connects space and time. This let him to coin a replacement of the time derivative to a fractional time derivative, the degree of which is directly related to the Hausdorff dimension of the fractal at which electromagnetic wave interactions take place (le Méhauté, 1995). Another approach is followed by Nottale (1996, 1995, 1992) who argues that space-time is fractal. As a consequence he replaces the co-moving time derivative by a generalized version containing a complex velocity and a reference to the time gauge Δt. As will become clear later, the temporal gauge occurs in a generalized scale dependent "diffusion" coefficient that is found in the "fractal source term" in the conservation laws. Let me now quickly illustrate the concept of time in relation to fractal scaling before going back to the constitutive relations.

E.6.3 The role of time in relation to fractals: le Méhauté's approach

In Euclidean/Minkowskian space-time (Nottale, 1995) there exists a trivial scale invariant relationship between space and time. This relationship is expressed by the notion of velocity. In this section it is argued that in fractal space-time such a relationship does not have a meaning, i.e. a scale invariant velocity does not exist. In order to illustrate this fundamental observation I will give a simple example on a race between a mouse and an elephant. First the mouse and the elephant will run along a line, i.e. a smooth Euclidean trajectory in space-time. After that they will run along a non-Euclidean monofractal curve.

But before regarding their race, a recipe will be given on how to construct a simple fractal curve, followed by how to associate the notion of a generalized fractal dimension to it.

Fractal dimension

The monofractal curve along which the mouse and the elephant are going to race is the Von Koch curve. It is constructed in the following way. Start with a straight line of unit length, see figure E.5 (a), and proceed by dividing it into three equal parts. Replace the middle part with an equilateral triangle and throw its base away, figure E.5 (b). Repeating this procedure on every line segment of the new curve, the curve in figure E.5 (c) is obtained and iterating once more yields figure E.5 (d). Figure E.5 (e) is obtained after a few more iterations. Now by taking the number of iterations to infinity one defines the Von Koch curve. It is clear that the (Euclidean) length of the curve after n iterations

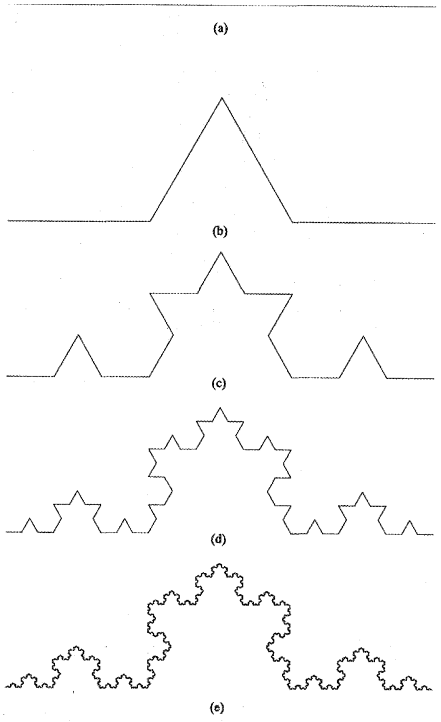

Figure E.5 *The construction of the Von Koch curve.*

becomes

$$L_n = \left(\frac{4}{3}\right)^n, \tag{E.38}$$

yielding a *divergence* as the limit $n \to \infty$ is taken. Is a diverging length realistic? Yes, it is indeed. The problem of diverging lengths was coined by Mandelbrot (1967), who wondered how long the coast-line of Great Britain is. He showed that when the coast-line of Great Britain is measured with gauges of length λ, the number of gauges N needed follows a power law

$$N(\lambda) \sim \lambda^{-\mathrm{D}}, \quad \lambda \downarrow 0, \tag{E.39}$$

where the exponent D is called the fractal dimension of the curve. Here, as le Méhauté (1991, 1995) points out, the limit of λ towards zero must be taken in a relative sense. It is the way in which the number of gauges increases or decreases that determines the rate of fractality in a scale range. Equation (E.39) gives the rate of change of the number of gauges with respect to the resolution. On the other hand, in case the curve is a straight

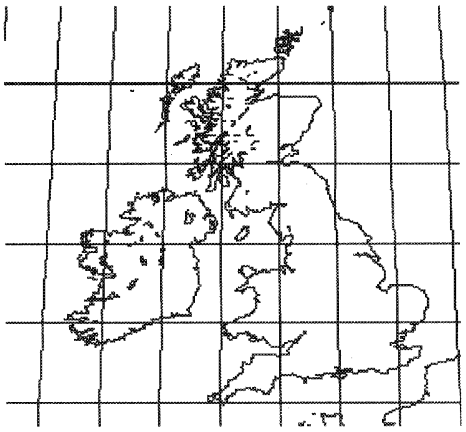

Figure E.6 *The coast-line of Great Britain.*

line or a smooth curve, then the number of gauges needed is only doubled when the length of the gauges is halved. This means that the dimension of a line $D = 1$. However, when this procedure is repeated for the highly irregular coast-line of Great Britain the exponent D is found to be greater than one, indicating that the number of gauges is more than doubled when the length of the sticks is halved, see also figure E.6.

Since the total length of the curve is given by the number of sticks multiplied with the length of the stick, the total length is

$$L(\lambda) = \lambda N(\lambda) \sim \lambda^{1-D}. \tag{E.40}$$

It is clear that when $D > 1$, the length of the curve diverges if the gauges are taken smaller and smaller. In other words, if the resolution of the measurement is taken finer, the outcome of the measurement of the length increases. This shows that the outcome of the measurement depends on the scale at which the measurement is made. If the dimension $D = 1$, i.e. the curve behaves as a line, then equation (E.40) shows that the outcome of the measurement is independent of the scale. Now, return to the Von Koch curve. It is not hard to show that the fractal dimension of this curve is given by

$$D = -\frac{\log N(\lambda)}{\log \lambda} = \frac{\log 4}{\log 3} = 1.2618\ldots, \tag{E.41}$$

which is indeed larger than one. This confirms the observation that the length of this curve diverges when it is viewed on increasingly finer scales.

The race of the mouse and the elephant

As I already briefly mentioned in the beginning of section E.6.3, I will try to convey the implications of fractal space-time by considering a race between a mouse and an elephant.

This race consists of traversing a trajectory between two points in Euclidean space-time, the start and finish. Now suppose that it took them both exactly the same time T to cover the distance. Then, from the Euclidean standpoint, everyone would agree that both participants must have travelled with the same average speed, despite the fact that the mouse and the elephant live on a different scale. So clearly they should both earn the gold medal, because it took them exactly the same time T to cover the track.

As already mentioned, the mouse and the elephant live on a different resolution, i.e. their step size is different, or, with other words, their gauges are of different size. Suppose that the size of the step Δx_m of the mouse is a times smaller than the size of the step Δx_e of the elephant, $a > 1$, i.e.

$$\frac{\Delta x_m}{\Delta x_e} = a^{-1}. \tag{E.42}$$

As a consequence of the smoothness of the track, think, e.g., of a cinder track, the number of steps they have to make is inversely proportional to the size of their steps,

$$N_e(\Delta x_e) \sim {\Delta x_e}^{-1} \quad \text{and} \quad N_m(\Delta x_m) \sim {\Delta x_m}^{-1}. \tag{E.43}$$

The total time T they have walked is now of course the number of steps multiplied by the average time to take a step

$$T = N_e(\Delta x_e)\Delta t_e = N_m(\Delta x_m)\Delta t_m, \tag{E.44}$$

or

$$\frac{\Delta t_m}{\Delta t_e} = \frac{N_e(\Delta x_e)}{N_m(\Delta x_e)} \sim \left(\frac{\Delta x_e}{\Delta x_m}\right)^{-1} = a^{-1}. \tag{E.45}$$

Equation (E.45) shows that when the gauge of space is made a times smaller, then the gauge of time has to be made a times smaller. Now clearly both the mouse and the elephant have walked the same distance. This distance is independent of both space and time and this is a mere reflection of the fact that the cinder track lives in Euclidean space.

Now let me consider a second battle. Again the mouse and the elephant are the two sole competitors. But now they will have to cover a fractal trajectory, such as the Von Koch curve, from start to finish. As a shear coincidence they finish again at exactly the same time and one might wonder who has won? The elephant will say that both have won, but the mouse will argue that he covered a larger distance than the elephant in the same amount of time, so that he has beaten the elephant. This shows that in the fractal case it will require some careful deliberations by the jury to decide who has won the race.

Let me state some arguments the jury must take into account. Running along a monofractal trajectory with fractal dimension D, the number of steps the mouse and the elephant had to make is proportional with, cf. equation (E.39),

$$N_e(\Delta x_e) \sim {\Delta x_e}^{-\mathrm{D}} \quad \text{and} \quad N_m(\Delta x_m) \sim {\Delta x_m}^{-\mathrm{D}}. \tag{E.46}$$

Hence, the ratio of the times of their steps becomes

$$\frac{\Delta t_m}{\Delta t_e} = \frac{N_e(\Delta x_e)}{N_m(\Delta x_e)} \sim \left(\frac{\Delta x_e}{\Delta x_m}\right)^{-\mathrm{D}} = a^{-\mathrm{D}}. \tag{E.47}$$

So, making the gauge of the space a times smaller induces that the gauge of the time has to be made a^{D} times smaller. This means that the relation between space and time is not longer a linear one as it was in case of the Euclidean trajectory. Hence, there does not longer exist a unique scale invariant velocity. In fact the relation between time and space is found to be

$$\Delta t = \frac{T}{N(\Delta x)} \sim \Delta x^{\mathrm{D}}, \tag{E.48}$$

or

$$\Delta x \sim \Delta t^{\frac{1}{\mathrm{D}}}. \tag{E.49}$$

The scaling of the time is caused by the broken dimension of the space. Equation (E.49) implies immediately that it is impossible to define a scale invariant speed as soon as $\mathrm{D} > 1$, since this speed would be given by

$$\lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \sim \lim_{\Delta t \to 0} \Delta t^{1-\frac{1}{\mathrm{D}}}, \tag{E.50}$$

which diverges as soon as $\mathrm{D} > 1$. Notice that Brownian motion has a fractal dimension $\mathrm{D} = 2$ (le Méhauté, 1991). With equation (E.49) this shows indeed that $x(t)$ has singularities of strength $\alpha = \frac{1}{2}$ everywhere. So the conclusion is that *a broken dimension in space yields singularities in time.* Notice that it is possible to define a generalized velocity

$$v(t) = \frac{\partial^{\frac{1}{\mathrm{D}}}}{\partial t^{\frac{1}{\mathrm{D}}}} x(t), \tag{E.51}$$

which reduces to the usual notion of velocity as $\mathrm{D} = 1$.

As a final remark I would like to state, following Mandelbrot (1982) and le Méhauté (1991, 1995), that the afore mentioned fractal space-time behaviour can be associated with a generalized diffusion of the form

$$p(x,t) \sim \mathcal{F}^{-1}\{e^{(-j\omega\tau_0)^{\frac{1}{\mathrm{D}}} x}\}, \tag{E.52}$$

where τ_0 is a constant depending on a reference scale. It is clear that in the Euclidean case, $\mathrm{D} = 1$, this equation obtains the form of a wave, whereas in case $\mathrm{D} = 2$ it obtains the form of a diffusion. For other values of D the equation becomes the solution of a generalized diffusion (le Méhauté, 1991). Notice the constant Q behaviour of equation (E.52) (Bickel, 1993; Hargreaves, 1992; Kjartansson, 1981).

Finally notice that in this example the self-similar situation has been considered solely. For the self-affine situation, where the track is a fractal "function" $f(x)$ of the coordinate x, one has to do with a spatial *gauge*

$$\Delta x \sim \Delta t \tag{E.53}$$

rather than

$$\Delta x \sim \Delta t^{\frac{1}{D}}, \tag{E.54}$$

see le Méhauté (1991).

E.7 A possible conjecture

E.7.1 The story of the mouse and the elephant continued

In the previous story on the race between mouse and the elephant one had to do with a coast-line that was homogeneous in its scaling behaviour. This assumption allowed for an establishment of a non-trivial relationship between space and time governed by one single fractal dimension. As a consequence of the non-triviality it was observed that a dispersion entered into the match, which made it harder for the mouse to win. Clearly, high and low frequencies correspond to "mouseness" or "elephantness" and this led Mandelbrot (1982), le Méhauté (1991) and Hargreaves (1992) to coin a generalized diffusion process to be associated with wave motion in monofractal media.

Of course, the underlying assumption of monofractality withstands a full comprehension of the dynamics of waves in media that display a heterogeneous type of scaling behaviour. Furthermore, I find it difficult to still assimilate translation as a means of transportation, because dilatation has taken over, and to comprehend specular "reflections". I attribute these reflections to the existence of dominating singularities. Therefore, I would like to postulate a generalized version for the tale of the elephant and the mouse. As a basic assumption I take the number of steps per time interval to be constant for all "creatures" participating in the race. I used the quotes around creatures to indicate that, as a consequence of spatial dilatations, the participants may transform from one species to the other. That is to say that the spatial gauges, the spatial step size being linked to the characteristic size of the creatures, are allowed to change during the race, so a mouse may turn into an elephant and vice versa. These changes in the spatial gauge are induced when crossing singularities appearing in the coast-line.

The physical interpretation to be given to the above metaphor is that I perceive wave packets to dilate when encountering singularities in the medium properties. On the other hand, the temporal gauge is taken to be constant. It is interesting to note that in this approach I do not longer think of the wave problem in terms of individual geodesics but rather in terms of the congruent behaviour of possibly infinitely many geodesics,

which can not be determined individually. To continue the above metaphor, one may think of a geodesic as being the solution to a "classical" Lagrangian, i.e. the eikonal equation, defined in terms of the conventional wave equation. This solution is obtained by attributing the elementary volume to, say, the characteristic size of a single cell amoeba. And if on this scale the environment is assumed to be non-scaling, then the amoeba and its "dilated" counterpart will not experience the consequences of scaling, because its environment is ruled by *functions*. However, an ant, a mouse or an elephant, which are conglomerates of cells, certainly live through the effects induced by the scaling. Their environments are formed by scale indexed *functionals*, i.e. the scaling medium representation.

Now the question is, *how must waves physically be embodied?* Are they amoebae, or are they multicellular?

E.7.2 A starting point?

As a consequence of the above line of reasoning the time has been set as a fixed gauge with which space is going to be measured. Since I am dealing with a self-affine situation, the medium profiles depend on the vertical coordinate, the spatial gauge is simply, following le Méhauté (1991), proportional to the time, $\sigma(t) \sim t$. It is this choice which potentially solves the problem of how to define the spatial gauge which has to be related to the time in one way or another. Let me now demonstrate what the implications are of this choice on the role of the time and the time derivative in conjunction with the constitutive relations.

In the current formulation one was able to replace, by virtue of Reynolds' transport theorem, the time derivative acting on, say, the functional expressing the observed mass flow density $\Phi_x = \rho v_x$, i.e. $\frac{\mathrm{d}}{\mathrm{d}t}\langle\Phi_x, \phi_{\sigma(t),x}\rangle$, by a partial[26] time derivative sandwiched between the mass density, ρ, and the field quantity, v_x, $\frac{\mathrm{d}}{\mathrm{d}t}\langle\Phi_x, \phi_{\sigma(t),x}\rangle = \langle\rho\partial_t v_x, \phi_{\sigma(t),x}\rangle$. This identity holds since there exists a relation expressing the conservation of mass for the elementary volume. This conservation law states that the time derivative of the functional $\langle\rho, \phi_{\sigma(t),x}\rangle$ equals zero, i.e. $\frac{\mathrm{d}}{\mathrm{d}t}\langle\rho, \phi_{\sigma(t),x}\rangle = 0$, in the absence of annihilation and creation of mass (de Hoop, 1995). To summarize, the derivation of the constitutive relations manifests the material properties, $\{\rho, \kappa\}$, with respect to the observed field quantities $\{v_x, p\}$. In order to establish a physical meaning to these constitutive relations I demand passivity, following de Hoop (1995), implying that $\frac{\mathrm{d}}{\mathrm{d}t}\langle\{\Phi_x, \Theta\}, \phi_{\sigma(t),x}\rangle \to 0$ as $\frac{\mathrm{d}}{\mathrm{d}t}\langle\{v_x, p\}, \phi_{\sigma(t),x}\rangle \to 0$.

In equation (E.17) I postulated that, in case of a scaling number density, one may have to alter the form of the pertaining conservation law. For the conservation of mass this

[26] I linearized the co-moving time derivative, i.e. $D_t \mapsto \partial t$.

alteration implies a recast of the conservation law into

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \rho, \phi_{\sigma(t),x}\rangle = \langle \partial_t \rho, \psi_{\sigma(t),x}\rangle \qquad \Delta t \to 0 \tag{E.55}$$

rather than

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \rho, \phi_{\sigma(t),x}\rangle = 0 \qquad \Delta t \to 0. \tag{E.56}$$

The implication of this postulate is that the source free *equation of motion*,

$$\langle \partial_x p, \phi_{\sigma(t),x}\rangle + \frac{\mathrm{d}}{\mathrm{d}t}\langle \Phi_x, \phi_{\sigma(t),x}\rangle = 0 \tag{E.57}$$

– being, as a consequence of equation (E.56) and Reynolds' transport theorem, equivalent to

$$\langle \partial_x p, \phi_{\sigma(t),x}\rangle + \langle \rho D_t v_x, \phi_{\sigma(t),x}\rangle = 0, \tag{E.58}$$

– becomes a system of equations reading

$$\langle \partial_x p, \phi_{\sigma(t),x}\rangle + \frac{\mathrm{d}}{\mathrm{d}t}\langle \Phi_x, \phi_{\sigma(t),x}\rangle = 0 \tag{E.59}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \rho, \phi_{\sigma(t),x}\rangle = \lim_{\Delta t \to 0} \frac{\Delta \sigma}{\sigma}\langle \partial_t \rho, \psi_{\sigma(t),x}\rangle \tag{E.60}$$

In the modified situation the functionals can not, as in the current formulation, be dropped in order to obtain a *local* form for the equation of motion

$$\partial_x p + \rho \partial_t v_x = 0, \tag{E.61}$$

where I linearized the co-moving time derivative towards an ordinary partial time derivative. Instead I ended up with a functional relationship in which explicit reference is made to the spatial scale and indirectly to the temporal gauge, Δt. Moreover, an additional functional emerged expressing a contribution due to an infinitesimal dilatation induced during the time step, Δt.

What is to be learned from the additional equation (E.60)? First of all it refers to a change in the time rate of change in the density of mass, effectuated by an change in the resolution of the elementary volume during the time step Δt. This change is measured in terms of the wavelet coefficient $\langle \partial_t \rho, \psi_{\sigma(t),x}\rangle$ and, as a consequence, it becomes interesting to study the behaviour as a function of the temporal gauge, Δt. Let me first note that while taking $\Delta t \to \infty$ one obtains the static behaviour in which the density must be conserved a notion becoming manifest in the sense that the scale contribution drops out. Moreover, this scaling contribution becomes zero in case the density becomes smooth at the small scale range or anywhere else. Of course these are the two situations I am not

strictly interested in and let me therefore review what may occur during the time step Δt in case the density is taken as

$$\rho(x,t) \sim x^{\alpha}. \tag{E.62}$$

Then it may be expected that equation (E.55) would yield a behaviour like

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\langle \rho, \phi_{\sigma(t),x}\rangle &= \langle \partial_t \rho, \psi_{\sigma(t),x}\rangle \\ &\sim \partial_t t^{\alpha} \end{aligned} \tag{E.63}$$

This, on its turn, may imply a dissipation since

$$\mathcal{F}\{t^{\alpha-1}\} \sim (j\omega)^{-\alpha}, \tag{E.64}$$

yielding an *irreversibility* of the time. In this expression I made use of the fundamental property of the Fourier transform that maps powerlaws onto powerlaws and that leaves scale operator of the type $x\partial_x$ untouched, i.e. $x\partial_x + 1 \mapsto k\partial_k$ (le Méhauté, 1991; Holschneider, 1995). The irreversibility implies that the time has only one direction, a notion well reconcilable with the fact that the real and imaginary parts of $(j\omega)^{\alpha}$[27] form a Hilbert transform pair under certain conditions for the α and that implies a *causality*. There is even a more important observation to make, that concerns the conjectured ability of this formulation to transport spatial singularities from the medium to temporal singularities of the wavefield. At this point I can not prove this, but I know that something similar occured for the situation described in the previous section where the scale divergence of the "coast-line" mapped to singularities in the time. Another point favouring this idea is the fact that the algebraic singularities constitute eigenfunctions of the scale operator, i.e.

$$x\partial_x x^{\alpha} = \alpha x^{\alpha}. \tag{E.65}$$

In this way the scaling remains intact. This notion is in my opinion not shared by, for instance, operators of the form

$$\frac{1}{\rho}\partial_x \rho \sim \frac{1}{x^{\alpha}}\partial_x x^{\alpha} = \alpha\frac{1}{x} \tag{E.66}$$

where the logarithmic derivative, appearing in the reflectivity[28], seems to absorb the scaling! For me this observation is difficult to reconcile with the empirical findings concerning the measured seismic data.

In section E.4 I mentioned that the scale contribution would only matter in those cases where it exceeds the $o(1)$ term. If the scaling becomes important then the system depicted

[27] This can be associated with constant Q behaviour.

[28] I am aware that care must be taken while differentiating x^{α}.

in equations (E.59) and (E.60) comes into action. At this stage I do not oversee what the exact consequences would be. However, given the global/local Hölder exponent estimates it is relative straightforward to determine this transition, marking the point at which the scaling becomes important. As far I can judge, I expect the transition to occur at $\alpha = 1$. So for $\alpha < 1$ the scale contribution comes in. For $\alpha = 0$ one obtains a frequency independence, a property which may well be consistent with the fact that the Hölder exponent of a jump discontinuity equals zero while the pertaining reflection coefficient is frequency independent. Finally, if the scaling of the density is broken at the large or small scales[29], i.e. there is a transition from scale divergent behaviour to regular behaviour, see chapter 2, then the additional scaling contribution will drop out! Hopefully the future expressions for the reflectivity induced by isolated algebraic singularities will yield the same behaviour as found via the method of imposing boundary conditions.

Indeed the scaling medium representation I introduced in chapter 2 fits very well in the above picture. But again the presented conjecture is still in a very premature state of development. On the other hand, I would be surprised when the scaling medium representation I proposed will not find its way into the current formulation of the wave theory. The reason for this is that the scale derivative's role proved to be essential in unraveling the complexity displayed by the medium and that sparks the question why it is lacking in the current wave theory? This wave theory is namely used to describe how this complexity is carried to the space-time behaviour of the wavefield!

At this point I ask to myself whether the basic assumptions within the idealization of the continuum mechanics can be matched with the multiscale findings in this thesis. In this context I like to quote Sedov (1971)

> " ... in particular, such an idealization is necessary, because one wants to employ the apparatus of continuous *functions*, differential and integral calculus in the study of the motion of deformable bodies."

Let me stress that I refrain from rejecting differential and integral calculus, because I believe that these concepts are still deployable when the appropriate measures are taken to deal with the complexity. For me, these measures boil down to replace *functions* by *functionals* and to consider scale derivatives. Of course, the notions of displacement and dilatation are crucial concepts in describing physical phenomena, only in my opinion they should not be applied without a reference to scale, i.e. they should refer to functionals instead of functions. For me the prerequisite of applying functions solely hinders a true comprehension of complexity. The scaling medium representation may be a last remedy to unravel and deal with the complexity, given the above integro-differential theoretical framework.

Falling back on the metaphor, what is happening with the "embodied" wave packet as it hits a singularity? As I mentioned, spatial dilatations were allowed to take place. These

[29] May even be at the intermediate scales.

dilatations are in my opinion necessary to transgress a singularity. In fact what happens, is that the singularity is regularized by a stretching of the wave packet, counterbalancing a densification in the medium. Because of the dissymmetry in space-time that goes with this action I expect dissipation. Breaking the complexity by the regularization may cause a loss of energy[30]. Instead of dissipating energy into heat[31] one may envisage the energy to go into an increase of complexity in the time behaviour of the wavefield. The wave started of as a single event in time and as time passes the complexity is gradually built in to the signal due to the singular scattering[32]. Scattering is not well understood and maybe the above deliberations will shed some new light on this difficult mechanism. Furthermore allow me to remark that I find it difficult to understand how it is possible to pick up information on the medium's heterogeneities without some form of dissipation, i.e. I do not understand how a wave packet may interact with a singularity without a loss of energy. I can not prove that, but again it is a consequence of the dissymmetry, and of the general rule of thumb that almost nothing goes for nothing.

Let me now consider the case where a scale derivative enters into the formulation. I expect it to detect the singularities in the constitutive parameters. This detection is then followed by a congruent action of this scale derivative invoking not only a spatial dilatation, on the spatial gauge, but also a transfer of the singularities from space to time. In this, the scale operator acts as the infinitesimal generator for the dilatations. Please let me stress again, that I am fully aware of the speculative nature of this discussion. Still, I am convinced that as soon as a scale derivative enters into the current formulation for the space-time structure of the wave equation, one experiences an intricate scale/space/time coupling. I only do not have the answers yet what this structure exactly looks like. But what I do know, is that this succession of events strives well with the wave interactions at a discontinuity[33], which give rise to a congruent spatial dilatation of the wave packet and the emergence of an additional event in time, all caused by the spatial singularity.

It is interesting to note that I believe it is possible to associate the additional "fractal source" term in equation (E.60) to the "generalized scale dependent diffusion" coefficient appearing in the generalized co-moving/convective time derivative as coined by Nottale (1992, 1995, 1996)

$$\frac{\mathrm{d}}{\mathrm{d}t} \mapsto \frac{d}{\mathrm{d}t} = \frac{\partial}{\partial t} + \mathcal{V} \cdot \nabla - j\mathcal{D}(x,t,\Delta t)\Delta. \tag{E.67}$$

In this operator one finds a generalized complex velocity $\mathcal{V}$, to be attributed to the fractal

[30]This observation does not imply that the conservation law of energy no longer holds. It merely states that the wave's energy transforms into complexity rather than into heat.

[31]I do not expect heat dissipation at the seismic temporal frequency range. This does not apply to ultrasonic wave interaction.

[32]That is to say, specular scattering rather than the scattering referred to in scattering theory, where the spectral measure is taken to be absolutely continuous. This continuity precludes a dispersion.

[33]This is done via supplementing boundary conditions, which in my opinion work well in case the singularities can be regarded as being isolated.

nature of space-time, and a additional complex term corresponding to a diffusion type of behaviour. The coefficient $\mathcal{D}$[34] represents the generalized scale dependent diffusion coefficient.

In the proposed approach presented in this epilogue I only paid attention to the fractal source term. Whether the velocity, defining the co-moving derivative, requires a generalization, has not been investigated. Because the time behaviour is supposed to be linear, I think this approach is justifiable. Concerning the diffusion term, with its Laplacian, I like to mention that it can be associated with a scale derivative. This is because the Gaussian and its derivatives constitute solutions to the diffusion equation, i.e.

$$\frac{\partial}{\partial\sigma}G = \frac{\partial^2}{\partial x^2}G, \quad \text{with } G = \frac{1}{\sigma}e^{\frac{x^2}{\sigma^2}}.$$

That may explain the possible link between the scale derivative in equation (E.60) and the term containing the Laplacian in equation (E.67). To be – speculatively – more specific, I associate the diffusion coefficient $\mathcal{D}(x,t,\Delta t)$ with my spatial gauge $\sigma(x,t,\Delta t)$. This may imply that the recipe that replaces the time derivative by a generalized one as a consequence of the scaling, becomes

$$\frac{\mathrm{d}}{\mathrm{d}t} \mapsto \frac{d}{\mathrm{d}t} = \frac{\partial}{\partial t} - \sigma\frac{\partial}{\partial\sigma} \quad \text{evaluated at } \sigma = \sigma(x,t,\Delta t), \tag{E.68}$$

where I tacitly invoked a "linearization" by neglecting the velocity term. Now of course the crucial question is, what is the behaviour as a function of the temporal gauge? Clearly for an isolated discontinuity one may expect no temporal frequency dependence, because the Hölder exponent equals zero. In "all" other cases, where there is a scale dependence exceeding the $o(1)$, I expect a dissipation type of behaviour.

Perhaps superfluously let me stress again that within the above deliberations no physical assumptions have been made whatsoever concerning scale ranges that lie outside the scale range of observation. Only during the conversion to the "language" of mathematics, one may get the impression that such assumptions have been made. But notice that exactly a similar type of assumption is made by representing the data in terms of – continuous – functions. Given the empirical findings it appeared that tempered distributions are more appropriate to represent the observed data. The implication of this latter choice is that a reference to scale is made and that may yield different predictions for the wave motion in cases where the medium heterogeneities behave as tempered distributions. Of course the formulation has to comply with the "correspondence" principle I mentioned earlier. To summarize, I think it has been useful to come up with the preceding discussion because it coins an initial discussion on a formulation that

- replaces functions by functionals, thereby explicitly referring to the notion of scale.
- does something with the time in relation to space. In that sense the discussion goes along similar lines as the work presented by Nottale (1995) and le Méhauté

[34] Remark that this $\mathcal{D}$ does not refer to the dilatation operator.

(1995), who argue that the time becomes irreversible as a consequence of dissipation induced by the scale divergence and its counter action regularization. In their work this yields a replacement of the ordinary time derivative by a generalized one, breaking the space-time symmetry. Hence, it is argued that the time plays a key role in measuring the spatial complexity, yielding, as a consequence, a velocity that is not longer unique.

- allows for a hopefully more suitable statement of the inverse problem. It is my conjecture that it is perhaps better to aim, during the inversion, for an achievement of information on the nature of the singularities occuring in the time. That is to say, that I envisage an inversion scheme yielding estimates for the local scaling exponents, that characterize the main singularities occurring in the medium properties. Since these singularities express the complexity, it is to be expected that the characterization of the singularities serves the purpose of obtaining a litho-stratigraphical indicator.

At this point I may have over fed the reader and left him or her puzzled with the question why does seismic exploration work? Indeed, from the kinematical point of view it is quite understandable why it works. The reason is that during the forward modelling, it only concerns itself with a mapping of the location pertaining to a singularity in space to the location of the singularity in time, while the reverse process is strived for during migration. This means that for cases where one has a general idea on the average space-time behaviour one is all set, a notion striding well with the inherent redundancy of seismic measurements. For the second dynamic part perceived by me to be responsible for the wave interactions I experience much more difficulties in coming to terms with the intricate mechanism being responsible for the apparent dispersion and specular reflections.

I like to conclude this thesis by quoting Pierce (1981), who makes the following remark concerning the use of the acoustic wave equation as a model

> "Although this model is approximate and gives no account of sound absorption, its predictions are often good an approximation to reality. Because of its simplicity, it is the one most often used unless there is some positive indication that the refinements contained in more complicated models are necessary for the problem at hand."

It may be clear that for me the notion of scaling is an indication, while the working hypothesis gave me motivation to look for a more complicated model. Note, however, that the aim was not to make things more complicated as they are but that the ultimate goal was and still is to invoke the complexity at the level of the constitutive relations rather than trying to deal with it at the level of the wave equation itself. Adding the complexity at a later stage may be much more complicated, judged, for instance, by the current findings of localization theory (Faris, 1995), where people try to come to terms

with a certain randomness in the wave or Schrödinger equation. Therefore I am reluctant in adding complexity at levels transgressing the level of the founding first principles, because of the possible mathematical implications. If one does not keep this in mind, one may end up solving problems that lost their bearings with the problem one initially set out to solve. It might very well be, that adding a large amount of complexity to the coefficients of the wave equation is an example of such a situation. Hopefully the observations made in this thesis trigger a re-evaluation of the founding principles behind the derivation of the wave equation.

Bibliography

Alexander, S. (1989), Vibrations of fractals and scattering of light from earogels, *Physical Review B*, **40**(11).

Anderson, P. W. (1958), Absence of diffusion in certain random lattices, *Physical Review Letters*, **109**:1492–1505.

Arneodo, A., Grasseau, G. and Holschneider, M. (1988), Wavelet transform of multifractals, *Physical Review Letters*, **61**(20):2281–2284.

Asch, M., Kohler, W., Postel, M. and White, B. (1991), Frequency content of randomly scattered signals, *SIAM Rev.*, **33**:519–625.

Asch, M., Papanicolaou, G., Postel, M., Sheng, P. and White, B. (1990), Frequency content of randomly scattered signals I, *Wave motion*, **12**:429–450.

Auriault, J.-L. (1991), Heterogeneous medium. Is an equivalent macroscopic description possible?, *International Journal of Engineering Science*, **29**(7).

Bacry, E., Muzy, J. F. and Arneodo, A. (1993), Singularity spectrum of fractal signals from wavelet analysis: Exact results, *Journal of Statistical Physics*, **70**(3/4):635–674.

Baluni, V. (1985), Transmission of acoustic waves in a random layered medium, *Physical Review A*, **31**(5).

Banik, N. C., Lerche, I. and Shuey, R. T. (1985a), Stratigraphic filtering, Part I: Derivation of the O'Doherty-Anstey formula, *Geophysics*, **50**(12):2768–2774.

Banik, N. C., Lerche, I. and Shuey, R. T. (1985b), Stratigraphic filtering, Part II: Model spectra, *Geophysics*, **50**(12):2768–2774.

Berkhout, A. J. (1982), *Seismic migration. Imaging of acoustic energy by wave field extrapolation*, Elsevier, Amsterdam.

Berkhout, A. J. (1987), *Applied seismic wave theory*, Elsevier, Amsterdam.

Beylkin, G. (1992), On the representation of operators in bases of compactly supported wavelets, *SIAM J. Numer. Anal.*, **29**(6):1716–1740.

Beylkin, G., Coifman, R. and Rokhlin, V. (1991), Fast wavelet transforms and numerical algorithms I, *Comm. Pure Appl. Math.*, **44**:141–183.

Beylkin, G., Coifman, R. and Rokhlin, V. (1992), Wavelets in numerical analysis, in: Ruskai, M. B., Beylkin, G., Coifman, R., Daubechies, I., Mallat, S., Meyer, Y. and Raphael, L., eds., *Wavelets and their applications*, pp. 181–210, Jones and Bartlett.

Bickel, H. (1993), Similarity and the inverse Q filter: The Pareto-Levy stretch, *Geophysics*, **58**:1629–1633.

Bleistein, N. (1984), *Mathematical methods for wave phenomena*, Academic press Inc.

Bony, J. (1983), Propagation et interaction des singularités pour les solutions de équations au derivées partielles non-liniaires, in: *Proc. of the Internations Congress of Mathematicians*, pp. 1133–1147.

Borowitz, S. (1967), *Fundamentals of quantum mechanics*, Benjamin Inc.

Burridge, B. and Chang, H. W. (1989), Multimode, one-dimensional wave propagation in a highly discontinuous medium, *Wave Motion*, **11**:231–249.

Burridge, B., Papanicolaou, G. and White, B. S. (1988), One-dimensional wave propagation in a highly discontinuous medium, *Wave Motion*, **10**:19–44.

Burridge, R., de Hoop, M. V., Hsu, K., Le, Lawrence and Norris, A. (1992), Waves in stratified viscoelastic media with microstructure, in: *Expanded Abstracts*, pp. 1369–1372, Soc. Expl. Geophys.

Burridge, R., de Hoop, M. V., Hsu, K. and Norris, A. (1993), Waves in stratified viscoelastic media with micro structure, *JASA*, **94**(5).

Carmona, R. and Lacroix, J. (1990), *Spectral theory of random Schrödinger operators*, Birkhäuser.

Chui, C. K. (1992), *An introduction to wavelets*, Academic Press, San Diego.

Claerbout, J. F. (1971), Toward a unified theory of reflector mapping, *Geophysics*, **36**:467–481.

Cohen, L. (1993), The scale representation, *IEEE Trans. Signal Process.*, **41**(12).

Cohen, L. (1995), *Time-frequency analysis*, Prentice Hall.

Collet, P. (1986), *Dynamical Systems 1986*, chap. Haussdorff dimension of the singularities for invariant measures of expanding dynamical systems, Springer-Verlag.

Collet, P., Lebowitz, J. and Porzio, A. (87), The dimension spectrum of some dynamical systems, *J. Stat. Phys.*, **47**:609–644.

Collier, M. R. (1993), On generating kappa-like distributions functions using velocity space Lévy flights, *Geophysical Research Letters*, **20**(15):1531–1534.

Daubechies, I. (1988), Orthonormal bases of compactly supported wavelets, *Communications on pure and applied mathematics*, **41**:909–996.

Daubechies, I. (1992), *Ten lectures on wavelets*, SIAM, Philadelphia.

Dautray, R. and Lions, J. L. (1988), *Mathematical analysis and numerical methods for science and technology*, vol. 2, Springer-Verlag.

Dautray, R. and Lions, J. L. (1990), *Mathematical analysis and numerical methods for science and technology, spectral theory and applications*, vol. 3, Springer-Verlag.

Dautray, R. and Lions, J. L. (1992), *Mathematical analysis and numerical methods for science and technology, evolutions problems I*, vol. 5, Springer-Verlag.

Davis, A., Marchak, A. and Wiscombe, W. (1994), *Wavelets in geophysics*, chap. Wavelet-based multifractal analysis of non-stationary and/or intermittent geophysical signals, pp. 249–298, Academic Press.

Delyon, F., Levy, Y. and Souillard, B. (1985), Anderson localization for one- and quasi-one-dimensional systems, *Journal of Statistical Physics*, **41**(3/4):375–388.

Dessing, F. J. (1997), Ph.D. thesis, Delft University of Technology, in preperation `http://wwwak.tn.tudelft.nl/~frankd`.

Duistermaat, J. J. (1993), Distributies, Lecture notes.

Einstein, A. (1905), Über die von der Molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, *Annalen der Physik*, **17**:549.

Esmersoy, C., Hsu, K. and Schoenberg, M. (1989), Quantitative analysis of fine layering effects: Medium averaging versus synthetic seismograms, in: *Expanded Abstracts*, p. 1083, Soc. Expl. Geophys.

Farge, M., Hunt, J. C. R. and Vassilicos, J. C. (1993), *Wavelets, fractals, and Fourier transforms*, Clarendon Press.

Faris, W. G. (1995), Random waves and localization, *Notices of the American Mathematical Society*, **42**(8):848–853.

Feller, W. (1950), *An introduction to probability theory and its applications*, John Wiley and Sons, Inc.

Fokkema, J. T. and van den Berg, P. M. (1993), *Seismic applications of acoustic reciprocity*, Elsevier, Amsterdam.

Folstad, P. G. and Schoenberg, M. (1993), Scattering from a set of anisotropic layers to second order in frequency, in: *55th Mtg. Eur. Assoc. Expl Geophys., Abstracts*, p. Session:P105, Eur. Assoc. Expl. Geophys.

Fürstenberg, H. (1963), Non commuting random products, *Trans. Am. Math. Soc.*, **108**(3):377.

Gabor, D. (1946), Theory of communication, *J. Inst. Elect. Eng.*, **93**:429–457.

Gel'fand, I. M. and Shilov, G. E. (1964), *Generalized functions*, vol. 1, Academic press.

Ghez, J. M. and Vaienti, S. (1992), Integrated wavelets on fractal sets: II. The generalized dimensions, *Nonlinearity*, **5**:791–804.

Gilbert, F. and Backus, G. E. (1966), Propagator matrices in elastic wave and vibration problems, *Geophysics*, **v**(31):326–332.

Golub, G. H. and van Loan, C. F. (1984), *Matrix computations*, The John Hopkins University press.

Gretener, P. E. F. (1961), An analysis of the observed traveltimes discrepancies between continuous and conventional well velocity surveys, *Geophysics*, **26**:1–11.

Griffel, D. H. (1981), *Applied functional analysis*, Wiley and Sons.

Grossmann, A., Kronland-Martinet, R. and Morlet, J. (1989), Reading and understanding continuous wavelet transforms, in: Combes, J. M., Grossman, A. and Tchamitchian, P., eds., *Wavelets: time-frequency methods and phase space*, pp. 2–20, Springer-Verlag, Berlin.

Grossmann, A. and Morlet, J. (1985), Decomposition of functions into wavelets of constant shape, and related transforms, in: Streit, L., ed., *Mathematics and physics, lectures on recent results*, World Scientific Publishing, Signapore.

Halsey, T. C., Jensen, M. H., Kadanoff, L. P., Procaccia, I. and Shraiman, B. I. (1986), Fractal measures and their singularities: The characterization of strange sets, *Phys. Rev. A*, **33**:1141–1151.

Hargreaves, N. D. (1992), Simularity and the inverse Q filter; Some simple algorithms for inverse Q filtering, *Geophysics*, **57**:944–947.

Hausdorff, F. (1918), Dimension und äußeres Maß, *Math. Analen*, **79**:157–179.

Hentschel, H. G. E. and Procaccia, I. (1983), The infinite number of generalized dimensions of fractals on strange attractors, *Physica D*, **8**:435–444.

Herrmann, F. J. (1991), *The effect of detail on wave propagation, Towards an improved macro model parametrization*, Master's thesis, Delft University of Technology.

Herrmann, F. J. and Wapenaar, C. P. A. (1992), Macro description of fine layering: A proposal for an extended macro model, in: *Expanded Abstracts*, pp. 1263–1266, Tulsa, Soc. Expl. Geophys.

Herrmann, F. J. and Wapenaar, C. P. A. (1993), Wave propagation in finely layered media, a parametric approach, in: *Expanded Abstracts*, pp. 909–912, Tulsa, Soc. Expl. Geophys.

Herrmann, F. J. and Wapenaar, C. P. A. (1994), Scaling of the pseudo primary analyzed by the wavelet transform, in: *Expanded Abstracts*, pp. 1049–1052, Tulsa, Soc. Expl. Geophys.

Hoekstra, E. V. (1996), *Multiscale analysis of seismic data by the wavelet transform*, Master's thesis, Delft University of Technology, Delft, the Netherlands, `http://wwwak.tn.tudelft.nl/DELPHI/Publications.html`.

Holschneider, M. (1987), On the wavelet transform of fractal objects, *Journal of Statistical Physics*, **50**(5/6):963–993.

Holschneider, M. (1995), *Wavelets an analysis tool*, Oxford Science Publications.

Holschneider, M. and Tchamitchian, P. (1990), Régularité locale de la function 'non-différentiable' de Riemann, in: Lemarié, P. G., ed., *Les ondelettes en 1989*, pp. 102–124, Springer Verlag, Berlin.

't Hooft, G. (1994), *Under the spell of the gauge principle*, vol. 19 of *Advanced series in Mathematical Physics*, World Scientific.

de Hoop, A. T. (1995), *Handbook of radiation and scattering of waves*, Academic Press, London.

de Hoop, M. V., Burridge, B. and Chang, H. W. (1991a), The pseudo-primary field due to a point source in a finely layered medium, *Geophys. J. Int.*, **104**:489–506.

de Hoop, M. V., Burridge, B. and Chang, H. W. (1991b), Wave propagation with tunneling in a highly discontinuous medium, *Wave Motion*, **13**:307–327.

Hsu, K. and Burridge, R. (1991), Effects of averaging and sampling on the statistics of reflection coefficients, *Geophysics*, **56**(1):50–58.

Hsu, K., Burridge, R. and Walsh, J. (1992), P-wave and S-wave drifts in a slow formation, in: *Expanded abstracts*, pp. 185–188, Soc. Expl. Geophys.

Jaffard, S. (1989), Exposants de Hölder en points donnés et coefficients d'ondelettes, *C. R. Acad. Sci. Paris*, **308**:79–81.

Jaffard, S. (1991), Pointwise smoothness, two microlocalisation and wavelet coefficients, *Publicacions Mathematiques*, **35**.

Jannane, Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, R., Tarantola, A., Trezeguet, D. and Xie, M. (1989), Wavelength of earth structures that can be resolved from seismic reflection data, *Geophysics*, **54**(7):906–910.

Kennet, B. L. N. and Kerry, N. J. (1978), Seismic waves in stratified half space, *Geophys. J.R. astr. Soc.*, **57**:557–583.

Kerner, C. (1992), Anisotropy in sedimentary rocks modeled as random media, *Geophysics*, **57**(4):564–576.

Kjartansson, E. (1981), *Constant Q-wave propagation and attenuation*, Geophysics Reprint Series.

Klafter, J., Shlesinger, M. and Zumofen, G. (1996), Beyond brownian motion, *Physics Today*.

Kohler, W., Papanicolaou, G. and White, B. (1996), Localization and mode conversion for elastic waves in randomly layered media I, *Wave motion*, **23**:1–22.

Kolmogorov, A. N. (1941), Local structure of turbulence in an incompressible liquid for very large Reynolds numbers, *Dokl. Akad. Nauk.*, **30**:299–303.

Lapidus, M. L. and Pomerance, C. (1993), The Riemann-function and the one-dimensional Weyl-Berry conjecture for fractal drums, *Proc. Londin Math. Soc.*, **3**(66):41–69.

Lawecki, P., Burridge, R. and Papanicolaou, G. (1994), Pulse stabilization in a strongly heterogeneous layered medium, *Wave Motion*, **20**:177–195.

Lawecki, P. and Papanicolaou, G. (1994), Reflection of wavefronts by randomly layered media, *Wave Motion*, **20**:245–260.

Leary, P. (1991), Deep borehole log evidence for fractal distribution of fractures in crystaline rock.

Li, X. and Haury, J. C. (1995), Characterization of heterogeneities from seimic velocity mea-

surements using the wavelet transform, in: *Expanded Abstracts*, p. 1083, Soc. Expl. Geophys.

Lichtenberg, A. J. and Lieberman, M. A. (1992), *Regular and chaotic dynamics*, vol. 38 of *Applied Mathematical Sciences*, Springer-Verlag, 2nd edn.

Mallat, S. and Hwang, L. (1992), Singularity detection and processing with wavelets, *IEEE Transactions on Information Theory*, **38**(2):617–642.

Mallat, S. G. (1989), Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$, *Trans. Amer. Math. Soc.*, **315**(1):69–87.

Mallat, S. G. and Zhong, S. (1992), Characterization of signals from multiscale edges, *IEEE Trans. Patt. Anal. Mach. Intell.*, **14**:710–732.

Mandelbrot, B. B. (1967), How long is the coast-line of Britain, *Science*, **155**:636–638.

Mandelbrot, B. B. (1974), Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier, *J. Fluid Mechanics*, **62**:331.

Mandelbrot, B. B. (1982), *The fractal geometry of nature*, Freeman and Co.

Mandelbrot, B. B. (1989), *Statistical physics*, vol. 17, chap. Negative fractal dimension and multifractals, North Holland, Amsterdam.

Mandelbrot, B. B. and Wallis, J. R. (1969), Noah, Joseph and operation hydrology, *Water Research*, **5**:228.

Marr, D. (1982), *Vision*, Freeman and Company.

le Méhauté, A. (1991), *Fractal geometries*, CRC Press Inc.

le Méhauté, A. (1995), *Introduction to wave phenomena and uncertainty in fractal space–I*, Pergamon.

Messiah, A. (1958), *Quantum mechanics*, vol. I, North Holland.

Meyer, Y. (1990), *Ondelettes et opérateurs, I: ondelettes, II: opérateurs de Calderón-Zygmund, III:* (with R. Coifman), *opérateurs multilinéaires*, Hermann, Paris, English translation of first volume is published by Cambridge University Press.

Montroll, W. and Schlessinger, F. (unknown), On the wedding of certain dynamical processes disordered complex materials to the theory of the stable (Levy) distribution functions, Institute for Physical Science and Technology.

Morlet, J., Arens, G., Fourgeau, E. and Giard, D. (1982a), Wave propagation and sampling theory-Part I: Complex signal and scattering in multilayerd media, *Geophysics*, **V**(47):203–221.

Morlet, J., Arens, G., Fourgeau, E. and Giard, D. (1982b), Wave propagation and sampling theory-Part II: Complex signal and scattering in multilayerd media, *Geophysics*, **V**(47):222–236.

Muller, J., Bokn, I. and McCauley, J. L. (1992), Multifractal analysis of petrophysical data, *Ann. Geophysicae*, **10**:735–761.

Muzy, J. F., Bacry, E. and Arneodo, A. (1993), Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus maxima method, *Physical Review E*, **47**(2):875–884.

Nottale, L. (1992), *Fractal space-time and microphysics*, World Scientific.

Nottale, L. (1995), *Scale relativity, fractal space-time and quantum mechanics*, Pergamon.

Nottale, L. (1996), Scale relativity and fractal space-time: applications to quantum physics, cosmology and chaotic systems, *Chaos, solitions and fractals*, **7**(6):877–938.

O'Doherty, R. F. and Anstey, N. A. (1971), Reflections on amplitudes, *Geophysics*, **19**:430–458.

Osoledec, V. I. (1980), A multiplication ergodic theorem, Lyapunov characteristics numbers for dynamical systems, *Trans. Moscow Math. Soc.*, **19**:197.

Ott, E. (1993), *Chaos in dynamical systems*, Cambridge university press.

Painter, S. (1995), Random fractal models of heterogeneity: the Lévy-stable approach, *Mathematical Geology*, **27**(7):813–830.

Papanicolaou, G., Postel, M., Sheng, P. and White, B. (1990), Frequency content of randomly scattered signals II, *Wave Motion*, **12**:527–549.

Parisi, G. and Frisch, U. (1985), *Turbulence and predictability in geophysical fluid dynamics and climate dynamics*, chap. A multifractal model of intermittency, pp. 84–88, North Holland.

Pastur, L. (1994), Localization theory: Basic ideas and results (tutorial lectures), Lecture notes.

Pastur, L. and Figotin, A. (1991), *Spectra of random and almost-periodic operators*, Springer-Verlag.

Pearson, D. B. (1988), *Quantum scattering and spectral theory*, Academic Press.

Pecknold, S., Lovejoy, S., Schertzer, D., Hoohe, C. and Malouin, J. (1993), The simulation of universal multifractals, *Cellular Automata – Prospects in astrophysical applications*, **10**:10.

Pierce, A. D. (1981), *Acoustics: An introduction to its physical principles and applications*, Acoustic Society of America.

Priestley, M. B. (1981), *Spectral analysis and time series*, Academic press.

Redheffer, R. (1961), *Modern mathematics for the engineer (second series)*, chap. Difference equations and functional equations in transmission theory, McGraw-Hill New York.

Reed, M. and Simon, B. (1978), *Analysis of operators*, vol. IV of *Methods of modern mathematical physics*, Academic Press.

Reed, M. and Simon, B. (1979), *Scattering theory*, vol. III of *Methods of modern mathematical physics*, Academic Press.

Reed, M. and Simon, B. (1980), *Functional analysis*, vol. I of *Methods of modern mathematical physics*, Academic Press.

Renyi, A. (1970), *Probability theory*, North Holland.

Resnick, J. R., Lerche, I. and Shuey, R. T. (1986), Reflection, transmission, and the generalized

primary wave, *Geophys. J. Roy. Astr. Soc.*, **87**:349–377.

Richardson, L. F. (1922), *Wheather prediction by numerical process*, Cambridge University Press.

Samarodnitsky, G. and Taqqu, M. (1994), *Stable non-Gaussian random processes, stochastic models with infinite variance*, Chapman and Hall.

Saucier, A. and Muller, J. (1993), Use of multifractal analysis in the characterization of geological formation, *Fractals*, **1**(3):617–628.

Schertzer, D. and Lovejoy, S. (1987a), *Fractals' physical origin and properties*, chap. Nonlinear variability in geophysics multifractal simulations and analysis, pp. 49–79, Plenum Press, New York.

Schertzer, D. and Lovejoy, S. (1987b), Physical modeling and analyis of rain and clouds by anisotropic scaling multiplicative processes, *Journal of Geophysical Research*, **92**(D8):9693–9741.

Schertzer, D. and Lovejoy, S., eds. (1993), *Nonlinear variability in geophysics 3, scaling and multifractal processes*, EGS, European Geophysical Society, American Geophysical Union.

Schmitt, F. (1993), *Turbulence develope et multifractals universels en soufflerie et dans L'Atmosphere*, Ph.D. thesis, L'Universite Pierre Curie, Paris.

Schmitt, F., Lavallee, D., Schertzer, D. and Lovejoy, S. (1992), Emperical determination of universal multifractal exponents in turbulent velocity fields, *Physical Review Letters*, **68**(3):305–308.

Schmitt, F., Lovejoy, S. and Schertzer, D. (1994), Multifractal analysis of Greenland ice-core project climate data, unpublished.

Schoenberg, M. and Muir, F. (1989), A calculus for finely layered anisotropic media, *Geophysics*, **54**(5):581–589.

Schwartz, L. (1957), *Théorie des distributions*, vol. I, Herrmann and Cie, Paris.

Schwartz, L. (1959), *Théorie des distributions*, vol. II, Herrmann and Cie, Paris.

Sedov, L. I. (1971), *A course in continuum mechanics, basic equations and analytical techniques*, vol. I, Wolters-Noordhoff, Groningen.

Shapiro, S. A. and Zien, H. (1993), The O'Doherty-Anstey formula and localization of seismic waves (short note), *Geophysics*, **58**(5):736–740.

Shapiro, S. A., Zien, H. and Hubral, P. (1994), Generalized O'Doherty-Anstey formula for waves in finely layered media, *Geophysics*, **59**(11):1750–1762.

Sheng, P. (1995), *Introduction to wave scattering, localization, and mesoscopic phenomena*, Academic Press.

Sheng, P., Zang, Z., White, B. and Papanicolaou, G. (1986), Multiple-scattering noise in one dimension: Universality through localization length scaling, *Physical Review Letters*, **57**(8):1000–1003.

Siebesma, A. P. (1989), *Multifractals in condensed matter*, Ph.D. thesis, Rijksuniversiteit

Groningen.

Souillard, B. (1986), Waves and electrons in inhomogeneous media, in: Souletie, J., Vannimenus, J. and Stora, R., eds., *Chance and matter*, vol. session XLVI, pp. 305–376, les Houches, North Holland.

Staal, J. J. (1995), *Characterizing the irregularity of measurements by means of wavelet transform: A preliminary discussion on the implications of scaling/nondifferentiability on the dynamics of waves*, Master's thesis, Delft University of Technology, `http://wwwak.tn.tudelft.nl/DELPHI/Publications.html`.

Strichartz, R. (1994), *A guide to distribution theory and Fourier transforms*, Advanced mathematics, CRC Press.

Tartarskii, V. I. (1971), *The effects of the turbulent atmosphere on wave propagation*, Israel Program for Scientific translations LTD.

Todoeschuck, J. P. and Jensen, O. G. (1989), Scaling geology and seismic deconvolution, *Pure and Applied Geophysics*, **131**(1/2).

Ursin, B. (1983), Review of elastic and electromagnetic wave propagation in horizontally layered media, *Geophysics*, **48**:1063–1081.

Vermeer, P. L. (1992), *The multiscale and wavelet transform with applications in well-log analysis*, Ph.D. thesis, Delft University of Tecnology, Delft, the Netherlands.

Walden, A. T. and Hosken, J. W. J. (1985), An investigation of the spectral properties of primary reflection coefficients, *Geophys. Prosp.*, **33**(3):400–435.

Wapenaar, C. P. A. (1996a), AVA for self-similar interfaces, *Geophysics*.

Wapenaar, C. P. A. (1996b), Inversion versus migration: A new perspective to an old discussion, *Geophysics*, **61**(3):804–814.

Wapenaar, C. P. A. and Berkhout, A. J. (1989), *Elastic wavefield extrapolation*, Elsevier.

Wapenaar, C. P. A. and Herrmann, F. J. (1996), True-amplitude migration taking fine-layering into account, *Geophysics*, **61**(3):795–803.

Wapenaar, C. P. A., Slot, R. and Herrmann, F. J. (1994), Towards an extended macro model that takes fine-layering into account, *Journal of Seismic Exploration*, **3**:245–260.

Webster (1988), *Webster's ninth new collegiate dictionary and Webster's collegiate thesaurus, first digital edition*, Merriam-Webster INC.

Weder, R. (1991), *Spectral and scattering theory for wave propagation in perturbed stratified media*, Springer-Verlag.

West, B. (1990), Sensing scintillation, *J. Optical Soc. of America*, **7**:1074–1100.

Wilcox, C. H. (1984), *Sound propagation in stratified fluids*, Springer-Verlag.

Wilson, K. G. (1983), The renormalization group and critical phenomena, *Rev. Mod. Phys.*, **55**:583–600.

Yaglom, A. M. (1987), *Correlation theory of stationary and related functions I: Basic results*, Series in statistics, Spingler Verlag.

Yilmaz, O. (1987), *Seismic data processing*, Society of Exploration Geophysicists.

Zemanian, A. H. (1965), *Distribution theory and transform analysis*, Dover Publications, INC.

Ziedler, E. (1991), *Applied functional analysis*, vol. 108 of *Applied mathematical sciences*, Springer-Verlag.

Acknowledgements

The author wishes to thank all those who contributed to the realization of this thesis. In particular I like to thank the committee, especially my promotor and toegevoegd promotor, for providing me with feedback and criticism. I also would like to thank the 'vakgroep', where I have spent some joyful and noisy years. I also would like to mention Francois Schmitt, whom I met during an EGS-meeting on multifractals and with whom I had the opportunity to discuss the scale concept in geophysics during my visits to Paris. During the same EGS-meeting I met Pier Siebesma, who first showed me a paper on localization theory, the contents of which I was only able to grasp a few years later. Via Pier I met Harm Jonker who appeared to be an expert on the use of the wavelet transform to conduct a multifractal analysis. Harm, it was and still is a joy to work with you. I also like to thank Laurent Nottale for allowing me to visit him in Paris and to stimulate me to think about the role of scale and scale dynamics in wave motion. I also would like to mention one of my students, Joes Staal, together with whom I developed a way of thinking on the potential role of scale in wave motion. Last, but not least, I would like to thank my parents, brother, and friends for supporting me all along the long traject of writing this thesis.

This thesis only appeared in time, because a large number of friends stood by me during the period where I was unexpectedly struck by ill-health. I would like to mention Frank, Harm, Edo and especially Joes.

Samenvatting

Een schalende-mediumrepresentatie: een discussie over boorgatmetingen, fractalen en golven

Recente ontwikkelingen in de seismische technologie hebben laten zien dat er een toenemende vraag bestaat naar de specifieke eigenschappen van het gesteente waarin zich de olie- en/of gasreserves bevinden. Om aan deze vraag te voldoen probeert men data die op verschillende manieren en op verschillende schalen zijn verkregen, met elkaar te verenigen. Binnen dit proefschrift betreft dit de grofschalige Verticale Seismische Profielen en seismische oppervlakte data aan de ene kant en de fijnschalige boorgatmetingen aan de andere kant. Het blijkt dat er nog veel vragen zijn omtrent de integratie van deze twee typen metingen. Het doel van dit proefschrift is een beter inzicht te verkrijgen in deze vragen.

De wijze van aanpak is de volgende. Allereerst heb ik een uitgebreide schaalanalyse gedaan op boorgatmetingen met behulp van de wavelet transformatie. Deze bestaat zowel uit een locale als een globale schaalanalyse; de locale schaalanalyse levert Hölder exponenten als afschattingen van de locale mate van differentieerbaarheid; de globale schaalanalyse geeft informatie omtrent de gegeneraliseerde fractale dimensies, de statistische momenten en het singulariteitenspectrum. Dit singulariteitenspectrum drukt de hiërarchie van globale Hölder exponenten uit. Toepassing van deze schaalanalyse op boorgatmetingen laat zien dat deze zich multifractaal gedragen. Dit betekent dat ze niet langer als ordinaire functies gerepresenteerd kunnen worden, maar dat men zijn toevlucht moet nemen tot een representatie in termen van functionalen. Functionalen kunnen een betekenis gegeven worden met behulp van een continue wavelet representatie. Door een juiste keuze voor het wavelet te maken kan het divergerende karakter van de boorgatmetingen geregulariseerd worden en kunnen bepaalde operaties weer betekenis gegeven worden, ook al moet hiervoor een afhankelijke parameter worden toegevoegd, namelijk de schaal.

Om de multi-schaalanalyse een beter theoretisch raamwerk te geven besteed ik vervolgens aandacht aan de spectrale representatie behorende bij de operatoren die de infinitesimale generator vormen voor de schuif- en dilatatie-operatie. Deze twee operaties zijn verantwoordelijk voor de vorming van de familie van continue wavelets. Deze operator benadering vormt ook de basis van het hoofdstuk dat gaat over akoestische golfbeweging. Hier is de Hamiltoniaan de belangrijkste operator in de zelf-geadjungeerde infinitesimale generator die de tijdevolutie van de Cauchy-data beschrijft. Het blijkt dat het lastig is om het singulier continue deel van de spectrale maat uit te sluiten, aangezien boorgatmetingen een schalend gedrag vertonen over een zeer groot schaalgebied. Merk op dat er aan de eigenfuncties behorende bij het singulier continue deel van het spectrum geen fysische betekenis gegeven kan worden. Deze observatie is lastig, zo niet onmogelijk, te rijmen met de noodzakelijke localiteit van de verstoringen die wordt verondersteld bij localisatietheorie. Localisatietheorie is binnen de conventionele golfvergelijking de enige robuuste theorie, die dispersie geïnduceerd door complexiteit beschrijft. De toepasbaarheid van deze theorie zou men in twijfel kunnen trekken gezien de complexiteit en het schalingsgedrag van de boorgatmetingen.

Om de lezer enig inzicht te verschaffen omtrent de schalingsanalyse en de bijbehorende fractale karakterisatie heb ik een aantal hoofdstukken opgenomen in het deel Capita Selecta. Deze hoofdstukken presenteren geselecteerde onderwerpen uit de distributietheorie en geselecteerde onderwerpen omtrent deterministische mono- en multifractalen, omtrent stochastische monofractalen en omtrent locale en globale multi-schaalanalyse door middel van de wavelet transformatie.

In de epiloog die geen officieel onderdeel van dit proefschrift vormt, schets ik mijn premature ideeën omtrent de mogelijke implicatie van schaal en schaaldynamica op akoestische golfbeweging. Dit inzicht wordt voornamelijk gevoed door het feit dat boorgatmetingen een separatie van schalen niet rechtvaardigen. Deze separatie van schalen is één van de belangrijkste pijlers van de continuum mechanica, maar daarin wordt in het algemeen geen referentie naar schaal gemaakt. Derhalve denk ik dat de vraag interessant is hoe aan de bestaande theorie volgens het correspondentieprincipe wel een referentie naar schaal en schaaldynamica toegevoegd kan worden. Dit correspondentieprincipe brengt tot uitdrukking dat de gegeneraliseerde theorie overgaat in de conventionele theorie in geval de constitutieve parameters geen schaalgedrag meer vertonen. De epiloog is een eerste aanzet voor de ontwikkeling van een gegeneraliseerde theorie.

Curriculum vitae

PERSONALIA

name	Herrmann, Felix Johan
nationality	Dutch
date of birth	June 27, 1964
place of birth	's-Gravenhage

SECONDARY EDUCATION

school	Thijmstra Atheneum, 's-Gravenhage
period	1976-1983

UNIVERSITY (M.Sc. program)

university	Delft University of Technology
faculty	Applied Physics
period	1983-1991
M.Sc. thesis	The effect of detail on wave propagation, towards an improved macro model parametrization
awards	KIVI Petroleumtechniek prijs 1991

UNIVERSITY (Ph.D. program)

university	Delft University of Technology
faculty	Applied Physics
period	1991 - 1996
Ph.D. thesis	A scaling medium representation: a discussion on well-logs, fractals and waves
awards	NWO-Talent Stipendium for the year 1997 to continue the research reported on in this thesis
presentations	• Invited speaker at NLS-workshop "Analysis of scaling phenomena with multifractals and wavelets", Doorn, The Netherlands, May 1996 • Invited speaker at AMS-meeting Benelux "Wavelets and their applications", Antwerp, Belgium, May 1996

WORK PLACEMENTS

1989	Summer research assistant at NATO SACLANT, Undersea Research Centre, La Spezia, Italy
1992	Summer research assistant at Schlumberger Cambridge Research, Great-Brittain