University of Alberta

KIRCHHOFF MIGRATION IN COMPRESSED SPACES

by

Bin Liu

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**.

in

Geophysics

Department of Physics

Edmonton, Alberta
Fall 1999

Acknowledgements

The author wishes to express her sincere thanks to her academic supervisor, Dr. Mauricio D. Sacchi, for his guidance and encouragement during the period of this research.

Special thanks are due to my parents for their unflinching support and understanding during the years of study.

Abstract

Migration is a crucial step in seismic data processing. Improving the computational efficiency of migration is very important because of the high cost associated to migrate seismic data. In this thesis, we discuss application of time-frequency analysis techniques to the problem of seismic data migration and propose two new methods which enable us to perform fast migration by using a few elements of the compressed data space.

In particular, we examine two different problems. First, we present an algorithm to migrate seismic data which have been filtered using wavelet transform. The wavelet transform is used to isolate the energy that need to be migrated. This fast migration algorithm is suitable for time migration. Our second approach entails the application of the matching pursuit algorithm to migrate post and prestack data. This algorithm is suitable for depth migration. In this case, the seismic energy that contributes to the migrate image is first located by means of a matching pursuit algorithm. The latter permits us to migrate only a few strong arrivals per trace instead of migrating the complete data volume. The approach speeds up the conventional matching pursuit depth migration by using searching and sorting methods to invert the associated travel time tables.

Contents

List of Figures

CHAPTER 1

Introduction

1.1 Background and motivation

Seismic methods play a prominent role in the search for hydrocarbons. The basic equipment for reflection seismic prospecting is a source, a receiver, and a multichannel waveform display system. A survey line is defined along the earth's surface. The source sends out impulsive sound waves, which are reflected upon subsurface layers. The receiver is used to receive the echoes of sound waves reflected from subsurface layers. The echoes can arrive to receivers from several directions at the same time. A layered material may be specified according to the reflection strength found on a record. Therefore, seismic records provide a way for geophysicists and geologists to study the earth's interior.

Seismic data processing entails the use of computers for the analysis of seismic data. The primary stages in seismic data processing are Deconvolution, Stacking, and Migration. Deconvolution acts along the time axis. It removes the seismic wavelet (the source time function modified by various effects of the earth and recording system) from the recorded seismic trace and thereby increases temporal resolution. Deconvolution achieves this goal by compressing the wavelet. Stacking also is a process of compression. In particular, the data volume is reduced to a plane of midpoint–time at zero offset (source and receiver are on the same location). The result is a stacked section. Migration is commonly applied to stacked data as a final

procedure. Migration is a spatial deconvolution process that moves the reflection events from seismic time records to their correct spatial locations. Because of its significance in interpretation, many advanced migration methods have been proposed in the past couple of decades. Migration can now be performed in different domains. such as space–time, space–frequency, wavenumber–time, or wavenumber–frequency (Claerbout, 1985).

Kirchhoff migration is often implemented in the $x-t$ domain. This technique is based on the wave equation. This method is widely used in the oil and gas industry as a result of its accuracy and low cost. More important, it is more versatile than any other migration method because it allows us to migrate the complete data or a subset of it.

There are two flavors of Kirchhoff migration: input and output based migration. Input based migration can be highly optimized by using compressed data as input. However, the situation is different for output based migration. The motivation of this thesis is to find efficient methods to perform Kirchhoff migration (input and output based) in compressed data spaces. These are data compressed by some compression technique (e.g. wavelet transform, a matching pursuit decomposition).

1.2 Scope and contribution of this work

The purpose of this thesis is to investigate new means to migrate seismic data. This thesis will be focused on Kirchhoff migration. In particular, the applications of Kirchhoff migration to data that have been compressed using the wavelet transform or a matching pursuit decomposition algorithm are studied. The goal is to recover the proper seismic image by using a few elements of the compressed data space.

This thesis also complements the work of Wang and Pann (1996) and Li et al. (1998) in matching pursuit migration. These researchers have applied a matching pursuit algorithm to migrate seismic data in the case where exists an analytical expression for the migration operator (constant velocity and $v(z)$ (lateral invariant velocity) time migration).

1.3 Thesis outline

This thesis is organized as follows:

In Chapter 2, several important aspects of seismic data migration based on the Kirchhoff integral are reviewed. This brief survey includes not only the theoretical aspects but also various important practical aspects that are needed for a proper numerical implementation.

In Chapter 3, I discuss the wavelet transform and introduce the concept of migration of wavelet transform filtered data. In this new approach, the seismic data are treated as a combination of horizontal events and dipping events. The wavelet transform is used as a tool to differentiate from the data the dipping events which are considered to be the component that contributes to hyperbola summation in migration. The new procedure allows us to find a way to properly migrate the events that contribute to the image formation process.

In Chapter 4, a new technique is devised to migrate seismic data using Kirchhoff methods, but in this case the signal is decomposed using a less elegant algorithm called matching pursuit decomposition. This algorithm is used to represent the seismic data as a superposition of isolated sources which can be migrated one at the time. The advantage is that migration is carried out on a subset of the complete data

set, the compressed space, and, therefore, the computational cost of the migration algorithm is reduced.

In the final chapter, I provide a summary of the techniques developed in this thesis. Applications that are worth further study are also discussed.

CHAPTER 2

Kirchhoff Migration

2.1 Background

Migration reveals the true structure of the subsurface of the earth. Up to the late 1960s, this was achieved by manual methods on a few picked horizons using ray tracing and timing calculations (Hagedoorn, 1954). Then around 1970 the first 'diffraction stack' migration methods became commercially available. They perform migration by summing the seismic amplitudes along a diffraction hyperbola whose curvature is governed by the medium velocity. Again, these were based on ray tracing concepts and scalar diffraction theory. In the 1970s several major developments took place. In particular, the diffraction stack method was put into the context of the Kirchhoff integral theory rather than ray theory. The Kirchhoff summation technique is similar to diffraction summation, with the addition of an amplitude and phase correction applied to the data before summation. These corrections make the summation consistent with the wave equation in that they account for spherical spreading, obliquity factor, and the phase shift factor inherent to Huygen's secondary sources. Another development (Claerbout and Doherty, 1972) was based on the idea that a stacked section can be modeled as an upcoming zero–offset wavefield generated by exploding reflectors. Using that model, migration can be conceptualized as consisting of wavefield extrapolation (in the form of downward continuation) followed by imaging. Downward continuation of wavefields can be

implemented conveniently using finite–difference solutions to the scalar wave equation. Migration methods based on such implementation are called finite–difference migration (Claerbout, 1985). After the developments on Kirchhoff summation and finite–difference migration, Stolt (1978) introduced migration by Fourier transform. This method involves a coordinate transformation from frequency (associated with time axis) to vertical wavenumber axis (associated with depth), while keeping the horizontal wavenumber unchanged. Another frequency–wavenumber migration is the phase–shift method (Gazdag, 1978). This method is based on the idea that downward continuation amounts to a phase shift in the frequency–wavenumber domain. The existence of a range of seismic migration techniques can now adequately solve most imaging problems. However this is not to imply that the appropriate migration is always used, but rather that for most imaging problems, the appropriate migration tool has been developed.

Regardless what kind of migration method is used, the purpose is the same: briefly stated, is to transform a seismic wavefield recorded at the earth's surface (time section) to an earth reflectivity map (depth section)[1]. This is done by collapsing diffractions and mapping dipping events on a stacked section to their true subsurface locations. Figure 2.1 shows that a dipping event $z_1 z_2$ in subsurface is recorded in a zero offset time section at another position $t_1 t_2$. It is clear that the reflection segment $t_1 t_2$ does not present the true position of the dipping event. When the seismic section is migrated, the segment $t_1 t_2$ is moved updip, steeped, shortened, and mapped onto its true subsurface location $z_1 z_2$. Another migration example is shown in Figure 2.2. Figure 2.2 (Top) is a synthetic zero–offset section that shows three reflection events, a bowtie shape event, a dip event, and a horizontal event. Figure 2.2 (Bottom) is the migrated section of Figure 2.2 (Top). The bowtie is untied

[1]The resulting reflectivity section is often convolved with a wavelet, since imprecise knowledge of the original recorded wavelet prevents perfect deconvolution.

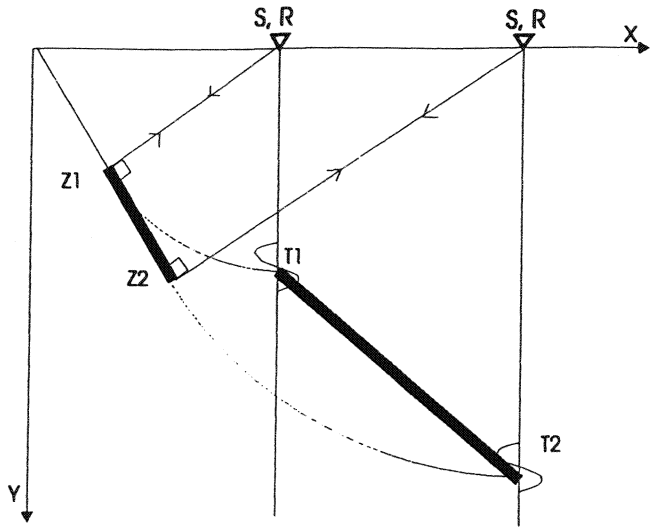


Figure 2.1: Migration of a dip event: After migration, the reflection segment t_1t_2 in time domain is mapped to the correct spatial location z_1z_2.

into a syncline; the dip event is moved updip, steeped, shortened, and moved to the true position; the horizontal event is unmoved. Figure 2.3 answers why a syncline looks like a bowtie on the stacked time section. Given the subsurface picture in Figure 2.3a, the normal–incidence rays can be computed to derive the zero–offset section in Figure 2.3b. Only five CMP locations are shown for clarity. At location 2 and 4, there are two distinct arrivals, while at location 3, there are three distinct arrivals. By filling in the intermediate raypaths and completing the procedure that trace the traveltime curve in Figure 2.3b, the bowtie character of the syncline can be constructed on the time section. Migration reveals the true structure underneath and therefore delineates the syncline. (Figure 2.2).

There are a number of concepts and assumptions made in migration theory which are fundamental to a clear understanding of present day practice; only two are introduced here. It was mentioned earlier that migration was a mapping process

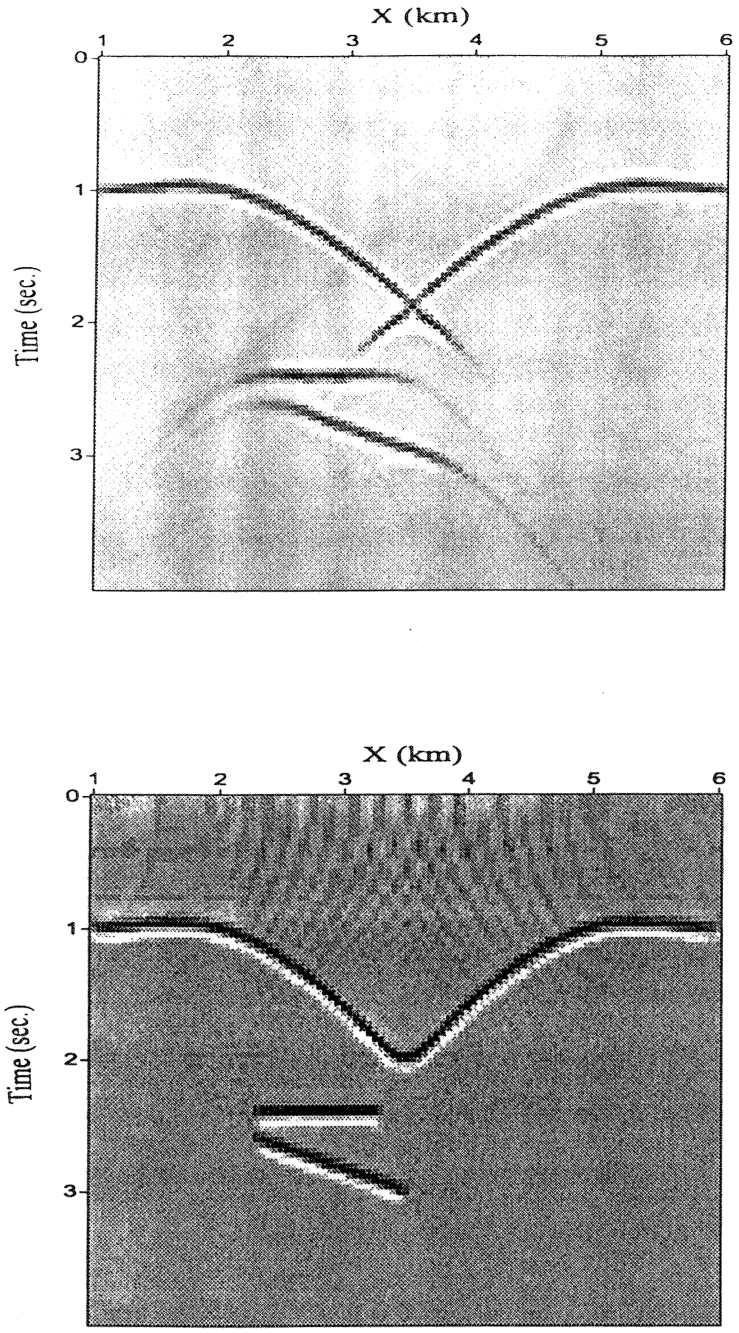


Figure 2.2: Top: Synthetic zero offset section. Bottom: Migration collapses diffraction and maps dipping events in a stacked section to their true subsurface locations.

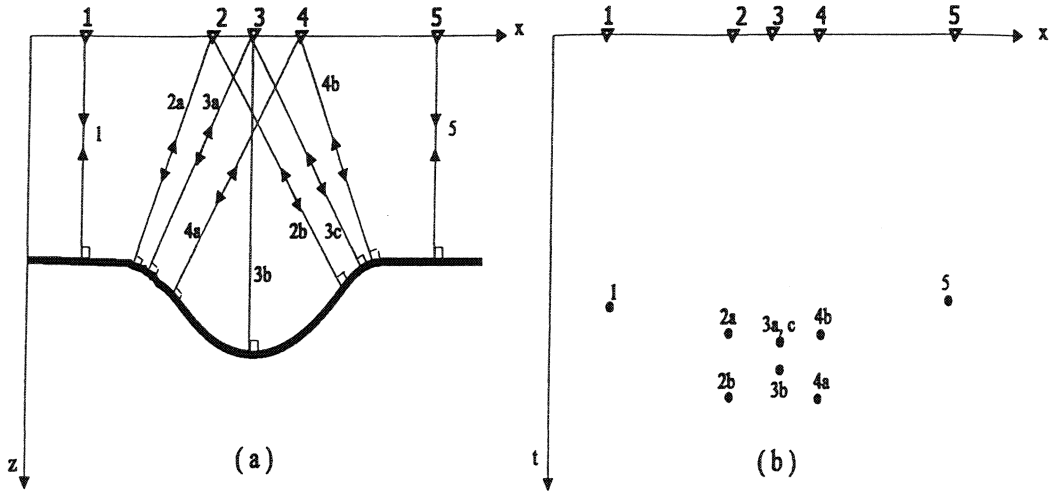


Figure 2.3: (a) A depth model consisting of a synclinal reflector. (b) The corresponding time section. Trace the bowtie in the time section. (Adapted from Yilmaz, 1987)

from surface recorded acoustic data to an earth reflectivity section. This process is known as 'depth migration'. However, the reason for the title is not the nature of the final section, but the fact that the migration process has tracked the wavefield in depth taking full account of the curvature of reflections. Another common presentation of migrated data is in terms of a time section. The time section is obtained by converting the earth reflectivity section using a suitable velocity field, or, in the case of a 'time migration', the time coordinates are the most natural output coordinates. In 'time migration' diffraction effects are considered, but not those refraction effects that are due to lateral changes in velocity (Hubral, 1977; Judson et al., 1978).

2.2 Theoretical and practical aspects of Kirchhoff migration

The Kirchhoff migration technique was developed from the Kirchhoff integral solution to the wave equation. This technique can handle most of the situations in migration (lateral variations in velocity, irregular sampling) and it is unique in its ability to migrate input traces selectively onto a pre–specified output volume.

2.2.1 From diffraction summation to Kirchhoff migration

Kirchhoff migration is closely related to diffraction summation. As mentioned before, diffraction summation is a straightforward summation of amplitudes along the hyperbolic trajectory whose curvature is governed by the velocity function. The equation for this trajectory can be derived from the geometry of Figure 2.4a. Assuming constant velocity v, from the triangle CX_AA in Figure 2.4a, we can derive:

$$t^2(x) = t_0^2 + 4x^2/v^2. \qquad (2.2.2.1)$$

Having computed the input time $t(x)$, the amplitude at input location B is placed on the output section at location A, corresponding to the output time $\tau = t_0$ at the apex of the hyperbola (Fig. 2.4b).

The diffraction summation method of migration that incorporates a weighting function w, is called Kirchhoff migration. The weighting function consists of three factors (Yilmaz, 1987):

- The obliquity factor, which describes the angle dependence of amplitudes and

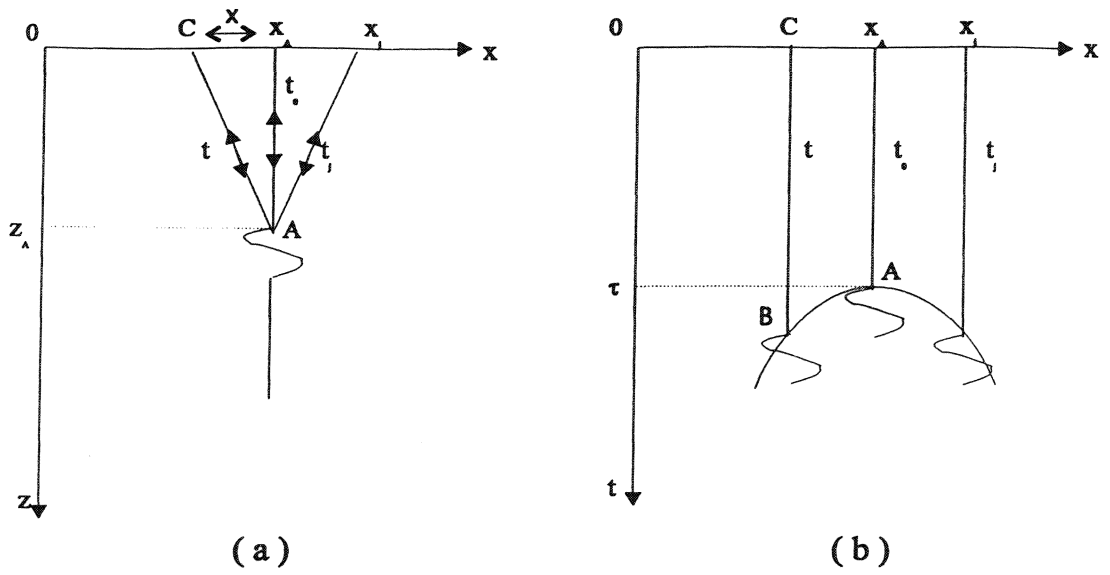


Figure 2.4: Migration considerations for a single migrated reflection event. (a) A single migrated event, (b) migration sums along a hyperbola.

is given by the cosine of the angle between the direction of propagation and the vertical axis z.

- The spherical spreading factor, which is proportional to $(1/vr)^{1/2}$ for 2–D wave propagation, and $(1/vr)$ for 3–D wave propagation.

- The phase factor (or the wavelet shaping factor), which is designed with a 45–degree constant phase spectrum and an amplitude spectrum proportional to the square root of frequency for 2–D migration. For 3–D migration, the phase shift is 90 degrees and the amplitude is proportional to frequency.

So Kirchhoff migration can be simply written as:

$$I(x,z) = \int_{r,s} w(r,s,x,z) \times D(r,s,T)\, ds\, dr. \qquad (2.2.2.2)$$

Where $I(x,z)$ is the migrated image, w is weighting factor, D is the input data at time T and $T = T_{s,x,z} + T_{r,x,z}$ is the traveltime along the raypath joining the source–receiver point (r,s) with the image point (x,z). If a zero–offset section (receiver and source are at the same location $(x',0)$) is considered and the media velocity v is constant, T in equation (2.2.2.2) can be replaced by r/v, where $r = \sqrt{(x-x')^2 + z^2}$ is the distance from the receiver (source) to the image point (x,z).

What was determined from a physical point of view in the foregoing discussion on Kirchhoff migration can be rigorously described by the integral solution to the scalar wave equation. Schneider (1978), Berryhill (1979) and Berkhout (1980) are excellent references for the mathematical treatment of the Kirchhoff migration method.

2.2.2 Integral formulation for Kirchhoff migration

Schneider (1978) discussed the mathematical formulation of migration as a solution to the scalar wave equation in which surface seismic observations are the known boundary values. In his approach, the migrated image is expressed as a surface integral over the known seismic observations when 3–D coverage exists. In the case of 2–D coverage, the wave equation migration is still possible by assuming that the subsurface and hence the surface recorded data do not vary perpendicular to the seismic profile. With this assumption, the surface integral reduces to a line integral over the seismic section. Neither the 2–D or 3–D integral migration algorithms require any approximation to the scalar wave equation (Schneider, 1978).

The scalar wave equation is

$$\nabla^2 U - \frac{1}{C^2} U_{tt} = -4\pi q(r,t). \qquad (2.2.2.3)$$

Where U is the pressure, C is the velocity and q is the source. The solution to the inhomogeneous wave equation in an arbitrary volume V_0 is given by a surface integral on S_0 enclosing V_0. A volume integral over V_0 involves both source terms and initial values. The latter can be ignored since the initial values are assumed to be zero before the shot instant, and there are no real sources in the subsurface image space, just reflectors and scatterers. In this case, only the homogeneous wave equation and inhomogeneous boundary conditions of Dirichlet type are left and the remaining surface integral is

$$U(r,t) = \frac{1}{4\pi}\int dt_0 \int dS_0 \Big[G\frac{\partial}{\partial n}U(r_0,t_0) - U(r_0,t_0)\frac{\partial}{\partial n}G\Big]. \qquad (2.2.2.4)$$

Here, the surface S_0 includes the recording surface $Z = 0$ plane and a hemisphere extending to infinity in the subsurface, n is the outward normal vector to the surface, and G is a Green function. The specific geometry of interest is shown in Figure 2.2.2.5. Contributions from the distant hemisphere are ignored, and the boundary value representation reduces to an integral over the surface involving the wavefield on S_0 and a suitable Green function G. Since $U(r_0,t_0)$ in equation 2.2.2.4 is equated to the observed seismic data, we will set $G = 0$ on S_0 in order to eliminate the gradient of U. A Green function having these properties at the free surface consists of a point source at r_0 and its negative image at r_0',

$$G(r,t \mid r_0,t_0) = \frac{\delta\left(t - t_0 - \frac{R}{C}\right)}{R} - \frac{\delta\left(t - t_0 - \frac{R'}{C}\right)}{R'}, \qquad (2.2.2.5)$$

where

$$R = \sqrt{(z-z_0)^2 + (x-x_0)^2 + (y-y_0)^2},$$

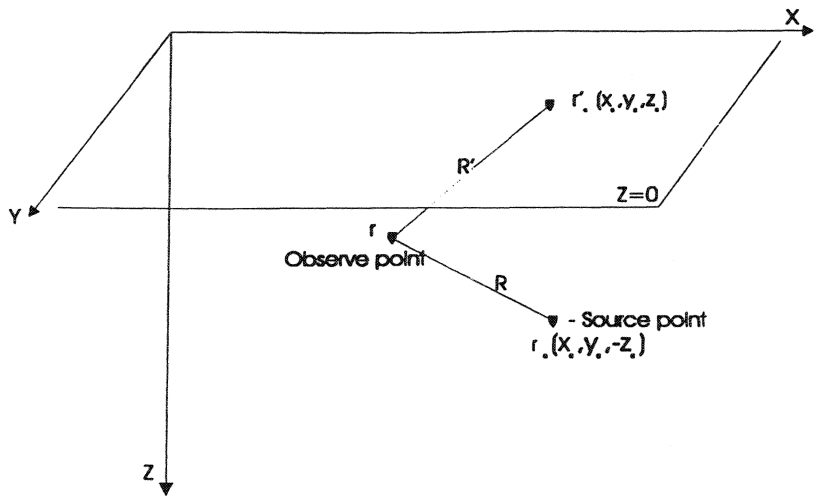


Figure 2.5: Geometry for boundary value solution.

and

$$R' = \sqrt{(z+z_0)^2 + (x-x_0)^2 + (y-y_0)^2}.$$

By substituting equation (2.2.2.5) into equation (2.2.2.4), equation (2.2.2.4) yields the following integral representation for the wavefield $U(r,t)$ at any point in the image space in terms of observations of the wavefield $U(r_0,t_0)$ on the surface:

$$U(r,t) = \frac{1}{2\pi}\int dt_0 \int dS_0 \cdot U(r_0,t_0)\frac{\partial}{\partial z_0}\left[\frac{\delta\left(t-t_0-\frac{R}{C}\right)}{R}\right]. \qquad (2.2.2.6)$$

By performing the indicated Z_0 differentiation and t_0 integration, equation (2.2.2.6) can be re–expressed as

$$U(r,t) = \frac{1}{2\pi}\int dS_0 \frac{\cos\theta}{RC}\left[U'(r_0,t_0) + \frac{C}{R}U(r_0,t_0)\right]_{t_0=t-(R/C)}. \qquad (2.2.2.7)$$

This is the 3–D Kirchhoff integral formula. This equation states Huygen's principle[2] and relates the wavefield $U(r_0, t_0)$ observed on the plane $Z = 0$ to its value at a point $U(r, t)$ in the earth's subsurface (see Figure 2.5) at an earlier time. In the seismic application, the second term in the brackets is normally ignored since it is small, but, as Schneider (1978) points out, its inclusion is trivially achieved[3]. The operations implied by equation (2.2.2.7) are simply weighting, scaling and phase shifting of data on a hyperbola. The term $cos\theta$ represents a directivity term (obliquity factor) which falls off from its value of unity at the apex of the hyperbola to a lesser value on the flanks. The factor $1/RC$ represents a true amplitude scaling factor to the derivative of the pressure field (data). The differentiation of the pressure with respect to time, when examined in the frequency domain, is equivalent to applying a $\pi/2$ phase shifting operation together with a linear high–frequency boost (a 'Newman' filter) to the pressure.

Equation (2.2.2.6) has another representation by interchanging the Z_0 derivative with a Z derivative which may then be taken outside the integral.

$$U(r,t) = -\frac{1}{\pi}\frac{\partial}{\partial z}\int dS_0 \frac{U(r_0, t - \frac{R}{C})}{R}. \qquad (2.2.2.8)$$

It can be recognized from this equation that migration can be viewed as downward continuation operation that transforms surface recorded data to deeper hypothetical recording (Schneider, 1978). The transformation is related to the point source solution to the wave equation via convolution, and it can be symbolically expressed

[2]Huygen's principle says: each point on a wavefront can be considered as the source of a small secondary wavelet that travels outward in every forward direction with the velocity of the medium at that point.

[3]In equation (2.2.2.7), because of the C/R or $1/t$ multiplier, the second term in brackets is frequently dropped giving the Rayleigh–Sommerfeld diffraction formula of optics (Goodman, 1968). To retain both terms, the seismic section needs to be differentiated and added with the same section scaled by $1/t$.

as

$$U(x,y,z,t) = U(x,y,z_0,t) * \frac{1}{2\pi}\frac{\partial}{\partial z_0}\left[\frac{\delta\left(t \pm \frac{r}{C}\right)}{r}\right], \tag{2.2.2.9}$$

where

$$r^2 = \Delta z^2 + x^2 + y^2.$$

The transformation becomes complex multiplication in frequency wavenumber domain. By taking the Fourier transform with respect to x, y and t, we obtain

$$\tilde{U}(k_x,k_y,z,\omega) = \tilde{U}(k_x,k_y,z_0,\omega)H(k_x,k_y,\Delta z,\omega), \tag{2.2.2.10}$$

where

$$H = \exp^{\pm i|\Delta z|\sqrt{\left(\frac{\omega}{C}\right)^2 - k_x^2 - k_y^2}}. \tag{2.2.2.11}$$

The transfer function H is seen to be a pure phase operator embodying the exact dispersion relation for the scalar wave equation. The operator H enables us to extrapolate waves in depth. The choice of sign in equation (2.2.2.11) determines the direction of extrapolation. However, in the operation of migration, the positive sign must be used to extrapolate converging waves back toward their origins.

Schneider also introduced the Kirchhoff integral equation in 2–D. In the 2–D case, the wavefield is independent of y,

$$U(x,y,0,t) = U(x,0,t).$$

This is true when two conditions are met: (1) subsurface geology must be independent of y, and (2) the source must either be a line source in the y–direction or the

source and receiver must be collocated as is approximately the case in CDP stack. Then from equation (2.2.2.7) the 2–D Kirchhoff integral formula can be obtained:

$$U(x,z,t) \approx \int_{-\infty}^{\infty} \frac{\cos\theta}{(R_2 c)^{1/2}} \left[\frac{\partial^{1/2} U(x_0,0,t_0)}{\partial t^{1/2}} \right]_{t_0=t+(R_2/c)} dx, \qquad (2.2.2.12)$$

where

$$R_2 = [(x_0 - x)^2 + z^2]^{1/2}.$$

In equation (2.2.2.12) the last term in the square brackets of equation (2.2.2.7) is dropped. The square root differentiation in equation (2.2.2.12) is not defined, except in the frequency domain, where it represents a nonlinear high–frequency boost followed by a $\pi/4$ phase shifting operation. The other factors appearing in equation (2.2.2.12) are the 2–D counterparts to those in equation (2.2.2.7).

In the practical implementation of equation (2.2.2.7) and equation (2.2.2.12), the integration is replaced by summation and the infinite limits in the integrals by finite limits. The replacement of integration by summation results in discretisation errors (due to the discrete nature of placing geophones on the ground), while the termination of the summation after a finite number of terms manifests itself as a 'truncation' error.

2.2.3 Practical aspects of Kirchhoff migration

Besides the errors from the discretisation of the Kirchhoff integral formula, several things can affect the quality of a migrated image. Some, such as spatial aliasing[4] and velocity uncertainty affect all migration methods. Other quality differences

[4]Spatial aliasing means insufficient sampling of data along the space axis. The data should be sampled at more than two points per wavelength (Claerbout and Black, 1993).

stem from the migration method. In Kirchhoff migration, the aperture width used in the summation and the maximum dip to migrate are also important parameters that affect the performance of Kirchhoff migration.

A migration algorithm can be tested by computing its impulse response. A band-limited impulse response is generated by using an input that contains an isolated wavelet on one trace only. The ideal migration process should produce a semicircle. Figure 2.6 (Top) shows an impulse response which indicates that Kirchhoff migration can accurately handle dips up to 90 degrees. The dip on a migration impulse response is measured as the angle θ between the vertical and a specified radial direction. The migration can be confined to a range of dips present on a seismic section. The impulse response for the dip-limited migration operator is a truncated semicircle. This is shown in Figure 2.6 (Bottom).

The migration aperture is measured in terms of the number of traces that the hyperbolic path spans. In theory, a diffraction hyperbola extends to infinite time and distance. In practice, a truncated hyperbolic summation path is used and the migration aperture is specified. The aperture width is an important parameter in the practical implementation of Kirchhoff migration. Using too small aperture width causes a dip filtering action during migration because a small aperture excludes the steeper flanks of the diffraction hyperbola from the summation. For any given time, the optimal value for the aperture width is defined by twice the maximum horizontal displacement for the steepest dip of interest in the input section (Yilmaz, 1987). The horizontal displacement in migration associated with a dipping event can be computed by the formula:

$$d_x = (v^2 t \tan\theta_t)/4. \qquad (2.2.2.13)$$

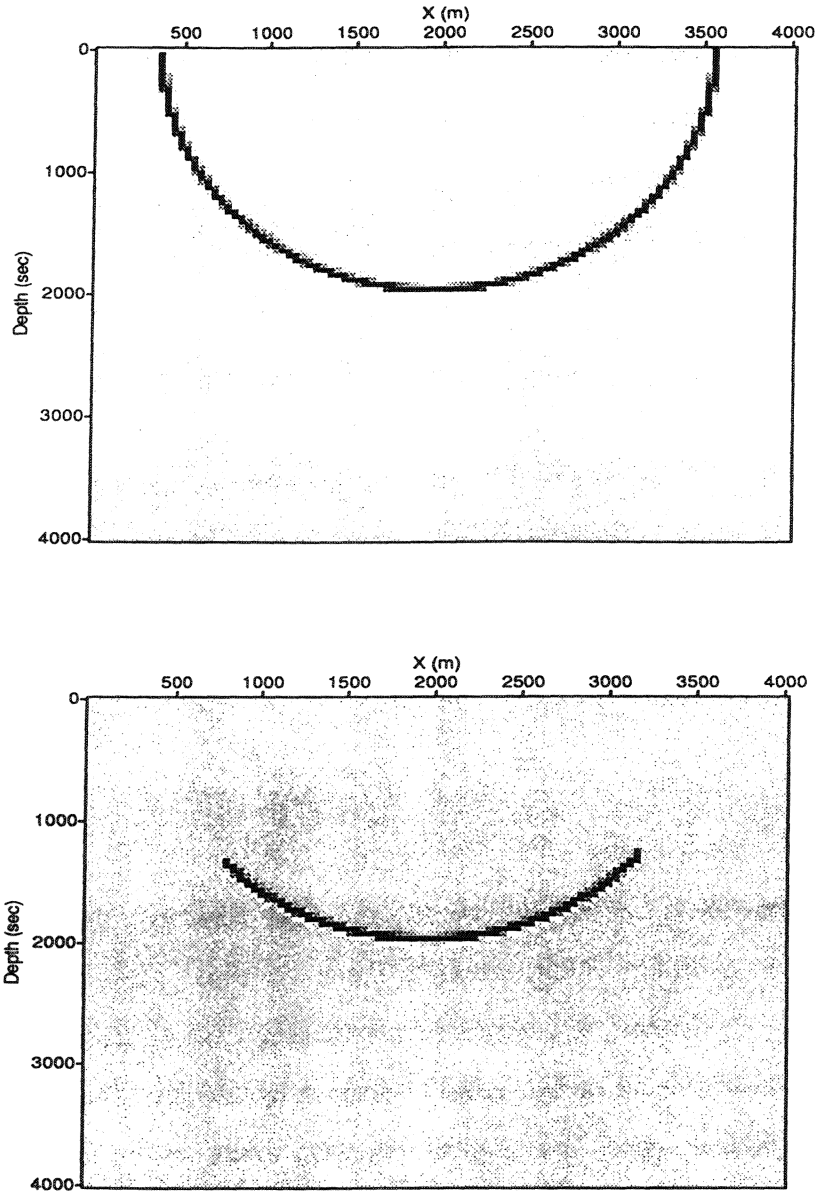


Figure 2.6: Top: A semicircle impulse response. Bottom: Dip–limited migration operator (a truncated semicircle).

Where d_x is the horizontal displacement, v is the medium velocity, t is traveltime, and θ_t the apparent dip of the reflector on the unmigrated time section. Another affect by using small aperture is that spurious horizontally dominant events may be produced. This is because the summation using very small aperture includes only the apex portion of the diffraction hyperbola, where the dip–filtering action passes flat or nearly flat events.

2.2.4 Input and output based migration techniques

In Figure 2.4, we assumed that the average velocity along the ray paths is given by v. From Figure 2.4, the following result can be drawn: the arrival times of all possible unmigrated events are given by the response at surface ($z_0 = 0$) due to a reflection at a point source (x_A, z_A). In mathematical terms,

$$\frac{(z_0 - z_A)^2}{(v/2)^2} + \frac{(x - x_A)^2}{(v/2)^2} = t^2. \qquad (2.2.2.14)$$

The left two terms are known values, the right item is a unknown variable. So this equation describes a hyperbola in the $x - t$ plane, the apex of which is given by $(x_A, 2z_A/v)$. Hence, if a single migrated event has to be constructed, then all sub–contributions have to be collected along a hyperbola. The latter addresses the following problem: given a migrated reflection event, where does the energy come from during the migration process. The problem can also be approached from the other end: where does the energy of a given unmigrated reflection event move during the migration process.

If we assume zero–offset data and consider one single reflection event with arrival time t_j at trace position x_i (Figure 2.7a), moreover, if the average velocity

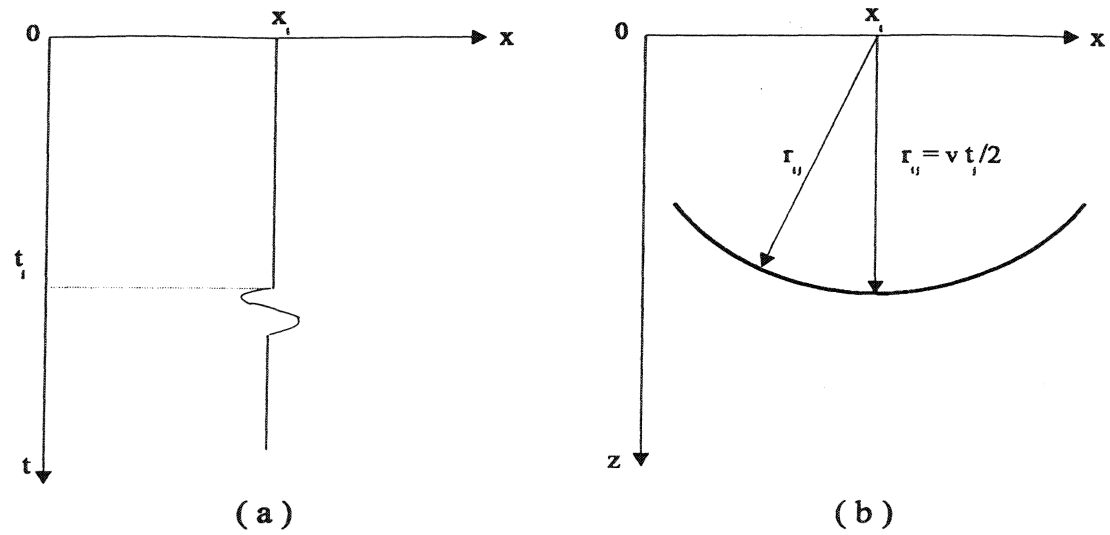


Figure 2.7: Output based Kirchhoff Migration. (a) A single reflection event. (b) Migration smears the event along a half circle.

along the travel path v is given and only straight travel paths are considered, then all possible reflection 'points' must be situated at a distance $r_{ij} = vt_j/2$ from the source–receiver position $(x_i, z_0 = 0)$ (Figure 2.7b). Mathematically speaking:

$$(z - z_0)^2 + (x - x_i)^2 = r_{ij}^2, \tag{2.2.2.15}$$

or

$$\frac{(z - z_0)^2}{(v/2)^2} + \frac{(x - x_i)^2}{(v/2)^2} = t_j^2. \tag{2.2.2.16}$$

The left two terms are variables, and the right term is a known value. So equation (2.2.2.15) describes a half circle in the $x - z$ plane, and equation (2.2.2.16) describes a vertically rescaled half circle in the $x - t$ plane, where $t = (z - z_0)/(v/2)$ represents the two way vertical travel time. Hence, if a single unmigrated reflection event is

given, the only sensible migration action that can be taken is to 'smear' the event along a migration half circle.

The above two examples illustrate the signature of Huygen's secondary source (Claerbout, 1985), which is a semicircle in the $x - z$ plane and a hyperbola in $x - t$ plane. Huygen's secondary source signature leads to two different practical migration techniques: input based and output based techniques.

Input based migration

According to Huygen's secondary source, the zero offset section can be considered that consists of a superposition of many hyperbolic traveltime responses. The migration based on this consideration consists of searching the input data in $x-t$ space for energy that would have resulted if a diffracting source (Huygen's secondary source) was located at a particular point in the output $x - z$ space. This search is carried out by summing the amplitudes in $x - t$ space along the hyperbolic curve that corresponds to Huygen's secondary source at each point in the $x - z$ space. The result of this summation is mapped onto the corresponding points in the $x - z$ space. The migration algorithm can be summarized as follows: First multiply the input data by the obliquity and spherical spreading factors. Then apply the filter $\sqrt{i\omega}$ in Fourier domain and sum along the hyperbolic path defined in equation (2.2.2.1). Finally, place the result on the migrated section at time $\tau = t_0$ corresponding to the apex of the hyperbola. This method uses the 'floating time reference' concept (Berkhout, 1984). This concept is very handy in time migration (wavefields are extrapolated with fixed extrapolation time steps ΔT, instead of fixed depth steps Δz). The imaging plane can be chosen at an integer amount of a suitable time sampling interval,

i.e.

$$T_m = m\Delta T.$$

This could be accomplished by interpolation but a more efficient approach would be to make Δz a variable such that

$$\Delta z = v_m \Delta T,$$

where ΔT is a fixed value. This technique is referred as an input based technique. Figure 2.8a shows a migration scheme for zero offset data using the input based technique.

Output based migration

In an output based technique, a reflector in the subsurface is visualized as being made up of many points that act as Huygen's secondary sources. Migration is viewed as the superposition of point source responses. To perform output based Kirchhoff migration, the filter $\sqrt{i\omega}$ can be first applied to all input traces. Then while holding the output space (the image) in the computer memory, the traces in time section are read one by one, and the points on a trace are smeared onto corresponding depth positions along with a weighting factor w. The factor w is usually computed in the inner loop of the migration algorithm. However, it may be computed in the outer loop of the migration algorithm (Gray, 1998). A migration scheme for zero offset data using an output based technique is shown in Figure 2.8b.

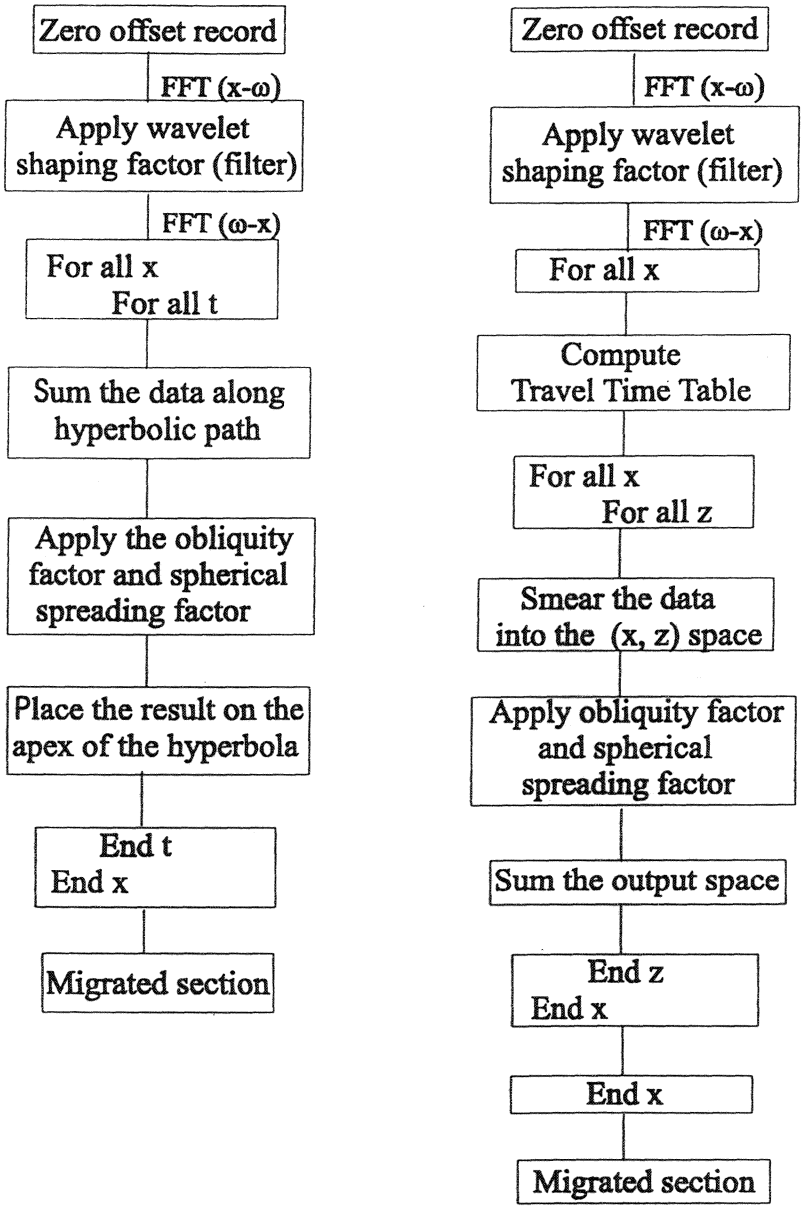


Figure 2.8: Two migration schemes. (a) Input based scheme. (b) Output based scheme.

2.2.5 Kirchhoff migration on compressed data

It was mentioned that efficiency is a critical issue in migration. Compressing the seismic data volume to a smaller one (which contains most of the energy of the data) and performing the migration on this smaller volume of data (the compressed space), may result in a fast migration algorithm. However, the situations are different for input and output based migration methods.

Input based migration can be highly optimized by using compressed data as input. For example, using the scheme discussed in Figure 2.8a, hyperbolic summation is performed for each point on an input section. Due to the fact that the compressed data is a sparse representation of the original data, the required positions where the summations need to be computed is reduced which results in reducing the computational cost of the procedure. The input based technique is best suited for time migration, where a laterally unvarying velocity field is assumed.

In complex media (with lateral variations in velocity), depth migration is required. In depth migration, an output sample location $I(x, z)$ is computed by integrating along a traveltime curve (see equation (2.2.2.2)). It is clear that for each source–receiver pair there is an associated traveltime table $T(r, s, x, z)$. These traveltime tables coincide in size with the migration output grid, and impose the requirement of using an output based migration technique. In this technique, each input trace (uncompressed or compressed) must visit each output image point (see Figure 2.8b). Therefore, when using compressed traces as input, it is difficult to improve the computational efficiency of the algorithm.

In Chapters 3 and 4, I will address two methods to overcome the difficulties encountered when we migrate compressed data.

CHAPTER 3

Migration of wavelet transform filtered data

3.1 Introduction

In seismic data processing, we usually transform acquired data to various domains in order to discriminate against coherent and incoherent noise. As an example, we can mention the 1D Fourier transform that is used to reject high frequency noise and the 2D Fourier transform that can be used to reject coherent noise like ground roll (Yilmaz, 1987). The Radon transform (Fyfe and Kelamis, 1992) is another transform that serves to reject coherent noise in this case multiple interference that tends to mask primary reflections. In this chapter, we study the application of another transform, the wavelet transform (WT), which is used not only to reject noise if needed but also to compress the seismic data. In particular, we use the WT as a means to identify where the seismic energy needs to be migrated is localized in the data domain.

In the first part of this chapter, the theory of the wavelet transform is reviewed. In the final part of this chapter, an implementation of the WT to the problem of post–stack time migration is discussed.

3.1.1 A brief history of the wavelet transform

The wavelet theory involves representing general functions in terms of simple and fixed building blocks at different scales and positions. Wavelet have a brief history. In the 70's, atomic decompositions and some formulas which are similar to the continuous wavelet transform, were in use. In 1980, Grossman and Morlet studied the wavelet transform in continuous form and suggested the word 'wavelet' for the building blocks. It became clear that these techniques could be effective substitutes for Fourier series in numerical applications. In 1985, Stephane Mallat discovered some relationships between quadrature mirror filters, pyramid algorithms, and orthonormal wavelet bases. Based on these results, in 1986, Meyer et al. constructed a new orthogonal wavelet expansion (Meyer, 1989). A couple of years later, Ingrid Daubechies (Daubechies, 1988) used Mallat's work to construct a set of wavelet orthonormal basis functions that are perhaps the most elegant, and have become the cornerstone of wavelet applications today.

3.1.2 Applications of the wavelet transform in geophysics

In the last decade, the wavelet transform and related localized transforms have been applied in various branches of science. Much of the research has occurred in the fields of mathematics and signal processing, e.g. image and video compression (Hilton et al., 1995). In geophysics, we can summarize the application of the WT as follows:

- Compression: WT is applied for compression of seismic reflection data (Bosman and Reiter, 1993; Donoho et al. 1995).
- Time–frequency signal processing: WT is used to preprocess seismic reflection data. In particular to improve SNR (Miao and Moon, 1994).

- Operator representation and image analysis: WT is used for efficient migration schemes (Dressing, 1997).
- Attribute analysis: Characterization of patterns in signals using WT (Li and Ulrych, 1996).
- Inversion: Wavelet transform inversion has been used to incorporate scale and location information in the formulation of an inversion problem (Li et al., 1994; Li et al., 1996).

3.2 The theory of wavelet transform

For extracting specific information from a given function (signal) $f(t)$, the function is transformed into a suitable representation where the desired information can be easily read. The adopted transformation depends on the nature of the information in which we are interested. Moreover, the transform should be invertable. In this section, the basic theories of wavelet transform and multiresolution analysis are reviewed. The wavelet transform is interpreted from the continuous and the discrete point of view, with a special emphasis on orthonormal wavelet bases and their properties.

3.2.1 The time–frequency representation

The Fourier transform (FT) is widely used in signal processing. A signal can be transformed into the frequency domain via

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int f(t) e^{-i\omega x} dx. \qquad (3.3.2.1)$$

The frequency content of the signal is well represented after the Fourier transform. However, the Fourier transform does not provide the time information in $\hat{f}(\omega)$ directly. The windowed Fourier transform (WFT) is a well–known method that can provide time–frequency representation of a signal (Gabor, 1946). To perform WFT, first we multiply the function $f(t)$ by a fixed function $g(t)$ (a window function which can restrict the signal f in an interval), then we compute the Fourier coefficients of the product. This procedure is repeated and each time a shifted version of g is used for the multiplication. The result of WFT leads to a family of windowed Fourier coefficients. The WFT can be written as the inner product:

$$S_{m,n}(f) =< f, g_{m,n} >, \tag{3.3.2.2}$$

where

$$g_{m,n}(t) = e^{-im\omega_0 s} g(t - nt_0).$$

Each $g_{m,n}$ is a envelope function, shifted by nt_0 and then 'filled in' with oscillations. The indices n and m indicate the time and frequency localization of $g_{m,n}$.

The wavelet transform is similar to the WFT. It computes the inner products of f with a sequence of functions $w_{m,n}$, with m indicating frequency localization, and n time localization. The main difference between WT and WFT (Daubechies, 1992) is that the window functions $g_{m,n}$ have a fixed width and $w_{m,n}$ is generated by dilations and translation of a basic wavelet w :

$$w_{m,n}(t) = a_0^{-m/2} w(a_0^{-m} t - nb_0) \tag{3.3.2.3}$$

where the w is typically well concentrated in time and in frequency and has integral zero

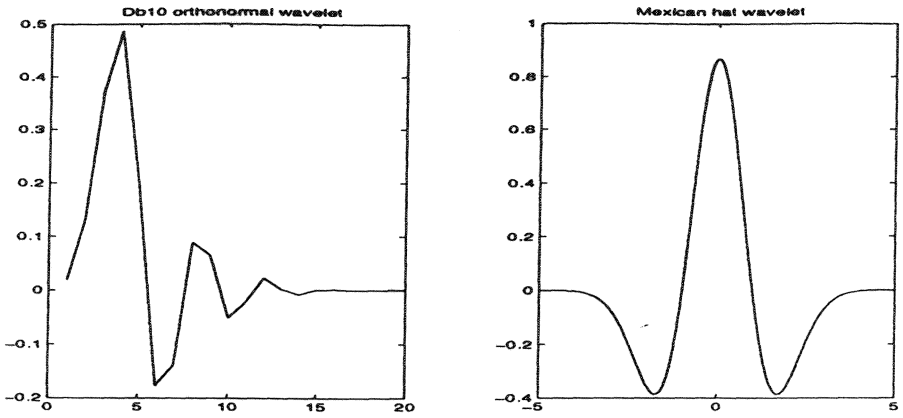


Figure 3.1: Left: Db10 wavelet. Right: Ricker wavelet

$$\int w(t)dt = 0, \tag{3.3.2.4}$$

which means it has at least some oscillations.

3.2.2 Definition of WT

The continuous wavelet transform

In pure mathematics, a wavelet has the following definition (Louis et al., 1997):
A function $w \in L^2(R)$ which satisfies the admissibility condition

$$0 < c_w := 2\pi \int_R \frac{\hat{w}(\omega)}{|\omega|} d\omega < \infty. \tag{3.3.2.5}$$

Figure 3.1 gives two examples of wavelet: Figure 3.1 (Left) is compactly–supported orthonormal wavelet 'Db10' wavelet and Figure 3.1 (Right) is 'Mexican hat', which in geophysics is called 'Ricker wavelet'.

The wavelet transform L_w of a function (signal) $f \in L^2(R)$ with respect to the wavelet w is defined as

$$L_w f(a,b) = \frac{1}{\sqrt{c_w}} |a|^{-1/2} \int_R f(t) w\Big(\frac{t-b}{a}\Big) dt, \tag{3.3.2.6}$$

where, $a, b \in R, a > 0$. Here the dilation and translation parameters a, b vary continuously over R. Another ingredient in this equation is a symmetry of concentration in $|\hat{w}(\omega)|^2$, with respect to the measure $|\omega|^{-1} d\omega$, on positive and negative frequency axes. This requirement is automatically satisfied if w is real.

The inverse wavelet transform is given by:

$$f(x) = \frac{1}{\sqrt{c_w}} \int_R \int_R L_w f(a,b) \frac{1}{\sqrt{|a|}} w\Big(\frac{x-b}{a}\Big) \frac{dadb}{a^2}. \tag{3.3.2.7}$$

Discrete wavelet transform: frames

The wavelet family (equation 3.3.2.3) and the wavelet transform (equation 3.3.2.6) can be viewed as discretized versions of the continuous wavelet transform, with a, b restricted to $a = a_0^m$ and $b = nb_0a_0^m$. In the discrete case, there does not exist, in general, a 'resolution of the identity' formula analogous to (3.3.2.7) for the continuous case (Daubechies, 1992). Reconstruction of f from the $L_{m,n}(f)$ raises the following questions:

1. Can we characterize f completely by knowing $L_{m,n}(f)$?

2. Can we reconstruct f in a numerically stable way from $L_{m,n}(f)$?

To answer these questions we have to introduce the concept of *Frames*. Daubechies (1992) gave the definition of a Frame as:

A family of functions $\{g_{mn}, m \in J\}$ *in a Hilbert space is called a frame if there exist* $A > 0, B < \infty$ *so that, for all* f *in the Hilbert space,*

$$A\|f\|^2 \leq \sum_{m \in J} | < f, g_{mn} > |^2 \leq B\|f\|^2. \tag{3.3.2.8}$$

A and B are called the frame bounds. Four types of frames are given (Daubechies, 1992):

1. $\{g_{mn}\}$ is called a *tight frame* for $A = B \neq 1$.
2. $\{g_{mn}\}$ forms an orthonormal basis if $A = B = 1$ and $|g_{mn}| = 1$.
   consequently,

$$f = \sum_{m} < f, g_{mn} > \underline{g}_{mn}. \tag{3.3.2.9}$$

3. $\{g_{mn}\}$ is called a *snug frame* for $A \neq B$ and B/A close to 1.
4. For $B/A \gg 1$, the decomposition and reconstruction become less stable.

Daubechies (1990) showed that since the wavelets $w_{m,n}$ constitute a wavelet frame for a wide range of values a_0 and b_0, the reconstruction of f from its wavelet representation $\{f, w_{m,n}\}$ may be carried out.

3.2.3 Wavelets and filter banks

The wavelet transform $L_w f$ can be interpreted as a convolution of f with the dilated wavelet $w(\cdot/a)$ for every fixed value of a. Because of the admissibility condition in

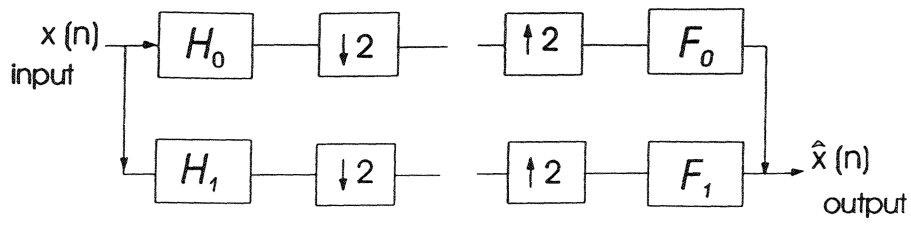


Figure 3.2: A two–channel filter bank.

equation (3.3.2.5), we can show that $\hat{w}(0) = 0$ and $w \in L^1(R) \cap L^2(R)$. It follows that $\lim_{\omega \to \infty} \hat{w}(\omega) = 0$ and w is actually a band pass filter (Louis et al., 1997). Therefore, the properties of the wavelet transform can be analyzed via filter banks. Figure 3.2 shows a two–channel filter bank. It is a set of filters, linked by sampling operators and sometimes by delays (Strang, 1997). The down–sampling operators are decimators and the up–sampling operators are expanders. In this two–channel filter bank, the analysis filters H_0 and H_1 are respectively lowpass and highpass filters. We can use the Haar filter banks[1] as an example to illustrate some properties of the filter banks.

In what follows, I will develop the basis to understand the idea behind filter banks. Let us start with a simple model composed of two signals $x(n)$ and $y(n)$. Let $x(n)$ be an input signal and $y(n)$ the output signal,

$$y(n) = \frac{1}{2}x(n) + \frac{1}{2}x(n-1). \tag{3.3.2.11}$$

[1]The Haar filter banks are based on the Haar wavelet which is given by

$$w(x) = \begin{cases} 1 & 0 \le x \le 1/2 \\ -1 & 1/2 \le x < 1. \\ 0 & \text{otherwise} \end{cases} \tag{3.3.2.10}$$

In equation (3.3.2.3), if we choose $a_0 = 2$ and $b_0 = 1$, the set $\{w_{m,n}\}$ is the Haar wavelet and forms an orthonormal basis for $L^2(R)$.

This model defines a lowpass filter. Its output at time $t = n$ is the average of the input $x(n)$ at that time and the input $x(n-1)$ at the previous time. In matrix structure, $y = Hx$:

$$\begin{bmatrix} . \\ y(-1) \\ y(0) \\ y(1) \\ . \end{bmatrix} = \begin{bmatrix} . & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ & \frac{1}{2} & \frac{1}{2} & & \\ & & \frac{1}{2} & \frac{1}{2} & \\ & & & . & . \end{bmatrix} \begin{bmatrix} . \\ x(-1) \\ x(0) \\ x(1) \\ . \end{bmatrix}. \qquad (3.3.2.12)$$

In this filter, the filter coefficients are $h(0) = \frac{1}{2}$, $h(1) = \frac{1}{2}$. In general, for a lowpass filter, the total output is

$$\begin{aligned} y(n) &= h(0)x(n) + h(1)x(n-1) + h(2)x(n-2) + \ldots \\ &= \sum_k h(k)x(n-k). \end{aligned} \qquad (3.3.2.13)$$

In frequency domain, the filter has the following Fourier transformation:

$$H(\omega) = \left(\cos\frac{\omega}{2}\right)e^{-i\omega/2}, \qquad (3.3.2.14)$$

The magnitude and the phase are given by:

$$|H(\omega)| = \cos\tfrac{\omega}{2} \quad \text{and} \quad \phi(\omega) = -\tfrac{\omega}{2}. \qquad (3.3.2.15)$$

Figure 3.3 (Left) is the plot of the magnitude $|H(\omega)|$ against the freqency ω. The cosine of $\frac{\omega}{2}$ drops to zero at $\omega = \pi$, the high frequency is wiped out. Figure 3.3 (Right) shows graph of $\phi(\omega) = -\frac{\omega}{2}$. It is a straight line, which means 'linear phase'. It reflects the fact that the filter coefficients $\frac{1}{2}$ and $\frac{1}{2}$ are symmetric.

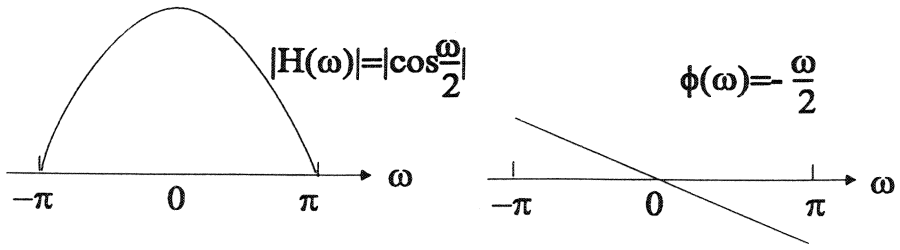


Figure 3.3: Left: The magnitude $|H(\omega)| = \cos\frac{\omega}{2}$. Right: The phase of $H(\omega)$.

A lowpass filter takes the 'average'. A highpass filter takes the 'difference'. The latter enhances the high frequency components in the signal:

$$y(n) = \frac{1}{2}x(n) - \frac{1}{2}x(n-1). \tag{3.3.2.16}$$

The filter coefficients are $h(0) = \frac{1}{2}$ and $h(1) = -\frac{1}{2}$. Equation (3.3.2.16) is a convolution $y = h * x$, which in matrix form can be written as follows:

$$\begin{bmatrix} \cdot \\ y(-1) \\ y(0) \\ y(1) \\ \cdot \end{bmatrix} = \begin{bmatrix} \cdot & & & & \\ -\frac{1}{2} & \frac{1}{2} & & & \\ & -\frac{1}{2} & \frac{1}{2} & & \\ & & -\frac{1}{2} & \frac{1}{2} & \\ & & & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot \\ y(-1) \\ y(0) \\ y(1) \\ \cdot \end{bmatrix}. \tag{3.3.2.17}$$

At frequency ω, the input vector can be expressed as $x(n) = e^{in\omega}$. The highpass output is

$$\begin{aligned} y(n) &= \frac{1}{2}e^{in\omega} - \frac{1}{2}e^{i(n-1)\omega} \\ &= (\frac{1}{2} - \frac{1}{2}e^{i\omega})e^{in\omega} \\ &= H_1(\omega)e^{in\omega}. \end{aligned} \tag{3.3.2.18}$$

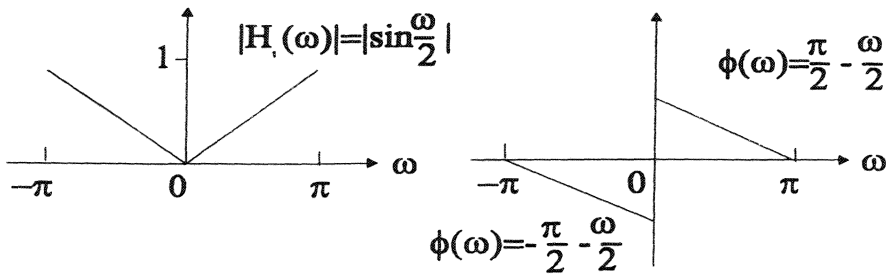


Figure 3.4: Left: The magnitude $|H_1|$. Right: The phase of highpass filter.

The quantity $\frac{1}{2} - \frac{1}{2}e^{i\omega}$ is the highpass response and

$$H_1(\omega) = \left(\sin\frac{\omega}{2}\right) ie^{-i\omega/2}. \tag{3.3.2.19}$$

The magnitude, $|H_1(\omega)| = |\sin\frac{\omega}{2}|$, has a zero response at $\omega = 0$. The phase factor of the filter has a discontinuity at $\omega = 0$, and is linear at any other point (Figure 3.4).

In any reasonable sense, the lowpass and highpass filters separately are not invertable. H_0 removes the highest frequency $\omega = \pi$ and H_1 removes the lowest frequency $\omega = 0$. However, together, they separate the signal into frequency bands and constitute the beginning of a filter bank. If the action of filtering produces a new signal with a length equal to the length of the original signal, the action of applying a high pass and low pass filter will double the amount of samples of the original signal. The filters are applied independently to $x(t)$, not in cascade. In matrix form

$$\begin{bmatrix} H \\ H_1 \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} y_L \\ y_H \end{bmatrix}, \tag{3.3.2.20}$$

the size of the left, middle and right matrixes are $2n \times n$, $n \times 1$, and $2n \times n$, respectively. Now we have a system with more equations that unknowns. We will

see that decimating the output serves to render a system of $n \times n$ equations.

The operation of downsampling is called decimation. It is represented by symbol $(\downarrow 2)$. Decimation removes a vector every other component. e.g.

$$(\downarrow 2)\begin{bmatrix} \cdot \\ x(-1) \\ x(0) \\ x(1) \\ x(2) \\ \cdot \end{bmatrix} = \begin{bmatrix} \cdot \\ x(-2) \\ x(0) \\ x(2) \\ x(4) \\ \cdot \end{bmatrix}. \tag{3.3.2.21}$$

This operation is not invertable. The recovery of x from $(\downarrow 2)x$ is possible if $X(\omega)$ is zero over a half–band of frequencies. Such a signal is 'band–limited'. It may be limited to the upper half–band or the lower half–band: $X(\omega) = 0$ for $0 \leq |\omega| < \frac{\pi}{2}$ or $X(\omega) = 0$ for $\frac{\pi}{2} \leq |\omega| < \pi$. For band–limited signals the odd–numbered components can be recovered from the even–numbered components (Shannon Sampling Theorem). Below we show a 'downsampling matrix ' which is the identity matrix with the odd–numbered rows removed:

$$\begin{bmatrix} \cdot & 1 & & & & & \\ & 0 & 0 & 1 & & & \\ & & & 0 & 0 & 1 & \\ & & & & & \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot \\ x(-2) \\ x(-1) \\ x(0) \\ x(1) \\ x(2) \\ \cdot \end{bmatrix} = \begin{bmatrix} \cdot \\ x(-2) \\ x(0) \\ x(2) \\ \cdot \end{bmatrix}. \tag{3.3.2.22}$$

Since the rows of this downsampling matrix are orthonormal, the matrix multiplication $(\downarrow 2)(\downarrow 2)^T$ must be the identity matrix. The operator $(\downarrow 2)^T$ is important. The transpose of downsampling is upsampling:

$$(\downarrow 2)^T = (\uparrow 2). \tag{3.3.2.23}$$

Upsampling places zeros into the odd–numbered components, e.g.

$$(\uparrow 2)\begin{bmatrix} \cdot \\ y(-1) \\ y(0) \\ y(1) \\ \cdot \end{bmatrix} =) \begin{bmatrix} \cdot \\ y(-1) \\ 0 \\ y(0) \\ 0 \\ y(1) \\ 0 \\ \cdot \end{bmatrix}. \tag{3.3.2.24}$$

The matrix form of upsampling is:

$$\begin{bmatrix} \cdot & & \\ 1 & 0 & \\ & 0 & \\ & 1 & 0 \\ & & 0 \\ & & 1 \quad 0 \\ & & \quad \cdot \end{bmatrix} \begin{bmatrix} \cdot \\ y(-1) \\ y(0) \\ y(1) \\ \cdot \end{bmatrix} = \begin{bmatrix} \cdot \\ y(-1) \\ 0 \\ y(0) \\ 0 \\ y(1) \\ 0 \\ \cdot \end{bmatrix}. \tag{3.3.2.25}$$

Multiply the matrixes of $(\downarrow 2)$ and $(\uparrow 2)$, the result is an identity matrix which mean that after upsampling and then downsampling the original signal is recovered. I a filter bank, downsampling is first used, and then upsampling is applied. In thi particular order, $(\downarrow 2)$ removes the odd-numbered components and $(\uparrow 2)$ puts i zeros. So $(\uparrow 2)(\downarrow 2)$ replaces the lost components by zeros:

$$(\uparrow 2)(\downarrow 2) = \begin{bmatrix} 1 & & & & & \\ & 0 & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & 1 & \\ & & & & & \cdot \end{bmatrix}. \qquad (3.3.2.26)$$

The essential part of the filter bank is that the downsampling acts on the outputs of two separate lowpass and highpass filters. Therefore, although only even-numbered components of the two outputs are kept after downsampling, the information of the original function is still preserved. For compensating the loss of half of the components in $(\downarrow 2)$, a $\sqrt{2}$ should be applied. Therefore the lowpass and highpass filters can be rewritten as $C(\omega)$ and $D(\omega)$:

$$\begin{aligned} \text{lowpass: } C(\omega) &= \sqrt{2}H_0(\omega) \\ \text{highpass: } D(\omega) &= \sqrt{2}H_1(\omega). \end{aligned}$$

In discrete time domain, those filters are $c(n)$ and $d(n)$. The two steps, downsampling and lowpass filtering remove the odd-number rows of the filter matrix C. The combination of filtering by C and decimation by $(\downarrow 2)$ is represented by a rectangular matrix L:

$$L = (\downarrow 2)C = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & & & \\ & & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \\ & & & & \cdot\ \cdot \end{bmatrix}. \tag{3.3.2.27}$$

The entries are $c(0)$ and $c(1)$ but half the rows of C have disappeared. Similarly the decimated highpass filter is represented by a rectangular matrix $B = (\downarrow 2)D$. Removing half the rows of D leaves the matrix B with a double–shift:

$$B = (\downarrow 2)D = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & & & \\ & & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \\ & & & & \cdot\ \cdot \end{bmatrix}, \tag{3.3.2.28}$$

L and B can be fitted into a single square matrix:

$$\begin{bmatrix} L \\ B \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & & & \\ & & 1 & 1 & \\ & & & & \cdot\ \cdot \\ -1 & 1 & & & \\ & & -1 & 1 & \\ & & & & \cdot\ \cdot \end{bmatrix}. \tag{3.3.2.29}$$

This matrix represents the analysis bank in a filter bank. All rows and columns are unit vectors and they are mutually orthogonal. This combined square matrix is invertible and the inverse is the transpose:

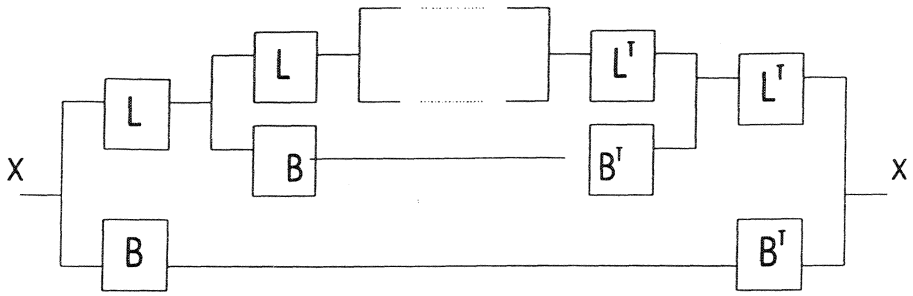


Figure 3.5: A tree of filter bank.

$$\begin{bmatrix} L \\ B \end{bmatrix}^{-1} = \begin{bmatrix} L^T & B^T \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 & & \\ 1 & 1 & & \\ & & 1 & -1 \\ & & 1 & 1 \\ & & & \\ & & & \end{bmatrix}. \qquad (3.3.2.30)$$

Because of the orthogonality,

$$\begin{bmatrix} L^T & B^T \end{bmatrix} \begin{bmatrix} L \\ B \end{bmatrix} = L^T L + B^T B = I. \qquad (3.3.2.31)$$

Equation (3.3.2.31) clearly suggests that a signal can be reconstructed by an inverse bank. This is also a two step procedure— upsampling and filtering. The inverse bank is also called synthesis bank.

In Figure 3.5, we illustrate a tree of filter banks. In this case, the lowpass part is further decomposed into 2 new signals. The relationship between filter banks and wavelets can be easily understood via the idea of multiresolution.

3.2.4 Multiresolution

Multiresolution analysis

A function $f(t)$ can be decomposed (projected) in a sequence of subspaces V_j. That means that there is a piece of $f(t)$ in each subspace. Those pieces (or projections) give finer and finer details of $f(t)$. A multiresolution decomposition of $f(t)$ can be described with the following requirements:

1. There is an increasing sequence of subspaces V_j (complete in L^2).
2. There is a wavelet subspace W_j which is the difference between V_j and V_{j+1}. Therefore
$$V_j + W_j = V_{j+1}.$$
3. If $f(t) \in V_j$ then $f(2t) \in V_{j+1}$.
4. If the basis $\phi(t) \in V_0$ then $\phi(t-k) \in V_0$ and if $w(t) \in W_0$ then $w(t-k) \in W_0$.
5. The basis $\phi(2^j t - k) \in V_j$ and $w(2^j t - k) \in W_j$.

In the theory addressed above, the first property tells us that each V_j is contained in the next V_{j+1}:

$$V_0 \subset V_1 \subset \ldots \subset V_j \subset V_{j+1} \subset \ldots$$

A function in the whole space has a piece in each subspace. Those pieces contain more and more of the full information of $f(t)$, e.g. the function $f(t)$ has a piece $f_j(t)$ in the subspace V_j and a piece $f_{j+1}(t)$ in the subspace V_{j+1}. Therefore, we can also write the first property as:

$$f_0(t) \subset f_1(t) \subset \ldots \subset f_j(t) \subset f_{j+1}(t) \subset \ldots$$

In the second property, by identifying the second family of subspaces W_j as

$$V_j \bigoplus W_j = V_{j+1},$$

(symbol $\bigoplus$ means 'direct sum' as V_j and W_j intersect only at zero vector) we can derive that the subspaces W_j contain the new information $\Delta f_j(t)$ and

$$f_j(t) + \Delta f_j(t) = f_{j+1}(t).$$

Further decomposition of the subspace V_j or the function f_j leads to:

$$V_0 \bigoplus W_0 \bigoplus W_1 \bigoplus \ldots \bigoplus W_j = V_{j+1},$$

or

$$f_0(t) + \Delta f_0(t) + \Delta f_1(t) + \ldots + \Delta f_j(t) = f_{j+1}(t).$$

The third property states a dilation requirement:

$$f \in V_j \Longleftrightarrow f(2x) \in V_{j+1}.$$

This means that if we assume $f(t) = c_{-j}e^{-i(-j)t} + \ldots + c_j e^{-ijt}$ (c_j are the Fourier coefficients),

$$f_j(t) = \sum c_k e^{ikt} \text{ for } |k| \leq 2^j,$$

$$f_{j+1}(t) = \sum c_k e^{ikt} \text{ for } |k| \leq 2^{j+1},$$

and Δf_j contains all frequencies between 2^j and 2^{j+1}:

$$\Delta f_j(t) = \sum c_k e^{ikt} \text{ for } 2^j < |k| \leq 2^{j+1}.$$

The fourth and fifth properties discuss the translation requirement. The translation leads to the fundamental requirement of time–invariance (shift invariance) in signal processing. The subspaces are shift–invariant: If $f_j(t)$ is in V_j then so are all its translates $f_j(t-k)$. If we suppose $f(t)$ is in V_0, then $f(2t)$ is in V_1 and so is $f(2t-k)$.

By induction, $f(2^j t)$ is in V_j and so is $f(2^j t - k)$. The requirement concerning a basis for each space V_j is that there exists $\phi(t)$ so that $\phi(t-k)$ is an orthonormal basis for V_0. When the functions $\phi(t-k)$ are an orthonormal basis for V_0, the rescaled functions $\sqrt{2}\phi(2t-k)$ will be an orthonormal basis for V_1. At scaling level j, the basis functions $\phi(2^j t - k)$ are normalized by $2^{j/2}$. The requirements discussed above are referred to as 'multiresolution analysis' (Mallat, 1989). From the multiresolution analysis, we can derive that a function $f_j(t)$ in V_j which has the basis $\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k)$ can be expressed as

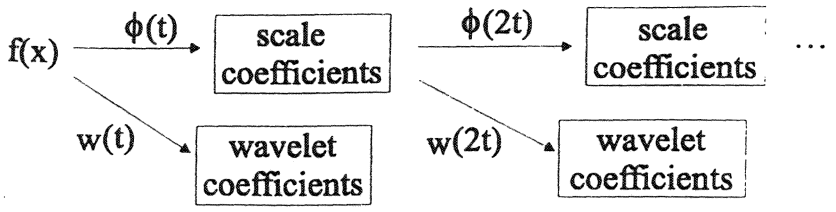
$$\begin{aligned} f_j(t) &= \sum_{k=-\infty}^{\infty} a_{jk}\phi_{jk}(t) \qquad (3.3.2.32) \\ &= \sum_k a_{0k}\phi_{0k}(t) + \sum_k b_{0k}w_{0k}(t) + \sum_k b_{1k}w_{1k}(t) + \ldots + \sum_k b_{(j-1)k}w_{(j-1)k}(t). \end{aligned}$$

Where ϕ is scale function, a_{jk} are the scale coefficients, w is wavelet function, and b_{jk} are the wavelet coefficients[2]. Equation (3.3.2.32) describes the nature of the wavelet transform. Figure 3.6 shows the parallelism between a filter bank tree in discrete time (Figure 3.6a) and multiresolution in continuous time (Figure 3.6b) is almost perfect.

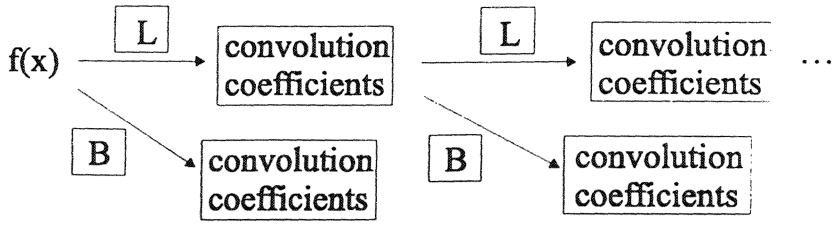
Relationship between wavelets and filter banks

If we suppose $\phi(t)$ is in V_0, then $\phi(t)$ is also in V_1 as the subspace V_0 is contained in V_1. Therefore, $\phi(t)$ must be a combination of the basis function $2^{1/2}\phi(2t-k)$ for the subspace V_1. The coefficients in the combination will be called $c(k)$. If we bring the factor $2^{1/2}$ outside:

[2]Scale coefficients a_{jk} and wavelet coefficients b_{jk} are the coefficients after the wavelet transform. They are also called approximation and detail coefficients, respectively.



(a) A filter tree



(b) Multiresolution

Figure 3.6: The parallelism between (a) a filter bank tree and (b) multiresolution.

$$\phi(t) = \sqrt{2}\sum_{k}^{N} c(k)\phi(2t-k). \tag{3.3.2.33}$$

This is the dilation equation. To find $c(k)$, we multiply the dilation equation (3.3.2.33) by $\sqrt{2}\phi(2t-n)$ and integrate after using the orthogonality condition:

$$\sqrt{2}\int_{\infty}^{-\infty} \phi(t)\phi(2t-n)dt = c(n). \tag{3.3.2.34}$$

From the orthogonality of the basis $\{\phi(2t)\}$ it can be proved that we have double shift orthogonality of the dilation coefficients $c(k)$ (Strang and Nguyen, 1997):

$$\text{Double-shift}: \sum c(k)c(k-2m) = \delta(m).$$

Because $\phi(t)$ has unit energy:

$$\text{Unit vector}: \sum |c(k)|^2 = 1.$$

Therefore, the dilation equation has brought us back to filter banks—where the key matrix is $L = (\downarrow 2)C$. Double–shift orthogonality becomes $LL^T = I$. The rows of L contain the double shifts $L_{ij} = c(2i - j)$ (see equation (3.3.2.27)). Since the subspace W_0 is in V_1, we can write the following equation for the wavelet:

$$w(t) = \sqrt{2} \sum_{k=0}^{N} d(k)\phi(2t - k). \tag{3.3.2.35}$$

The coefficients $d(k)$ are the coefficients of a highpass filter. Note that because of the orthogonality between V_0 and W_0, if $c(k)$ is lowpass filter then $d(k)$ must be a highpass filter.

3.2.5 Wavelet analysis

The wavelet transform is a mapping process. According to the theory of wavelet and multiresolution, the mapping is defined on the time–scale plane. There are two variables associated with this plane: time and scale. Therefore after the WT, the signal is divided into different scales of resolution, rather than different frequencies.

We will denote the transform level by j, Let us assume that we can decompose the signal in J levels. According to the multiresolution theory, a signal f can be represented as

$$f = A_0 = A_J + \sum_{j \leq J} D_j.$$

Where A_0 is the approximation at level zero (the signal), A_J is the approximation of the signal f at level J, and D_j $(j \leq J)$ are the details of the signal f in J levels. The words 'approximation' and 'detail' are justified by the fact that A_1 is an approximation of A_0 taking into account the 'low frequency' of A_0, whereas the detail D_1 corresponds to the high frequency correction. Figure 3.7 shows a real

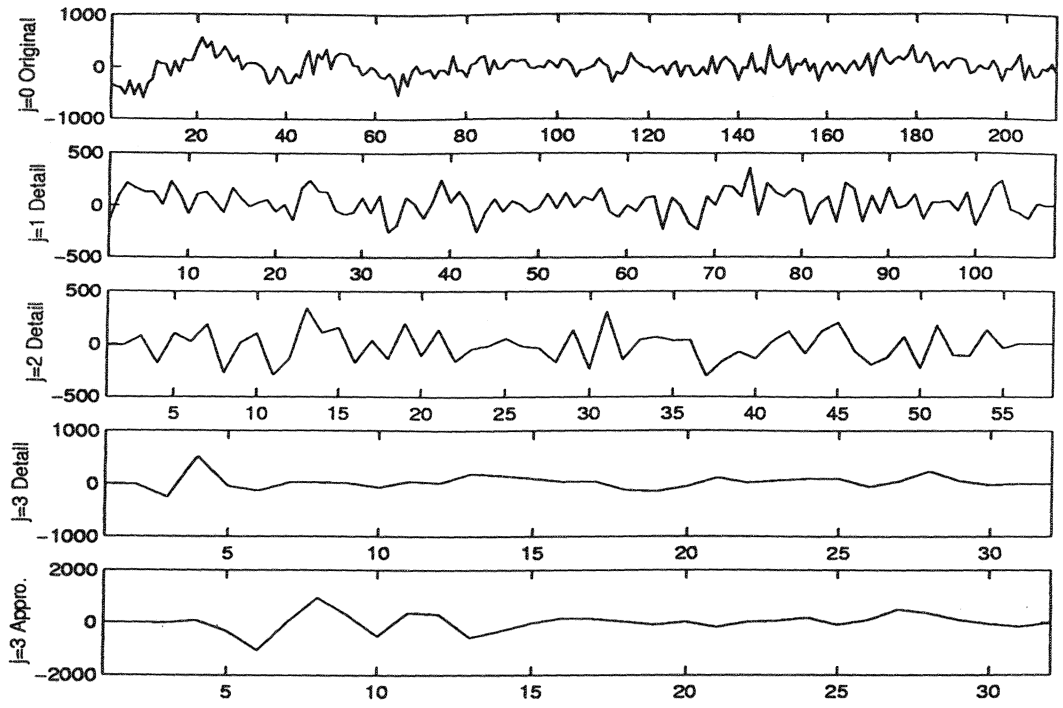


Figure 3.7: 1–D wavelet transform of a signal in $J = 3$ levels.

unprocessed seismic signal transformed in $J = 3$ levels. (*db*4 wavelet is used for the transformation). The original signal is shown in the first row. Second row shows the approximation coefficients. The third, forth and last rows show the three level detail coefficients of the signal.

The algorithm to compute the 1–D wavelet transform is also applicable to 2–D. The wavelet and scaling functions for 2–D can be obtained from one dimensional wavelets by tensorial product. This kind of 2–D WT leads to a decomposition of approximation coefficients at level j in four components: the approximation at level $j+1$ and the details in three orientations (horizontal, vertical, and diagonal). Figure 3.8 shows an image (a seismic section) and Figure 3.9 shows the decomposition of the image in $J = 2$ levels.

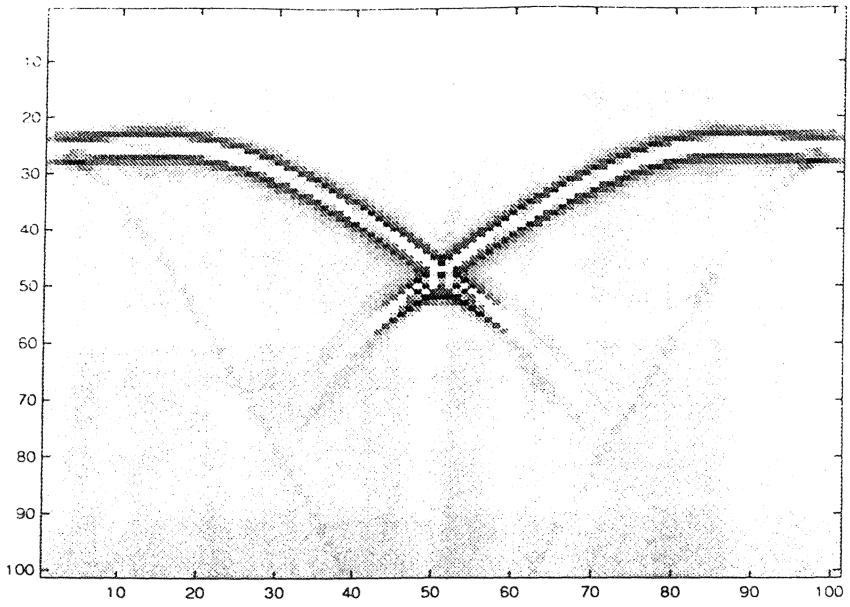


Figure 3.8: Synthetic data

The WT decomposition enables the wavelet analysis of a signal to be carried out. The wavelet analysis is capable of revealing aspects of the data that other signal analysis techniques miss, aspects like trends, breakdown points, discontinuities in higher derivatives, and self-similarity (Strang and Nguyen, 1997). Further, because it affords a different view of data from those presented by traditional techniques, wavelet analysis can often compress or de-noise a signal without appreciable degradation.

From the mathematical view, we can explain the ability of the WT of suppressing part of the signal as follows. Let w be a wavelet with at least $k+1$ vanishing moments (for $j = 0, 1, \ldots, k$, $\int_R x^j w(x)dx = 0$). If a signal f is a polynomial of degree k then the coefficients $C(a, b) = 0$ for all a and b (a, b correspond to scale and time level respectively). Such a wavelet automatically suppresses the polynomials. The degree of f can vary with time, provided that it remain less than k. If f is

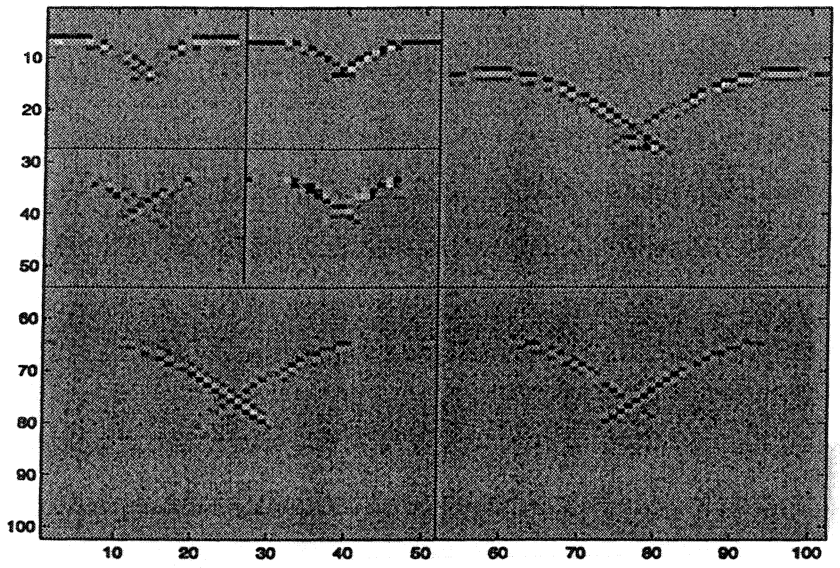


Figure 3.9: 2–D decomposition of the synthetic data in $J = 2$ levels. The top–right, bottom–left and bottom–right corners of the image are the details (horizontal, vertical, diagonal) of wavelet decomposition on $j = 1$. The approximation at level $j = 1$ at the top–left corner is further decomposed in four components at $j = 2$: approximation (top–left of the corner) and three details (top–right and bottom of the corner).

now a polynomial of degree k on the segment $[\alpha, \beta]$, then $C(a, b) = 0$ as long as the support of the function $\frac{1}{\sqrt{a}} w \frac{x-b}{a}$ is included in $[\alpha, \beta]$. In this case, the suppression is local. Likewise, suppose that on $[\alpha, \beta]$ to which 0 belongs, we have the expansion

$$f(x) = f(0) + xf'(0) + \frac{x^2 f''(0)}{2} + \ldots + \frac{x^k f^{(k)}(0)}{k!} + g(x).$$

The f and g signals then have the same wavelet coefficients. The signal g is the irregular part of the signal s. The wavelet transform systematically suppresses the regular part and analyzes the irregular part. Another way of suppressing a component of the signal consists of forcing certain coefficients $C(a, b)$ to be equal to 0. This is done by having selected a set E of indices (a threshold value), stipulate the $(a, b) \in E$, $C(a, b) = 0$, then synthesize the signal using the modified coefficients.

The compression features of a given wavelet basis are primarily linked to the relative sparseness of the wavelet domain representation for the signal. The notion behind compression is based on the concept that the regular signal component can be accurately approximated using the following elements: a small number of approximation coefficients (at a suitably chosen level) and some of the detail coefficients. The compression procedure contains three steps: decompose, threshold the detail coefficients and reconstruction. Figure 3.10 (Top) shows a seismic signal. If we apply the WT to the signal and threshold the approximation and the detail coefficients, the remaining coefficients can still well represent the original signal. Figure 3.10 (Bottom) shows the reconstructed signal using the remained coefficients. The reconstructed signal preserves most of the energy of the original signal.

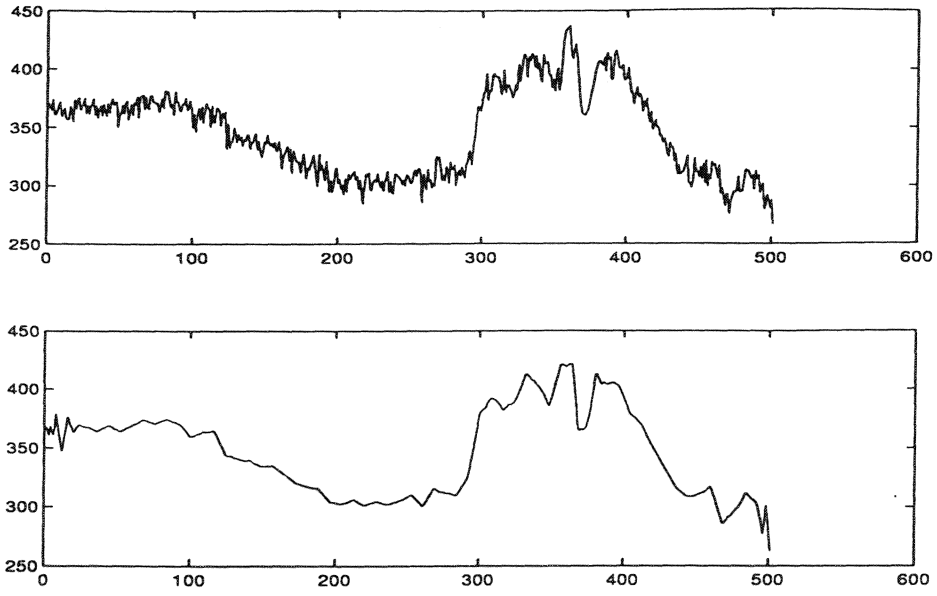


Figure 3.10: Wavelet compressing of a signal. Top: A original signal, Bottom: the reconstructed signal using the thresholded coefficients.

3.3 Migration on WT filtered data

Migration using input based technique can be highly optimized by compressing the input data. A new approach which implements this idea is discussed here. The new approach uses the wavelet transform as the filter to discriminate in the input section the dipping events, where the diffraction summation needs to be performed. The computational aspects of migration can then be improved by using those WT filtered data as the input for the migration.

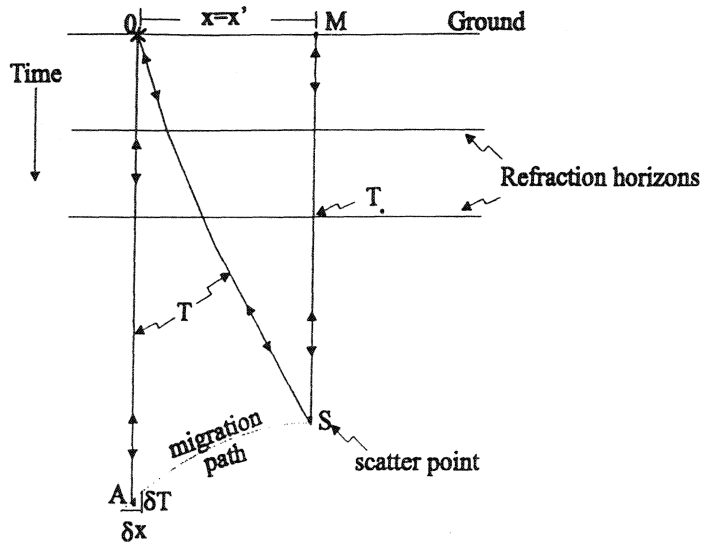


Figure 3.11: Schematic 2–D seismic time section diagram containing both a hypothetical source and receiver at ground position 0, a point scatterer S in the subsurface, an apparent scatterer location A recorded at CDP point O on the time section, and the CDP point M where the migrated recording of the scatter will be. (Adapted from Robinson and Robbins, 1978.)

3.3.1 Introduction

Seismic migration is a process of moving (or migrating) subsurface events on a seismic time section from their initial, erroneous common–depth point (CDP) positions to their true positions. Any deviation from a transversely homogeneous acoustical impedance structure in the subsurface produces seismic data which need to be migrated. We are basically only concerned with acoustical impedance translations or interfaces that are curved or dipping, although we make the usual assumption of transverse homogeneity from those interfaces up to the earth's surface. This point can be approved from the well–established normal–moveout relation (Robinson and Robbins, 1978). Assuming plane, horizontal refraction interfaces, one can easily arrive at the following relationship regarding Figure 3.11:

$$T = \left[T_0^2 + \frac{(2x)^2}{v^2}\right]^{1/2}, \tag{3.3.3.36}$$

where v represent velocity along a vertical path from the ground (position M) to the scatterer S. The apparent slope at image position A on the recorded seismic time section, due to the point scatterer S, can be expressed from a differentiation of equation (3.3.3.36):

$$\delta T/\delta x \simeq \frac{4x}{Tv^2}. \tag{3.3.3.37}$$

The migration correction consists of moving the recorded image point $A(0,T)$ to the point $S(x',T_0)$, where

$$x' = (1/4)Tv^2(\delta T/\delta x), \tag{3.3.3.38}$$

and

$$T_0 = \left[T^2 - \frac{(2x')^2}{v^2}\right]^{1/2}. \tag{3.3.3.39}$$

For a horizontal event at image position A, $\delta T/\delta = 0$. Then in equation (3.3.3.42), $x' = 0$. Substituting x' into equation (3.3.3.41), we get $T_0 = T$. Therefore, the image point A remains unmoved after migration. This implies that migration on the whole seismic data 'D' can be approximated by migration on dipping events 'D_d'.

We propose a method that uses the wavelet transform to isolate dipping events. Then migration is done by performing Kirchhoff summation on the dipping events.

3.3.2 Method and algorithm

In the Kirchhoff summation approach to migration, the output image is built according to:

$$I(x,z) = \int_{r,s} w(r,s,x,z) \times D(r,s,T) ds dr. \qquad (3.3.3.40$$

Where

$$T = T(x,z,r) + T(x,z,s).$$

T is sum of the travel time from the image point to the receiver and to the source, D is the input seismic data, w is the weighting factor (see Section 2.2.1). In the case of a post–stack time migration, the input data depend only on the midpoint coordinate $y = \frac{r+s}{2}$, and that equation (3.3.3.40) can be rewritten as

$$\hat{I}(y,T) = \int_y w(y,x,T) \times D(y,T) dy, \qquad (3.3.3.41)$$

or

$$\hat{I}(y,T) = \int_y w(y,x,T) \times D(y,T) dy + \int_y w(y,x,T) \times D(y,T) dr. \qquad (3.3.3.42)$$

Where $T = T(x,y)$, $D_d(y,T)$ and $D_h(y,T)$ are dipping and horizontal energy on input data, respectively. It was mentioned that horizontal energy on input data is unmoved after the migration. Therefore we may approximate $\int_y w(y,x,T) D(y,T) dy$ as the $\alpha D_h(y,T)$ and rewrite equation (3.3.3.42) as

$$\hat{I}(y,T) \approx \alpha \times D_h(y,T) + \int_y w(y,x,T) \times D_d(y,s,T) dr. \qquad (3.3.3.43)$$

Here α is a constant which is used for compensating the loss of the energy due t the fact that we do not perform a summation along hyperbola for the horizont events. We use the wavelet transform to filter the seismic section into two part one part contains the horizontal events and the events with small energy which ma be ignored in migration; another part is where the migration needs to be performe (dipping events). In our method, the 1-D WT is applied to each function $f(x)$ i the seismic data set $D(t,x)$ (along the spatial (x) direction). Therefore the dat space is mapped into the multiresolution approximation

$$\hat{D}_{t,x} = W D_{t,x}, \tag{3.3.3.44}$$

where W is a compactly supported, orthonormal *db4* wavelet (Daubechies, 1988). Assuming $f(x)$ is a signal in $D_{t,x}$, the horizontal events and the events with small energy are presented on $f(t)$ as certain continuous energy. The dipping events cause a rapid change of the energy of $f(x)$. This change is enhanced by the wavelet transform. Therefore the wavelet coefficients tend to be big for locations where dipping events exist ($\delta T/\delta x \neq 0$). If we discard all the approximation coefficients and assign a threshold value for the detail coefficients, a truncation version of $\hat{D}_{t,x}$, $\hat{D}'_{t,x}$, can be obtained. The inverse WT of $\hat{D}'_{t,x}$ is

$$\hat{D}''_{t,x} = W^T \hat{D}'_{t,x}, \tag{3.3.3.45}$$

where $\hat{D}''_{t,x}$ is considered to be the energy of dipping events (D_d). $D - \hat{D}''$ is considered to be the energy containing flat events (D_h). Therefore the migration according to equation (3.3.3.43) can be carried out.

If there are a total of $nx * nt$ points in the original data, the number of operation for the standard Kirchhoff time migration method is proportional to $nx *$

$nt*na$ (na is aperture width). By assuming that after WT filtering the points whe migration needs to be done is $\frac{1}{P}$ of the original points, the operation count of th new method is proportional to $nx * nt * na * \frac{1}{P}$.

3.3.3 Synthetic examples

In Figure 3.12 (Top Left), we illustrate a synthetic zero–offset seismic section. Figur 3.12 (Top Right) is the migrated image using the standard Kirchhoff migratio algorithm. All the points in the original data set (12100) are used in the hyperbol summation. Figure 3.12 (Bottom Left) is a migrated image using the new approach. The threshold is set to be 0.005, and 8296 points are used for summation. The migration is very similar to the standard migration using all the points. Figure 3.12 (Bottom Right) is also a migrated image using the new approach. In this case, the threshold is set to $thr = 1$, and 552 points are used for the migration. The migration result is poor since there are unsufficient remaining wavelet coefficients locate the dipping energy. We also examined the new approach by setting different threshold values. It was realized that the different threshold values affect the quality of the migrated image. Figure 3.13 (Top Left) is the migrated image by setting the threshold to 0.01. Figure 3.13 (Top Right) is the recovered detailed image which is used to isolate the dipping events for Figure 3.13 (Top Left). In this example, 6622 points are used for the summation. In Figure 3.13 (Bottom Left), the image is obtained by using $thr = 0.18$. Figure 3.13 (Bottom Right) is the recovered detailed image for the Figure 3.13 (Bottom Left). In this case, 2006 points are used for the summation. Figure 3.14 (Top Left) and (Top Right) are the migrated image and recovered detail coefficients, respectively, with $thr = 0.25$ in this example, and 1880 points are used for the summation. In the examples showed in Figure 3.14

(Bottom Left) and (Bottom Right), the threshold is 0.4 and 1127 points are used in the summation. In Figure 3.15 (Top) we used $thr = 0.5$, and 942 points were used in the summation. In Figure 3.15 (Bottom) we used $thr = 3$, and 76 points were used for the summation. By comparing those examples, we found that 2100 points, 1/6 of the total number of points (12100) in the original data are sufficient to properly image the zero–offset section (Figure 3.13 (Bottom)).

3.3.4 Conclusion

In this section, after reviewing the theory of the wavelet transform, a new approach for migration has been developed. This approach performs migration on WT compressed (filtered) seismic data. The computational cost of migrating data is reduced. The new approach is an output based Kirchhoff migration. This is suitable for post–stack time migration.

The wavelet transform filtered image provides a very sparse representation of our data. This representation enables us to minimize the number of points where summation along hyperbolic paths is required.

We tested the approach on synthetic zero–offset data. It is found that the quality of the migrated image is related to the threshold value used to filter the coefficients. By choosing a suitable threshold value, the migration can be efficiently carried out without degrading the final image.

It is important to stress that this technique does not attempt to replace existing post–stack migration algorithms, which are very efficient. This technique explores the plausibility of using WT filtered data to migrate seismic events.

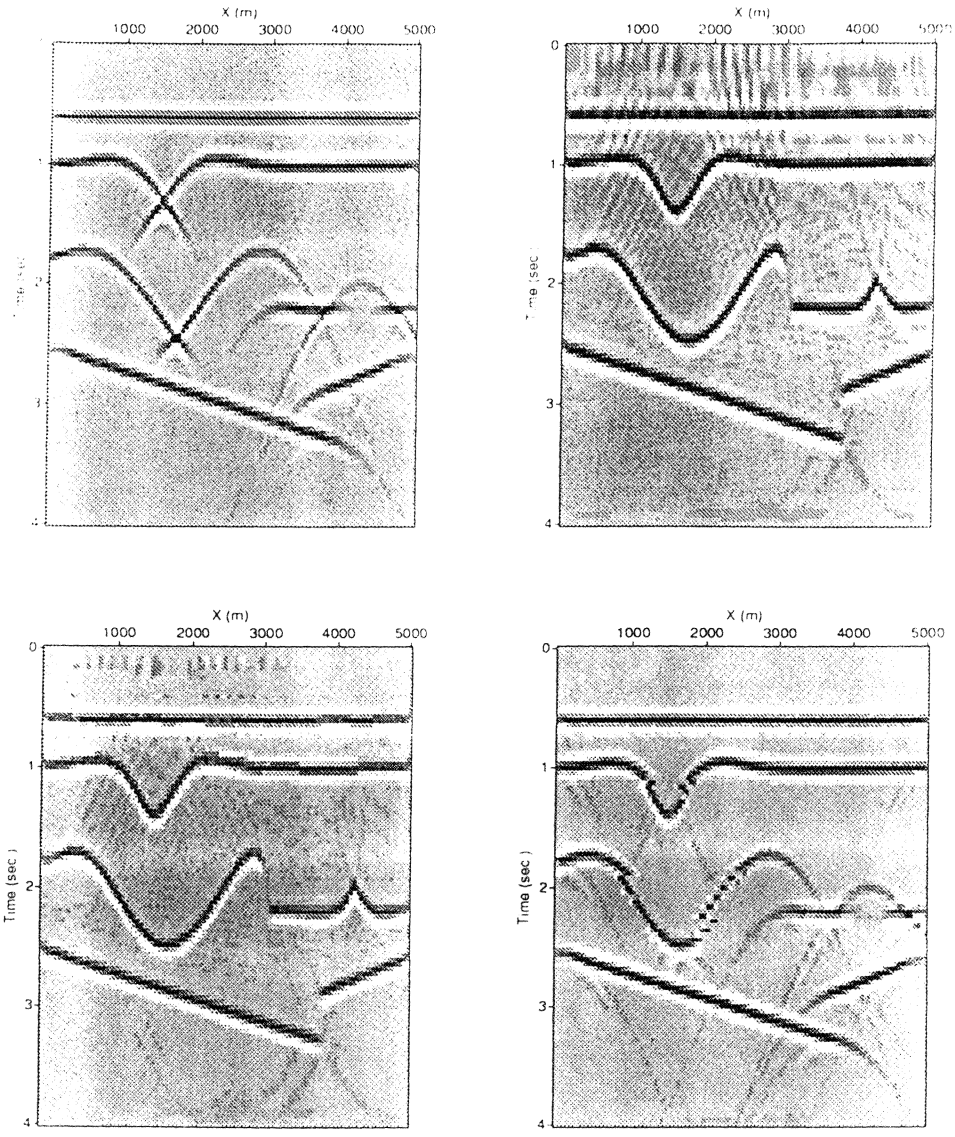


Figure 3.12: Top Left: Unmigrated seismic section. Top Right: The migrated image using ordinary Kirchhoff migration. Bottom Left: Migrated seismic section using $thr = 0.005$. Bottom Right: Migrated seismic section using $thr = 1$.

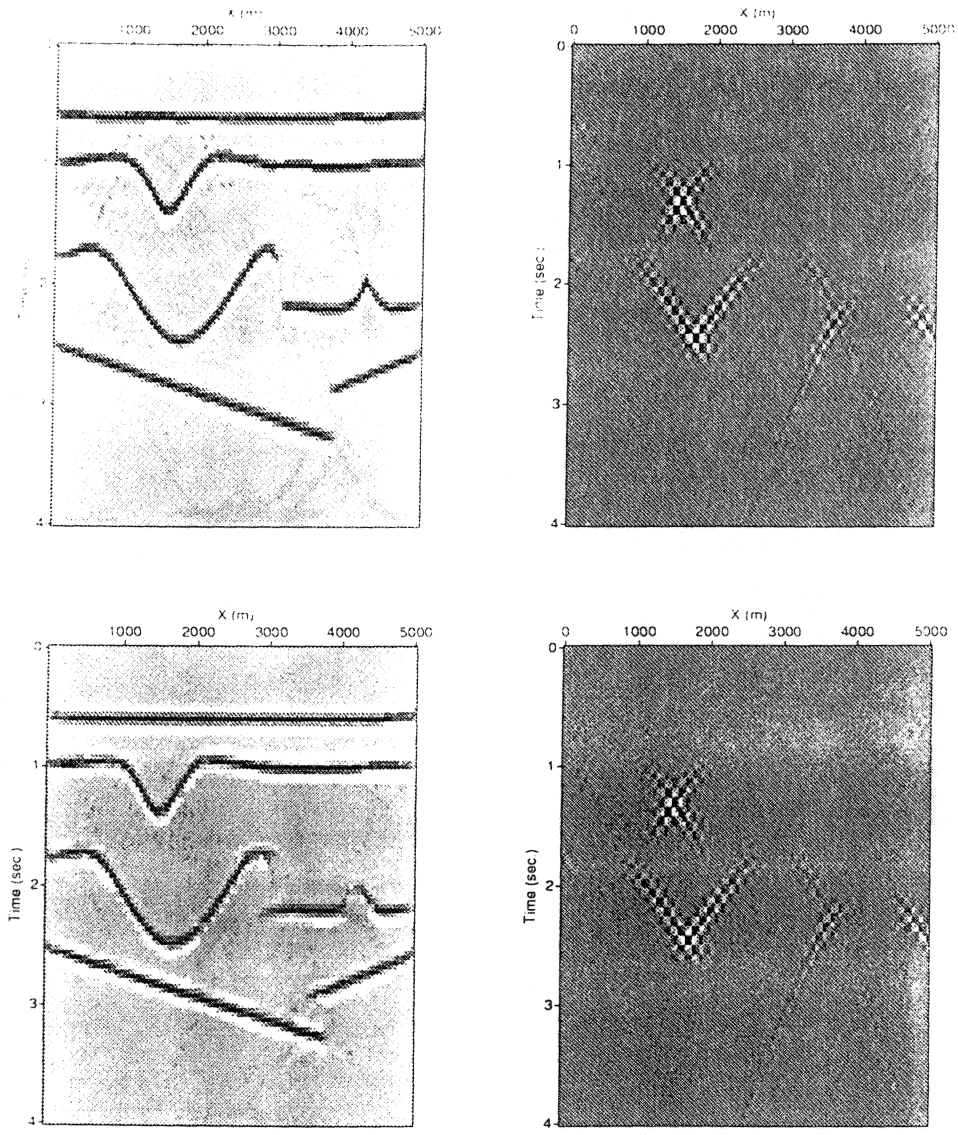


Figure 3.13: Top Left: Migrated seismic section using $thr = 0.01$, Top Right: the remained detail coefficients by setting $thr = 0.01$. Bottom Left: Migrated seismic section using $thr = 0.18$, Bottom Right: the remaining detail coefficients obtained by setting $thr = 0.18$.

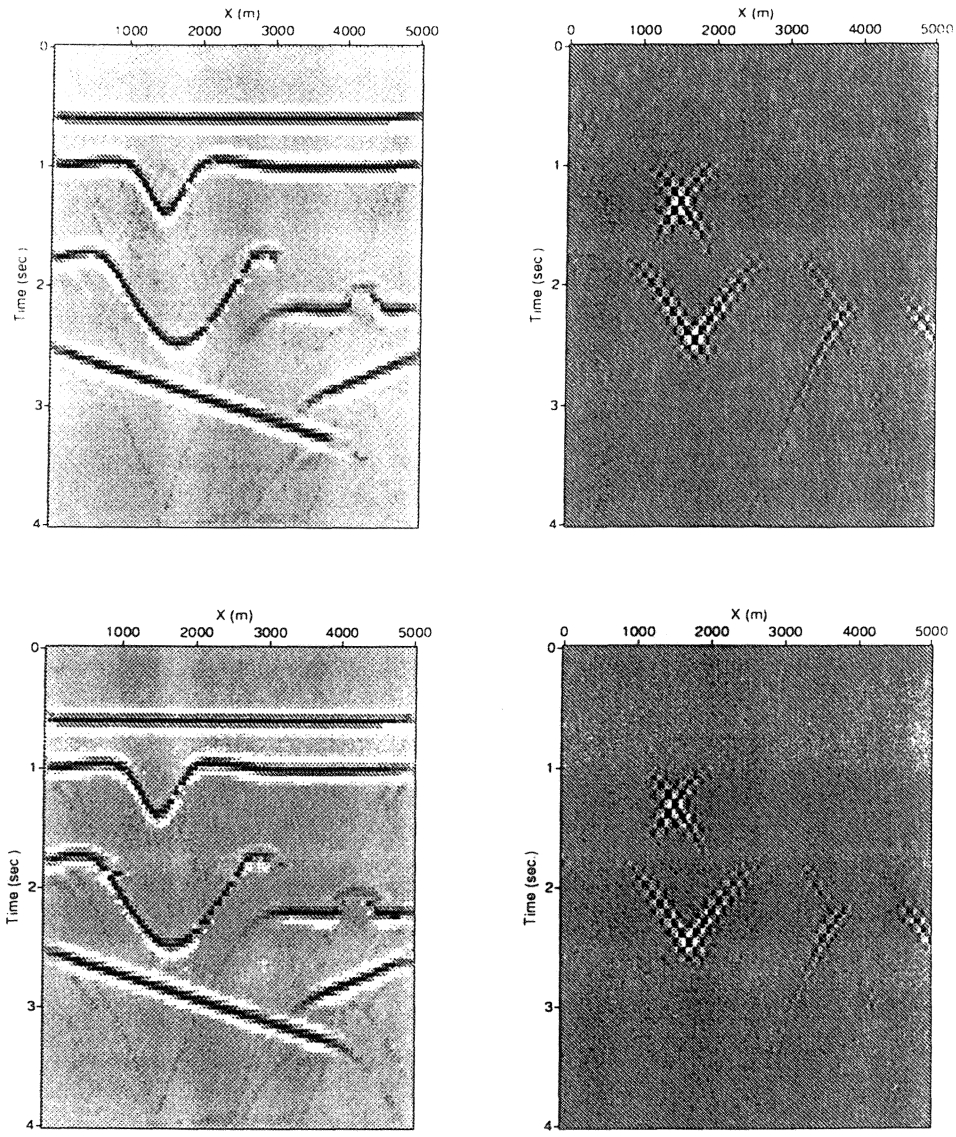


Figure 3.14: Top Left: Migrated seismic section using $thr = 0.25$, Top Right: the remained detail coefficients by setting $thr = 0.25$. Bottom Left: Migrated seismic section using $thr = 0.4$, Bottom Right: the remaining detail coefficients obtained by setting $thr = 0.4$.

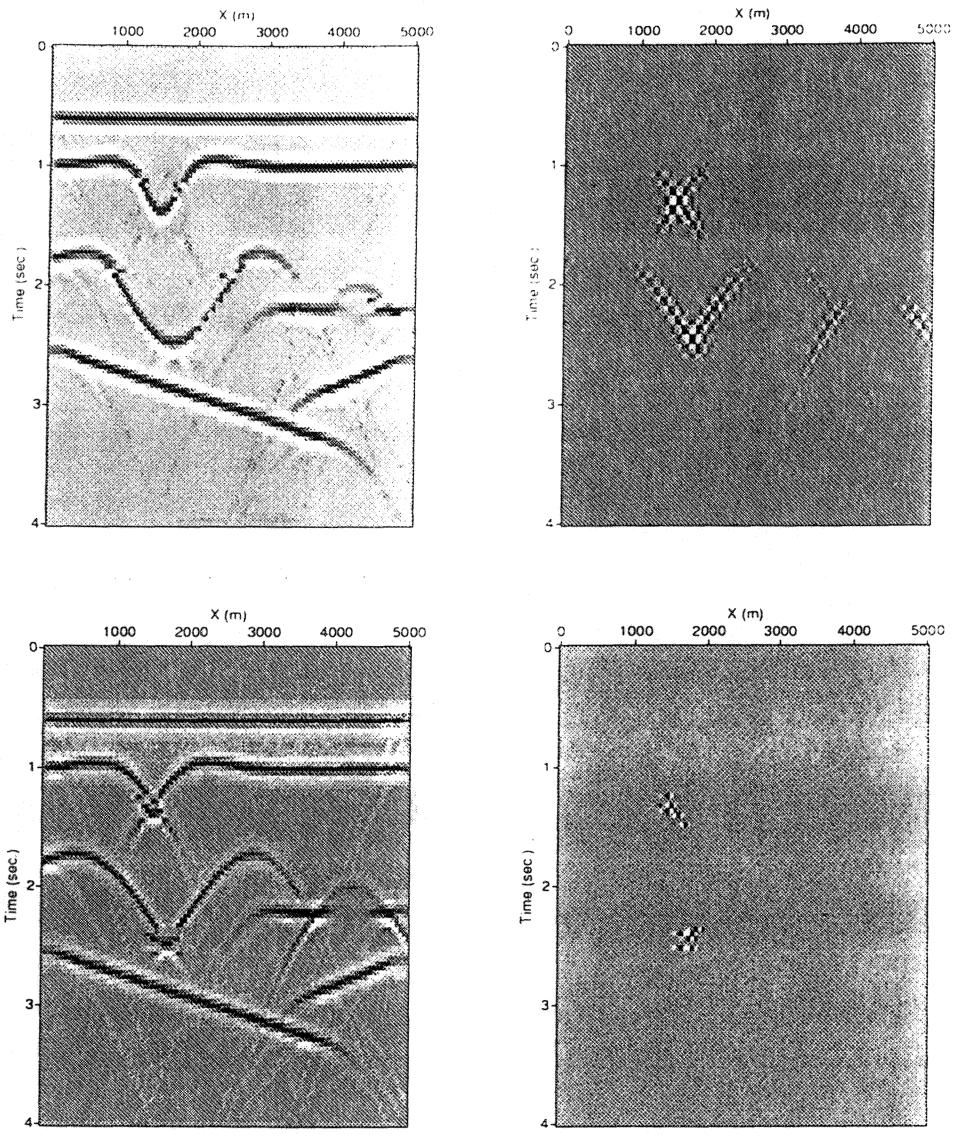


Figure 3.15: Top Left: Migrated seismic section using $thr = 0.5$, Top Right: the remained detail coefficients by setting $thr = 0.5$. Bottom Left: Migrated seismic section using $thr = 3$, Bottom Right: the remaining detail coefficients obtained by setting $thr = 3$.

CHAPTER 4

Matching Pursuit Migration and Inversion of Travel Time Tables

4.1 Introduction

In the previous chapter, we presented a technique to decompose a signal in terms of temporal and scale attributes. The wavelet transform provides a way to obtain a sparse representation of our data. In this chapter, we introduce a less elegant algorithm to decompose a signal in a set of waveforms or atoms which also seeks a sparse representation of the data. This algorithm, called matching pursuit (MP), was developed by Mallat and Zhang (1993) and applied in various scientific scenarios. In image processing, it was used by Neff et al. (1998), and Al–Shaykh et al. (1999) to compress video images. In seismology, the MP algorithm has been applied by Chakraborty and Okaya (1995) to visualize the non–stationary data behavior due to attenuation.

In seismic exploration, Wang and Pann (1996) have first suggested a MP algorithm to migrate seismic data. In their approach, the seismic traces are decomposed as a superposition of waveforms that are obtained via the MP algorithm. This approach focused on applications to post–stack time migration, where the migration operator has an analytical expression. This enables them to use an input migration technique. Li et al. (1998) introduce a new algorithm where the matching atom is

estimated from the data. They have also tested the algorithm in problems analytical expressions exist for the traveltime functions needed to migrate the data.

In this thesis, a novel MP algorithm to process pre–stack data is studied. In particular, we combine MP with the inversion of traveltime tables to perform seismic migration on the sparse representation of the seismic data.

4.2 The matching pursuit Algorithm

4.2.1 Introduction

The term *matching pursuits* refers to a *greedy algorithm*[1] which matches signal structures to a large, diverse dictionary of functions. The technique was proposed by Mallat and Zhang (1993) with an application to signal analysis. They give the definition of matching pursuit as an algorithm that decomposes any signal into a linear expansion of waveforms that are selected from a redundant dictionary of functions. These waveforms are chosen in order to best match the signal structure. Using the terminology introduced by Mallat and Zhang (1993), a dictionary is a collection of parameterized waveforms $D = (\phi_\gamma : \gamma \in \Gamma)$. The waveforms g_γ are discrete–time signals of length n called atoms. Depending on the dictionary, the parameter γ can have the interpretation of indexing frequency, in which case the dictionary is a frequency dictionary, or of indexing time/frequency jointly, in which case the dictionary is a time–frequency dictionary (Chen and Donoho, 1996). The dictionaries are complete in the case where they contain exactly n atoms, and is over complete in case where they contain more than n atoms. The wide scope of patterns

[1] As pointed out by Strang (1986), "greedy algorithm: Do the best thing at every step. In other words, ignore all difficulties that might come later and make the optimal choice now."

embedded in complex signals motivates decompositions over large and redundant dictionaries of waveforms. A complex signal may not be very well linearly expanded on a signal basis. For example, a signal f can be expanded on a Fourier basis which is a single basis via an inverse Fourier transform,

$$f(t) = \frac{1}{2\pi} \int_{+\infty}^{-\infty} F(\omega) e^{-i\omega t} d\omega. \tag{4.4.2.1}$$

It is clear that a Fourier basis provides a poor representation of a time–localized signal. Another example is the wavelet bases (Daubechies, 1990). As mentioned before, the decomposition of a signal via wavelet bases is a time–frequency decomposition:

$$\begin{aligned} (Wf)(a,b) &= < f, \psi^{a,b} > \\ &= \int f(t) a^{-1/2} \psi \overline{\frac{t-b}{a}} dt, \end{aligned} \tag{4.4.2.2}$$

where

$$\psi^{a,b}(t) = a^{-1/2} \psi(\frac{t-b}{a}),$$

with a indicates frequency localization and b indicates time localization. The wavelet transform decomposes a signal over time–frequency atoms of varying scales, and the scale factor is inversely proportionally related to the frequency parameter. The wavelet bases are not well adapted to represent functions whose Fourier transforms have a narrow high frequency support. In particular, signals that have important variances in their localized frequency attributes are a problem for wavelets. In this case, a more flexible representation is necessary.

4.2.2 The algorithm

A matching pursuit is a greedy algorithm (Mallat and Zhang, 1993) that chooses at each iteration a waveform that is best adapted to the signal. Matching pursuit starts from an "empty model" and builds up a signal model an atom at a time, at each step adding to the model only the most important new atom among all those not so far in the model. Mallat and Zhang (1993) gave a mathematical description of the matching pursuit procedure. In order to match the component of a signal, matching pursuit is done by successive approximations of signal f with orthogonal projections on waveforms $g_{\gamma 0}$ from a dictionary D. The signal f can be decomposed into

$$f =< f, g_{\gamma 0} > g_{\gamma 0} + R^0 f, \tag{4.4.2.3}$$

where $R^0 f$ is the residual of the signal after approximating f with $g_{\gamma 0}$. Since $g_{\gamma 0}$ is orthogonal to $R^0 f$, we can write

$$\| f \|^2 = |< f, g_{\gamma 0} >|^2 + \| Rf \|^2 . \tag{4.4.2.4}$$

To minimize $\| Rf \|$, $g_{\gamma 0} \in D$ is chosen such that $|< f, g_{\gamma 0} >|$ is maximized. The waveform $g_{\gamma 0}$ is said to be the best if

$$|< f, g_{\gamma 0} >| \geq \alpha \sup_{\gamma \in \Gamma} |< f, g_{\gamma} >|, \tag{4.4.2.5}$$

where α is an optimality factor that satisfies $0 < \alpha < 1$. A matching pursuit is an iterative algorithm that sub–decomposes the residue Rf by projecting it on a waveform $g_{\gamma 1} \in D$ that matches Rf almost at best. This procedure is repeated

while trying to minimize the energy of the residue $R^n f$ by selecting a best waveform $g_{\gamma(n+1)}$ each time and the following residue $R^{n+1} f$ is obtained

$$R^n f =< R^n f, g_{\gamma n} > g_{\gamma n} + R^{n+1} f, \tag{4.4.2.6}$$

which defines the residue at the order of $n+1$. Since the R^{n+1} is orthogonal to $g_{\gamma n}$

$$\| R^n f \|^2 = |< R^n f, g_{\gamma} n >|^2 + \| R^{n+1} f \|^2, \tag{4.4.2.7}$$

when the procedure is iterated up to the m^{th} order, the signal f is decomposed into the following concatenated sum

$$f = \sum_{n=0}^{m-1} (R^n f - R^{n+1} f) + R^m f. \tag{4.4.2.8}$$

Equation (4.4.2.6) yields

$$f = \sum_{n=0}^{m-1} < R^n f, g_{\gamma n} > g_{\gamma n} + R^m f, \tag{4.4.2.9}$$

and $\| f \|^2$ is decomposed in a concatenated sum

$$\| f \|^2 = \sum_{n=0}^{m-1} (\| R^n f \|^2 - \| R^{n+1} f \|^2) + \| R^m f \|^2 . \tag{4.4.2.10}$$

We can see that the energy is conserved, this can be proved from equation (4.4.2.7):

$$\| f \|^2 = \sum_{n=0}^{m-1} |< R^n f, g_{\gamma n} >|^2 + \| R^m f \|^2 . \tag{4.4.2.11}$$

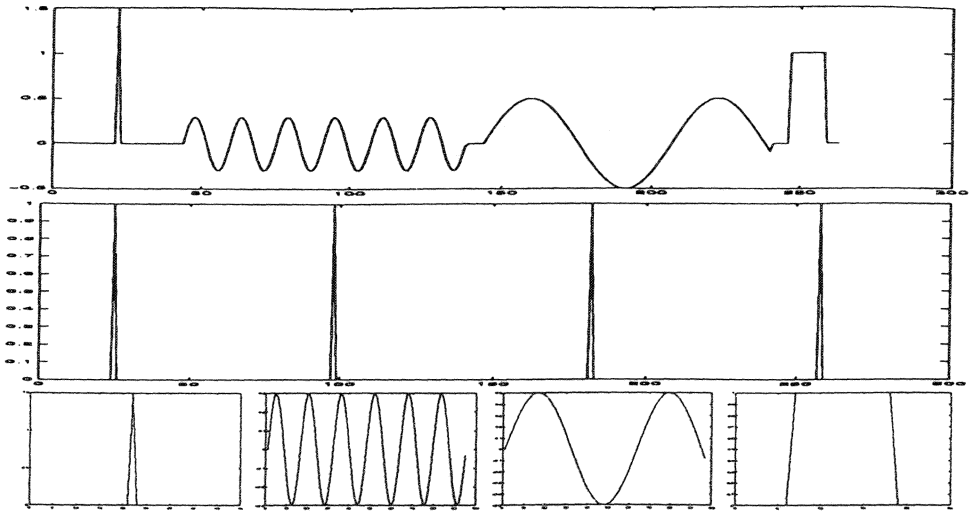


Figure 4.1: MP decompositions of a signal. First row shows a synthetic signal embedded with structures of spike, cosines and rectangular box. Second row shows the MP coefficients of the signal. The third row shows the atoms from a dictionary consisting of spikes, cosines of different angular frequencies, rectangular boxes. Those atoms are used for the MP projection and can well represent the signal.

Figure 4.1 shows a MP decomposition of a signal with atoms selected from a dictionary that consists of spikes, cosines of different angular frequencies and rectangular boxes.

4.2.3 Estimating the atoms using the maximum entropy method

In this thesis, the atoms (waveforms) used for MP are estimated using the maximum entropy method (MEM). Bear in mind, that in most MP applications the atoms are given and not estimated. The MEM is used to compute the power spectrum of the

wavelet. The phase is assumed to be zero[2]. This wavelet is considered to be th waveform that best matches the structure of the seismic trace. We have to stres that other atoms are plausible, but a zero phase function that resembles the seismi pulse will lead us to a very efficient representation.

The Maximum Entropy method

The MEM has been developed by Burg (1967) and also been suggested by Parzen (1969), who refers to it as an auto–regressive spectral estimator. Before the introduction of MEM, all of the usual methods of spectral analyses have been associated with window functions which are independent of the data or the properties of the random process analyzed. One difficulty with conventional methods is that the window function does not depend upon the true spectrum. Therefore misleading or false conclusions can sometimes be drawn when the estimated spectrum is used. The MEM method, unlike the conventional methods for power spectral density estimation, adapts to the actual characteristics of the signal under analysis (Lacoss, 1971). The maximum entropy method attempts to fit, in a least–squares sense, an autoregressive (AR) model to an input time series. Assuming that the data x_t (the seismic trace) are generated by the process

$$x_t = \varepsilon_t - a_1 x_{t-1} - a_2 x_{t-2} - \cdots - a_m x_{t-m}, \qquad (4.4.2.12)$$

where ε_t is uncorrelated random noise with zero mean and variance σ^2. Let $a_0 = 1$, equation (4.4.2.12) can be rewritten

$$x_t + a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_m x_{t-m} = \varepsilon_t. \qquad (4.4.2.13)$$

[2]We assume that after spiking deconvolution (Yilmaz, 1987) the resulting seismic wavelet is a zero phase time series.

The sequence a_t with $a_0 = 1$ is actually a prediction error filter. Those coefficients $a_0, a_1, \cdots, a_m$ are also called AR parameters, which the MEM approach attempts to estimate. It has been pointed out by Ulrych and Bishop (1975) that the AR representation is equivalent to that time series which is consistent with known autocorrelation measurements, but which has maximum entropy. From Shannon's theory of information, the entropy of a giving time series is proportional to the integral of the logarithm of its power spectrum. It seeks to maximize the integral

$$\int_{-\omega_n}^{+\omega_n} \log \Phi(\omega) d\omega, \tag{4.4.2.14}$$

with the constraint that

$$\int_{-\omega_n}^{+\omega_n} \Phi(\omega) e^{i\omega k \Delta t} d\omega = \phi_k. \tag{4.4.2.15}$$

Equation (4.4.2.15) is the Wiener–Khintchine theorem (Robinson, 1967) relating the power spectrum $\Phi(\omega)$ to the autocorrelation function ϕ_k. Here Δt is the time increment, k is a discrete time index, ω the angular frequency, and ω_N is folding frequency $\omega_N = \pi/\Delta t$. Based on these theorems, the MEM spectral estimate

$$\Phi(\omega) = \frac{\sigma_m^2}{A_m(\omega)}, \tag{4.4.2.16}$$

where A_m is $(m+1)$–length minimum delay prediction error (PE) filter, σ_m^2 is the error variance for the filter A_m. Therefore the maximum entropy representation of the observed data is actually an mth–order autoregressive (AR) process.

Burg (1967) developed a technique to estimate the PE filter A_m from the data rather than from the autocorrelation function. A two–term PE filter $(1, a)$ of the time series x_t is given by the choice of a which minimizes

$$E(a) = \sum_{t=1}^{N} \mid x_t + a x_{t-1} \mid^2 . \qquad (4.4.2.1'$$

To avoid $\mid a \mid$ being greater than unity, Burg uses another formula that minimizes th average of the sum of both the mean square forward and the mean square backwar prediction error. This is

$$E(a) = \sum_{t=1}^{N} \mid x_t + a x_{t-a} \mid^2 + \mid \overline{x}_{t-1} + a\overline{x}_t \mid^2, \qquad (4.4.2.18$$

which always leads to $\mid a \mid$ less than unity. The power spectral estimated associatec with this value of a is $R = 1/[(1 + \bar{a}/Z)(1 + aZ)]$.[3] Burg also noted that Levinsor recursion always gives minimum–phase filters. In the Levinson recursion a filter of order 3 is built up from an order of 2 by

$$\begin{bmatrix} 1 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 \\ a \\ 0 \end{bmatrix} - c \begin{bmatrix} 0 \\ a \\ 1 \end{bmatrix} .$$

Thus he takes a to be given from equation (4.4.2.18) and then does a least–squares problem to solve for c. This is done in such a way as to ensure that $\mid c \mid$ is less than unity, which guarantees that $A(z) = 1 + a_1 Z + a_2 Z^2$ is minimum–phase (Claerbout, 1976). Then equation (4.4.2.18) is rewritten as

$$E(a_1, a_2) = \sum_{t=2}^{N} \mid x_t + a_1 x_{t-1} + a_2 x_{t-2} \mid^2 + \mid \overline{x}_{t-2} + a_1 \overline{x}_{t-1} + a_2 \overline{x}_t \mid^2 \qquad (4.4.2.19)$$

or

$$\begin{aligned} E(c) \;=\; & \sum_{t=2}^{N} \mid x_t + a x_{t-1} - c(\bar{a} x_{t-1} + x_{t-2} \mid^2 \\ & + \mid \overline{x}_{t-2} + a\overline{x}_{t-1} - c(\overline{a\overline{x}}_{t-1} + \overline{x}_t) \mid^2 . \end{aligned} \qquad (4.4.2.20)$$

[3]Here, we use the Z transform, the Fourier transform can be obtained by replacing $Z = e^{j\omega}$.

Denote the error in forward prediction of $x_t - ax_{t-1}$ by e_+ and and error on backward prediction $x_{t-2} + \bar{a}x_{t-1}$ by e_-, equation (4.4.2.20) becomes

$$\begin{aligned} E &= \sum_t \mid e_+ - ce_- \mid^2 + \mid \bar{e}_- - c\bar{e}_+ \mid^2 \\ &= \sum_t \overline{(e_+ - ce_-)}(e_+ - ce_-) + \overline{(\bar{e}_- - c\bar{e}_+)}(\bar{e}_- - c\bar{e}_+). \end{aligned} \tag{4.4.2.21}$$

Setting the derivative with respect to $\bar{c}$ equal to zero

$$c = \frac{+\sum_t 2\bar{e}_- e_+}{\sum_t \bar{e}_+ e_+ + \bar{e}_- e_-} \tag{4.4.2.22}$$

Since the length of the vector $e_+ \pm e_-$ is always positive $|, c \mid$ is alway less than unity. If define e_+ and e_- as

$$\begin{aligned} e_+ &\leftarrow e_+ - ce_- \\ e_- &\leftarrow e_- - \bar{c}e_+ \end{aligned}$$

The forward and backward prediction errors of the three–term filter $(1, a_1', a_a') = (1, a_1 - c\bar{a}_1, -c)$. Return to equation (4.4.2.21) and proceed recursively, e_- and e_+ gradually become unpredictable, then a filter $A(Z)$ which filters $X(Z)$ with a output of white noise is found. Since the output has a constant spectrum, the spectrum of the input must be the inverse of the spectrum of the filter. Indeed, this effective mechanism of Burg spectral estimation is to compute a PE filter and look at the inverse of its spectrum. The most important property of a prediction–error filter is that its output tends to a white spectrum. No matter what the input to this filter, its output tends to whiteness as the number of the coefficients tends to infinity. Thus, the PE filter adapts itself to the input by absorbing all its color. The Burg's algorithm has a wave–propagation interpretation: in a layered medium the parameter c_k has the interpretation of reflection coefficients; the e_+ and e_- have the

interpretation of down–going waves; and the whole process of calculating c_k amounts to downward continuing surface seismograms into the earth, determining an earth model c_k as you go (Claerbout, 1976).

Examples

In our MP procedure, the MEM is used to compute the power spectrum of a seismic trace. The MEM filter is a minimum phase filter (Claerbout, 1976). However, as mentioned before, we prefer to use a zero phase wavelet. The MEM is a parametric spectral estimator which gives a smooth amplitude spectrum of the wavelet. The amplitude should be a smooth function because the wavelet is usually a short and smooth time series. Figure 4.2a shows an estimated wavelet (atom) using MEM from a real seismic trace (Figure 4.2b) with 1000 points. Figures 4.2c, e, and g show the resulting MP coefficients when 10, 20, and 50 atoms are used for the MP pursuit projection. Figures 4.2d, f, h are the reconstructed traces using 10, 20 and 50 atoms, respectively. This is done by convolving the MP coefficients with the wavelet (atom). The reconstructed traces are only an approximation of the original traces. We can see that when 10 atoms are used for MP projection, the reconstruction only retains the six strongest reflections of the original trace. When 50 atoms were used, the approximation was drastically improved. In Figure 4.3 (Top), we illustrate a real seismic section with 101 traces and 1000 time samples each. We performed the MP decomposition for each seismic trace using 100 atoms. The sparse representation is shown in Figure 4.3 (Bottom) which preserves most of the coherent energy of the original data (Figure 4.3 (Top)).

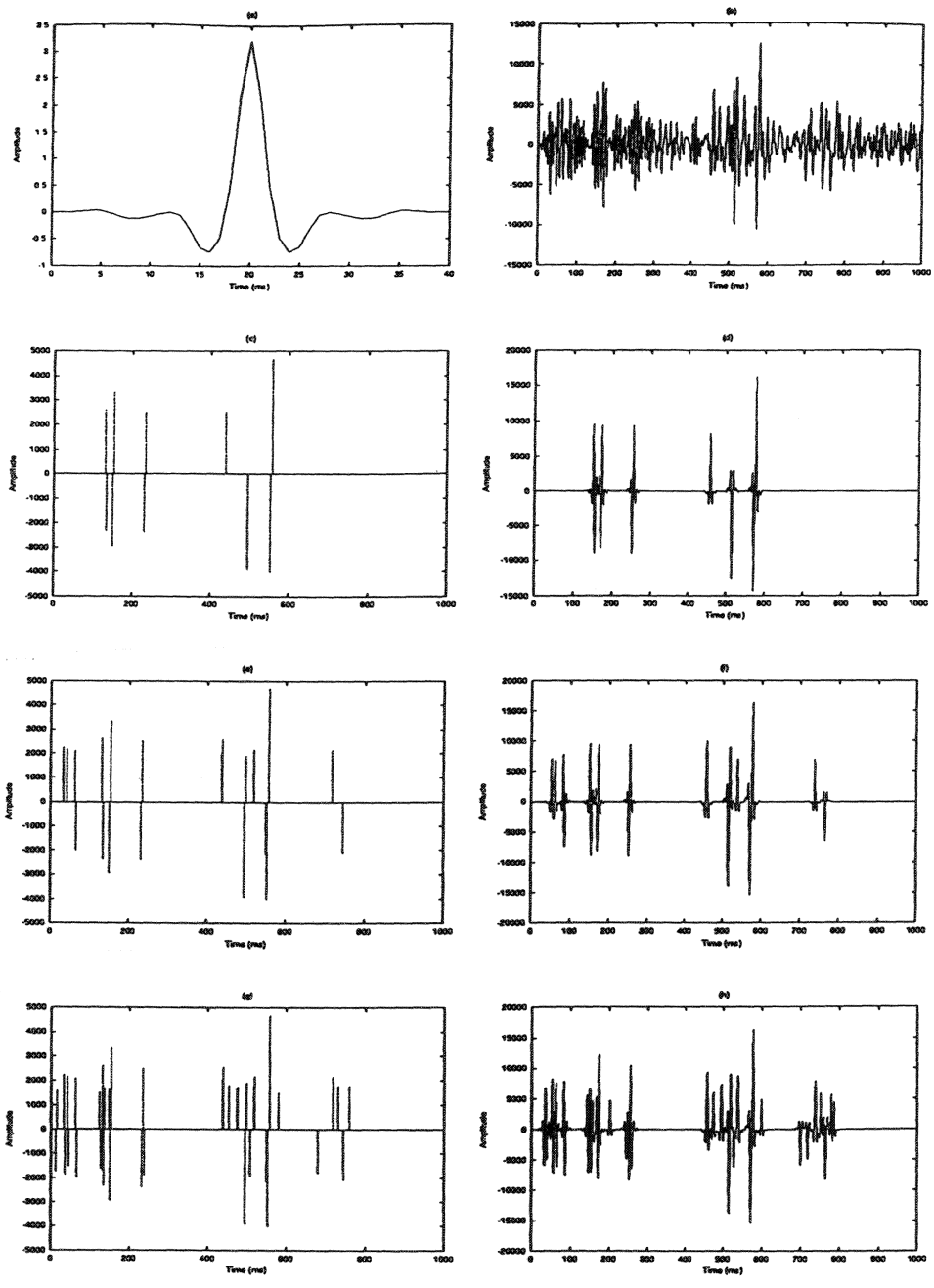


Figure 4.2: Matching pursuit of a seismic trace: (a) The estimated wavelet (atom) from a seismic trace using MEM. (b) The seismic trace. (c) and (d) MP coefficients and reconstructed trace using 10 atoms. (e) and (f) MP coefficients and reconstructed trace using 20 atoms. (g) and (h) MP coefficients and reconstructed trace using 50 atoms.

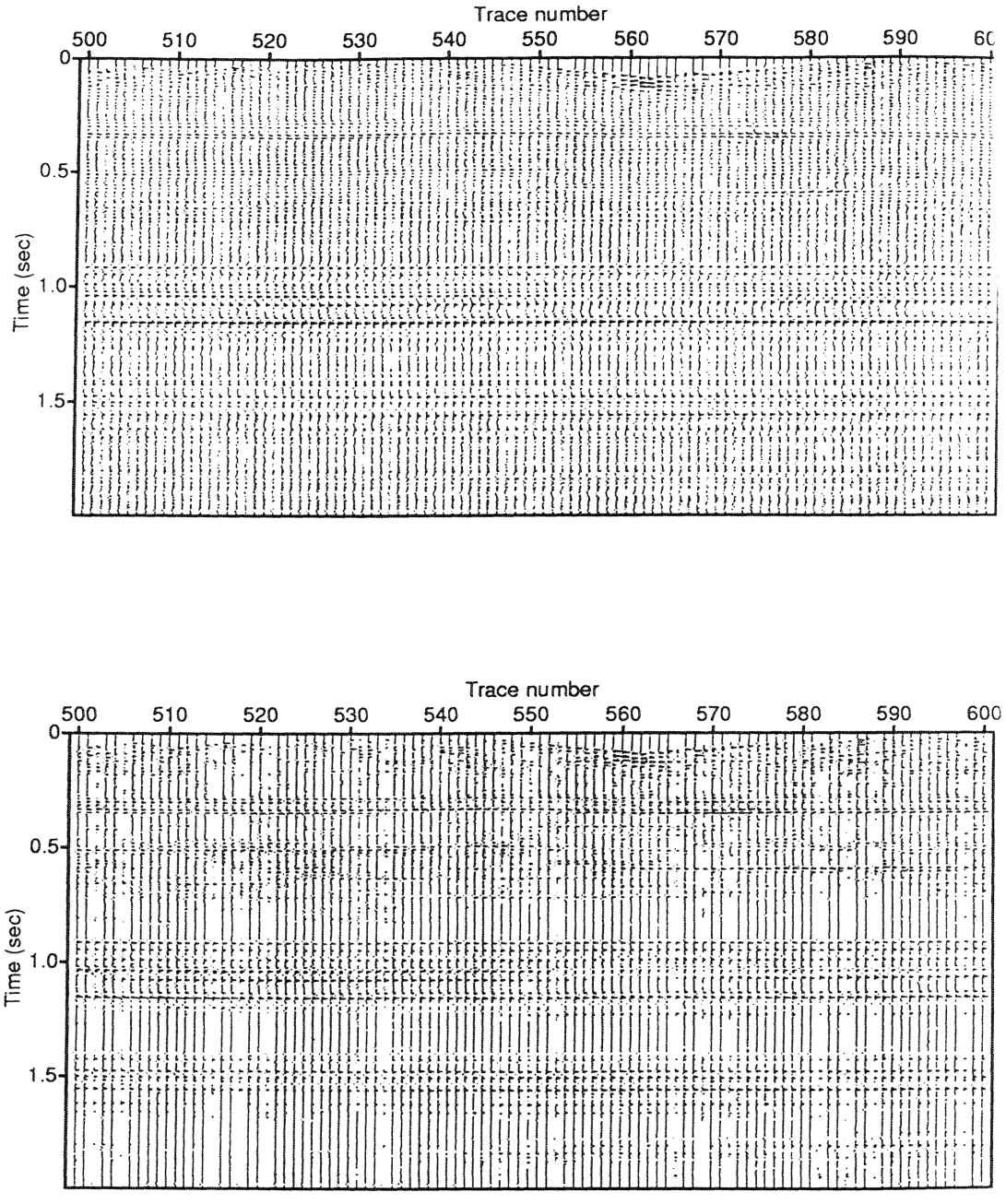


Figure 4.3: MP of a seismic section. Top: original seismic section. Bottom: after MP with 100 atoms.

4.3 Matching pursuit migration and inversion of travel time tables

It was shown that using a MP algorithm seismic data can be compressed into a sparse representation of the original data. Therefore, migration may be carried out efficiently by using MP compressed data. Wang and Pann (1996) and Li et al. (1998) have developed an approach where the MP algorithm is used for time migration. However, the situation is different for depth migration. In this section, we focus on an implementation of the matching pursuit algorithm to depth migration.

4.3.1 The MP migration algorithm

In the MP algorithm, the seismic traces are decomposed to a superposition of waveform or atoms. If we denote $D(r, s, t)$ the seismic trace obtained with a receiver r and a source s and apply the MP decomposition to the trace, we have

$$D =< D, g_{\gamma 0} > g_{\gamma 0} + R^0 D. \tag{4.4.3.23}$$

After m iterations of the procedure, the trace is decomposed into

$$D = \sum_{n=0}^{m-1} < R^n D, g_{\gamma n} > g_{\gamma n} + R^m D. \tag{4.4.3.24}$$

A seismic trace at a position (r, s) (receiver, shot) can be expanded in terms of wavelets using MP

$$\hat{D}(r, s, t) = \sum_{i=1}^{N(r,s)} A_i(r, s) w(t - t_i(r, s)). \tag{4.4.3.25}$$

$\hat{D}(r,s,t)$ is an approximation to the true seismic trace after matching. Where $N(r,s$ is number of atoms matched for the trace (r,s), $A_i(r,s)$ is the coefficient for the $i-t$ atom for trace (r,s), $t_i(r,s)$ is the time where we have matched the $i-th$ atom and $w(t)$ is the wavelet (atom). In our code, the wavelet is estimated from eac trace using the MEM. Therefore, the atom should be noted as $w(t,r,s)$. Back t the Kirchhoff formula that has been shown in Chapter 2:

$$I(x,z) = \int_{r,s} W(r,s,x,z) \times D[r,s,T(x,z,r)+T(x,z,s)]\,ds\,dr. \qquad (4.4.3.26)$$

Replacing $D[r,s,T(x,z,r)+T(x,z,s)]$ by its approximation given by equation (4.4.3.

$$\hat{D}(r,s,T(x,z,r)+T(x,z,s)) = \sum_{i=1}^{N(r,s)} A_i(r,s)w[T(x,z,r)+T(x,z,s)-t_i(r,s)]. \qquad (4.4.3.27)$$

By substituting equation (4.4.3.27) into the Kirchhoff formula (equation (4.4.3.26)), we obtain:

$$\hat{I}(x,z) = \int_{r,s} W(r,s,x,z) \times \sum_{i=1}^{N(r,s)} A_i(r,s)\omega[T(x,z,r)+T(x,z,s)-t_i(r,s)]\,ds\,dr \qquad (4.4.3.28)$$

The weighting operator $W(r,s,x,z)$ can be replaced by two operators, one acting on each input trace and another on each output sample. Therefore, the multiplication with $\omega(r,s,x,z)$ can be implemented as pre and post–summation operations (Gray, 1998). In this case, equation (4.4.3.28) can be replaced by

$$\hat{I}(x,z) = \sum_{r,s} \sum_{i=1}^{N(r,s)} A_i(r,s)\omega[T(x,z,r)+T(x,z,s)-t_i(r,s)]. \qquad (4.4.3.29)$$

Here integration is replaced by summation over all the traces.

At this stage, it is important to realize that the above summation can b carried out using 2 different algorithms. In the first algorithm, matching pursui migration (MPM), we have:

- Algorithm 1

```
for each x,y pixel
 for each triplet r,s,k
  image(x,z) = image(x,z) + A(k,r,s)* w (tray(r,s,x,z)-t(k,r,s))
 end
end
```

In this algorithm, `tray` denotes the traveltime. According to this algorithm, foi each trace the migration is performed by going through all the output (x, y) pixels to find where the `triplets` or atoms (which are used to represent the trace) need to be migrated. This algorithm is not efficient because of the loop on x, y. Another algorithm is given by the following pseudo–code. I will call it 'MPM&ITT' (matching pursuit migration and inversion of traveltime tables).

- Algorithm 2

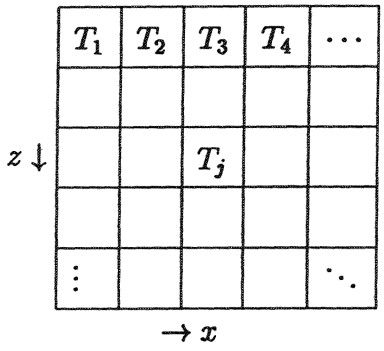
```
for each triplet r,s,k
   find the x,z pixel where the triplet needs to be mapped
    image(x,z) = image(x,z) + A(k,r,s)*w(tray(r,s,x,z)-t(k,r,s))
end
```

It is clear that to find the mapping $(r, s, k) \leftrightarrow (x, z)$ the traveltime tables need to b inverted. Algorithm 2 is an algorithm which combines MPM and traveltime tabl inversion (MPM& ITT), and the cost of finding where to map the atom needs to b analyzed.

4.3.2 A method for inverting traveltime tables

Traveltime tables

A traveltime table for a receiver–source position is a matrix that can be represented as



Given a time T we want to know which are the values $x - z$ that corresponds to the entry T_j in the matrix. Since we do not have an analytical manner to invert the traveltime table we will do it numerically using a search method. Bear in mind that we want to go from a traveltime table of the form $f(x, z, r, s) = T_j$ into a traveltime table of the form $F(T_j, r, s) = (x, z)$.

Shell method

After applying MP to a trace, the k atoms that have been matched, their correlation coefficients and time positions $t(k, r, s)$ are stored in the computer disk. The atoms are in an unordered array for each trace corresponding to a receiver–source pair (r, s). Therefore, a fast sorting method can help improve the efficiency of finding the (x, z) positions corresponding to these atoms.

There exist several sorting methods. *Quicksort* and *Heapsort* are known to be the best algorithms (Sedgewick, 1988). If one needs to sort N elements, $N > 1000$, Quicksort and Heapsort are good candidates to perform the task in an efficient way. If $N < 50$, roughly, *Shell's method* can be used. In the case of $N < 20$, *Straight insertion* is also a very fast alternative (Press et al., 1992). In our code, we usually match about 50 atoms, therefore, the Shell method is efficient enough for our sorting algorithm and its operation count is proportional to k^2 (here k is the number of atoms). In our algorithm, the atoms are sorted in time according to increasing order.

Index sequential method

In the procedure of inverting the traveltime table, we used the index sequential method to search in the traveltime table the (x, z) pixels where the triplets (k, s, r) need to be mapped. Figure 4.4 shows the scheme that we have implemented to search the traveltime table. The index sequential method first searches in the traveltime table the (x, z) pixels corresponding to the first and last atoms (the operation count is proportional to $2 * nx * nz$). Those pixels (x, z) are labeled (indexed). Then only between those labeled positions in the traveltime table, the next search is carried out. A new search is done for the second atom and the one before the last (the

operation count is proportional to $(nx * nz - l_1) * (nx * nz - l_2)$ where l_1 and l_2 are the labeled positions), and the new pixel positions are labeled to be used in the next search. This procedure is iterated until all the pairs (x, z) have been found. In each iteration, the search is performed in a smaller subset of the traveltime table, therefore the index sequential method is a $nk * nx * nz - (l_1 + l_2 + \cdots + l_{k-2})$ process and the computational cost of Algorithm 2 is $nk * nx * nz - (l_1 + l_2 + \cdots + l_{k-2}) + k^2$. Since k^2 is usually a small value compare to $l_1 + \cdots + l_{k-2}$, Algorithm 2 is faster than Algorithm 1 whose operation count is proportional to $nk * nx * nz$.

4.3.3 Examples

Marmousi example

We carried out output–based post–stack depth migration on the Marmousi model (Versteeg, 1994). The Marmousi model is a 2D synthetic data set that is useful in determining an algorithm ability to image seismic data. Figure 4.5 (Top) is the unmigrated Marmousi model which is computed using finite–differencing method with a Marmousi velocity model. The velocity model has been decimated and smoothed to a size of 243×495 (Figure 4.5 (Bottom)). We also compare the migrated section using the standard Kirchhoff migration method (Figure 4.6) and using MPM&ITT. Figure 4.7 and Figure 4.8 illustrate the migrated sections using MPM&ITT. Different numbers of atoms were used in the MP approximation. Figure 4.7 (Top) is the migrated image using 5 atoms. The image is not clear enough to display the main geological structures. Figure 4.7 (Bottom) is the migrated image using 10 atoms. Figure 4.8 (Top) and (Bottom) show the migrated images using 20 and 50 atoms. We can see that with $N = 50$ the MP migrated image starts to look very similar to the standard migration. The strong artifacts in the migrated image are

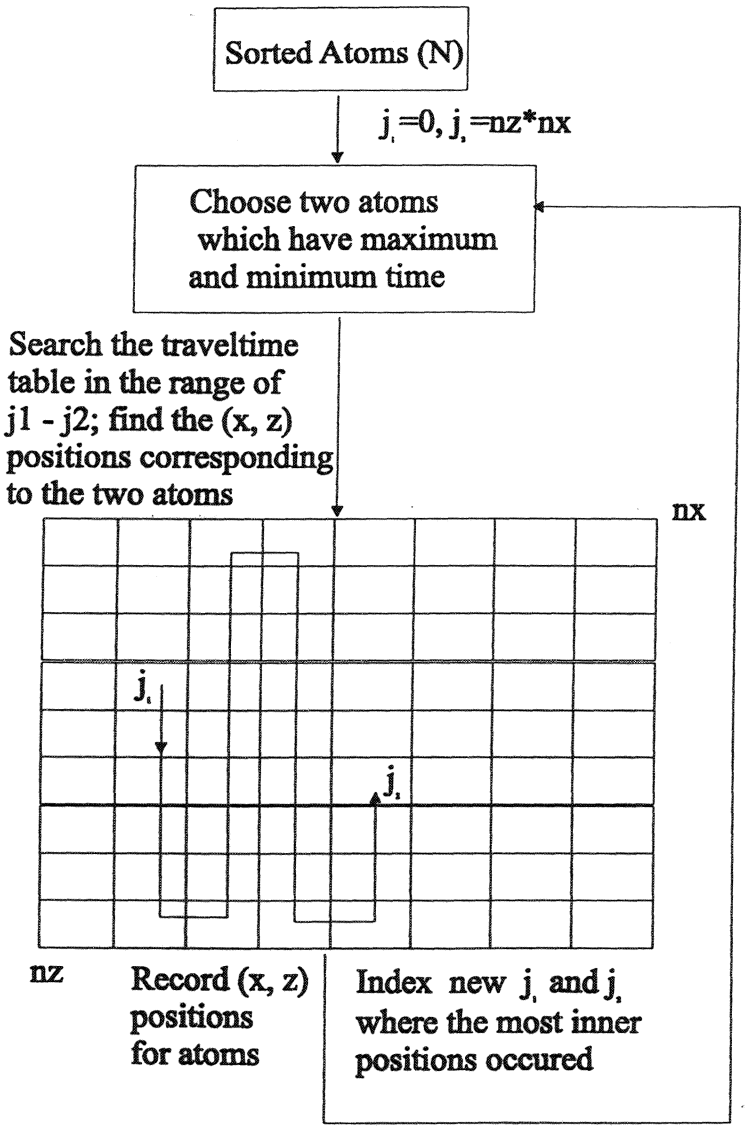


Figure 4.4: The searching scheme for TTT.

due to the traveltime tables which are not accurate enough. The traveltime tables were obtained by solving the Eikonal equation (Vidale, 1988). The problem with the grided Eikonal solver is its inability to control the propagation angle which may result in affecting a migration by producing artifacts (Gray, 1994).

Pre–stack example

We also carried out pre–stack depth migration on synthetic data with 10, 50 and 200 shots using the standard Kirchhoff method, MPM and MPM&ITT algorithm. In Figure 4.9, we illustrate a one shot gather (Left) and its approximation after matching pursuit with 8 atoms (Right). In Figure 4.10, we portray the migrated image using ordinary Kirchhoff migration (Top) and using MPM&ITT (Bottom). Although Figure 4.10 (Bottom) is only a approximation of Figure 4.10 (Top), it shows all the reflection events. The time reports for the standard Kirchhoff migration, MPM, and MPM&ITT are shown in Table 4.1. The first column shows the number of shots we used to migrate the data. The second column is the running time for the Eikonal equation solver, which is used to provide the traveltime tables. The 3th and 4th columns are the running times for matching pursuit and for migration in conventional MP migration. The 5th, 6th and 7th columns are the running times of matching pursuit, inverting the traveltime table and migration in our new method. The last column contains the time for the standard Kirchhoff migration. We see from Table 4.1 that the new approach MPM&ITT is faster than the other two algorithms.

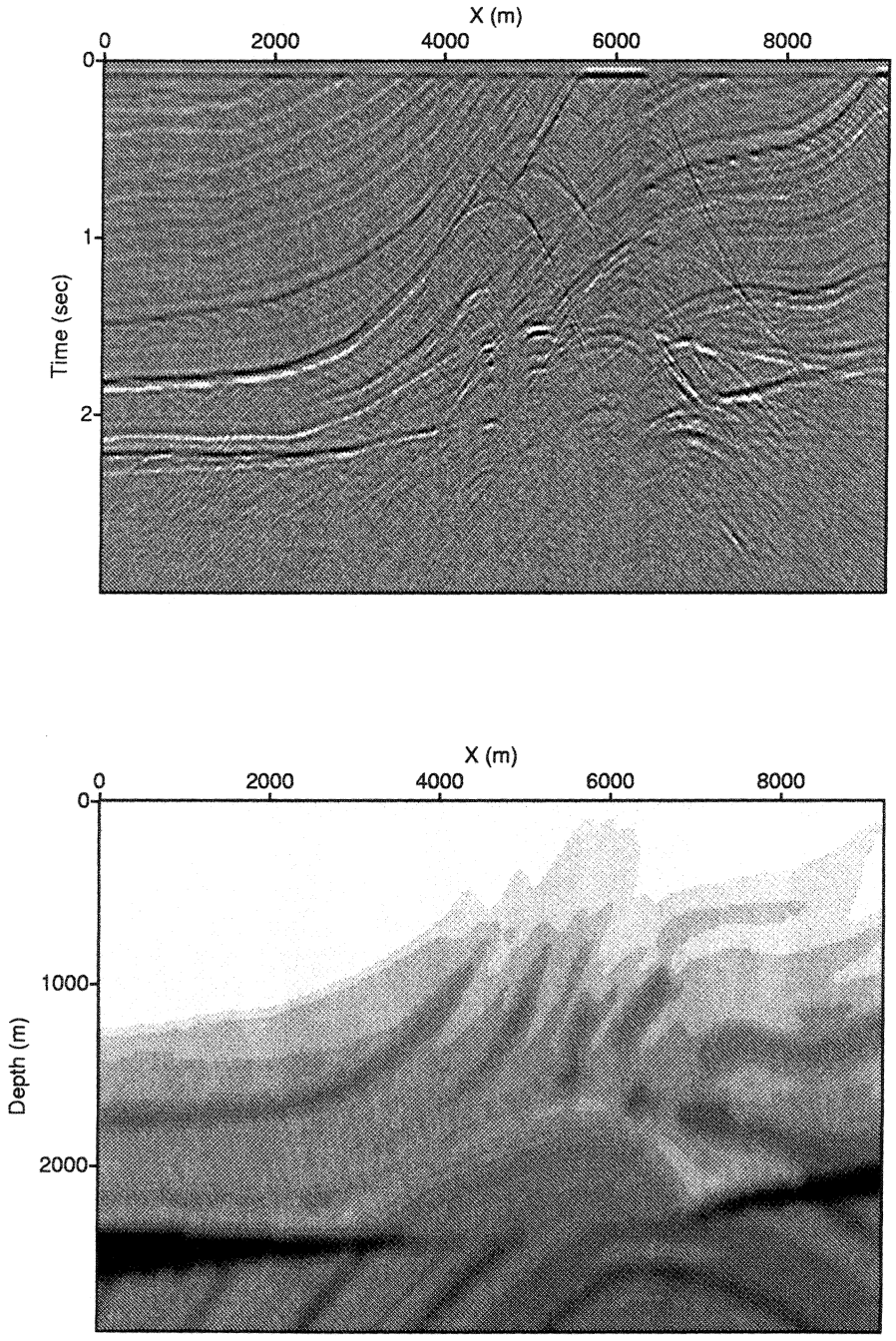


Figure 4.5: Top: Marmousi unmigrated section. Bottom: Smoothed Marmousi velocity model.

Number of shots	MP migration						Ordinary migration
	Eikonal	MPM		MPM&ITT			
		MP	Migration	MP	Search	Migration	
10	90	2.5	8.3	2.5	4.3	1.1	9.3
50	436.1	12.4	61.6	12.3	30.9	11.1	68.9
200	2037.3	51.4	279.1	53	139.6	50.9	301.4

Table 4.1: Time reports for the standard Kirchhoff migration, MPM, and MPM&ITT. (These tests were obtained using a SGI Origin 2000 computer.)

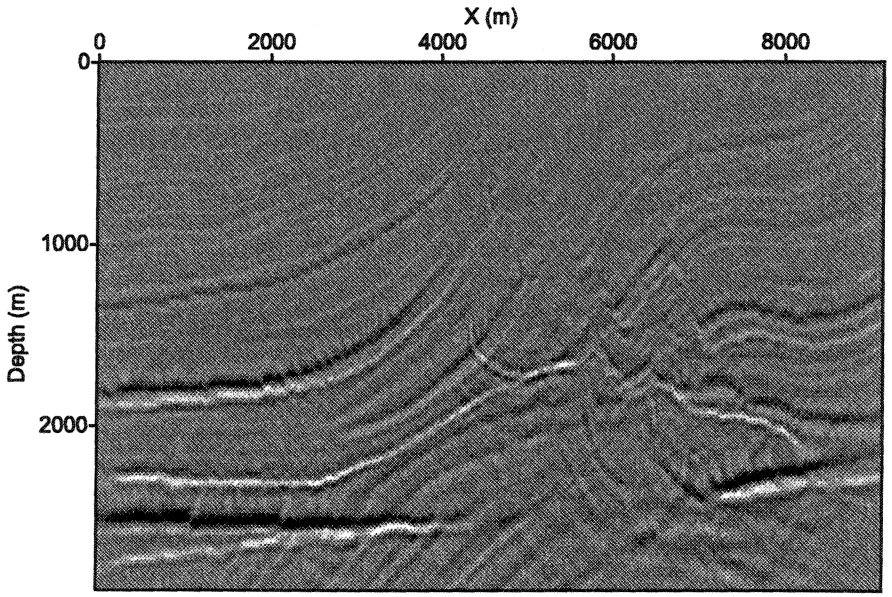


Figure 4.6: Migrated image using the standard Kirchhoff migration method.

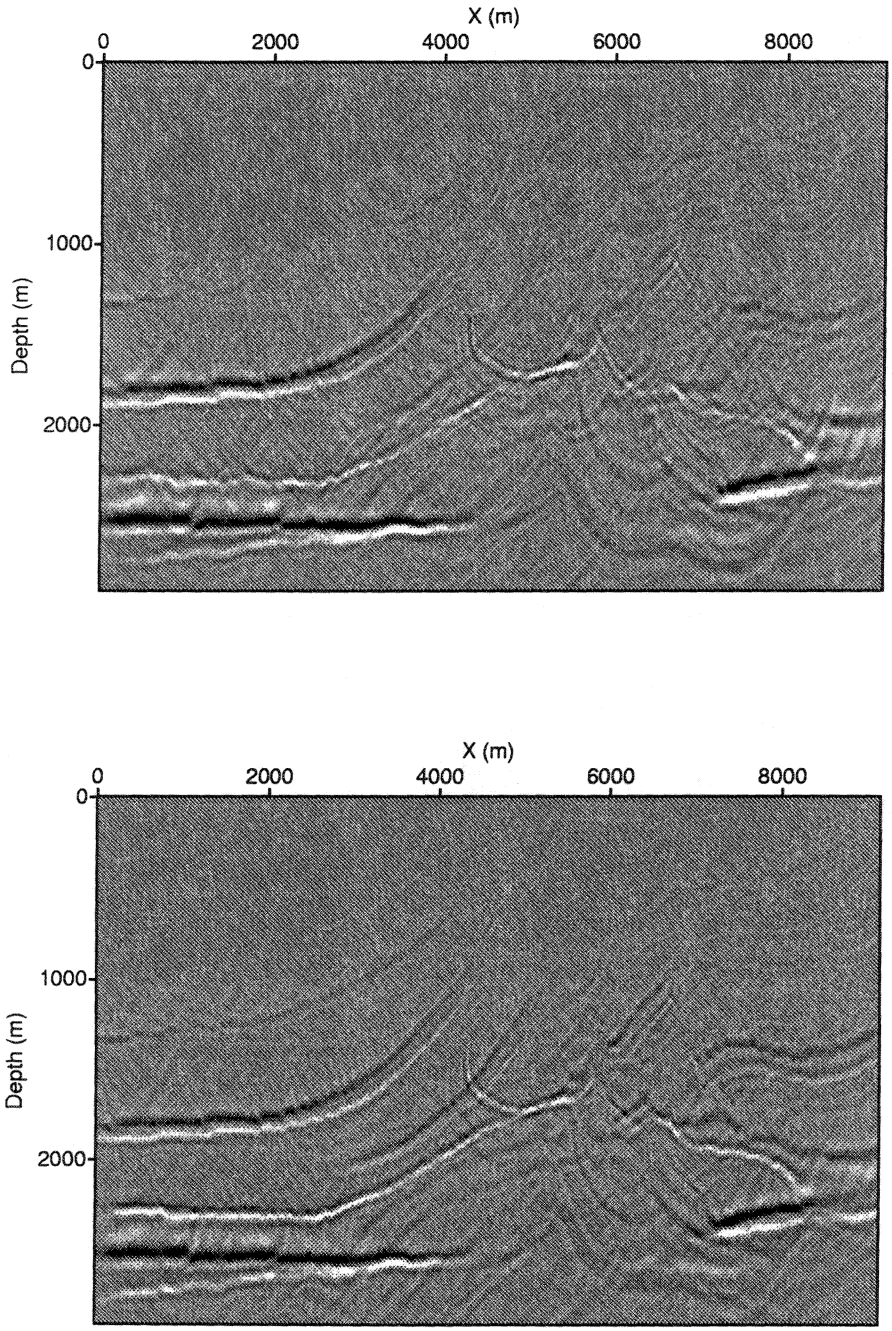


Figure 4.7: Top: Migrated image using MPM&ITT and 5 atoms are used for MP approximation. Bottom: Used 10 atoms for MP.

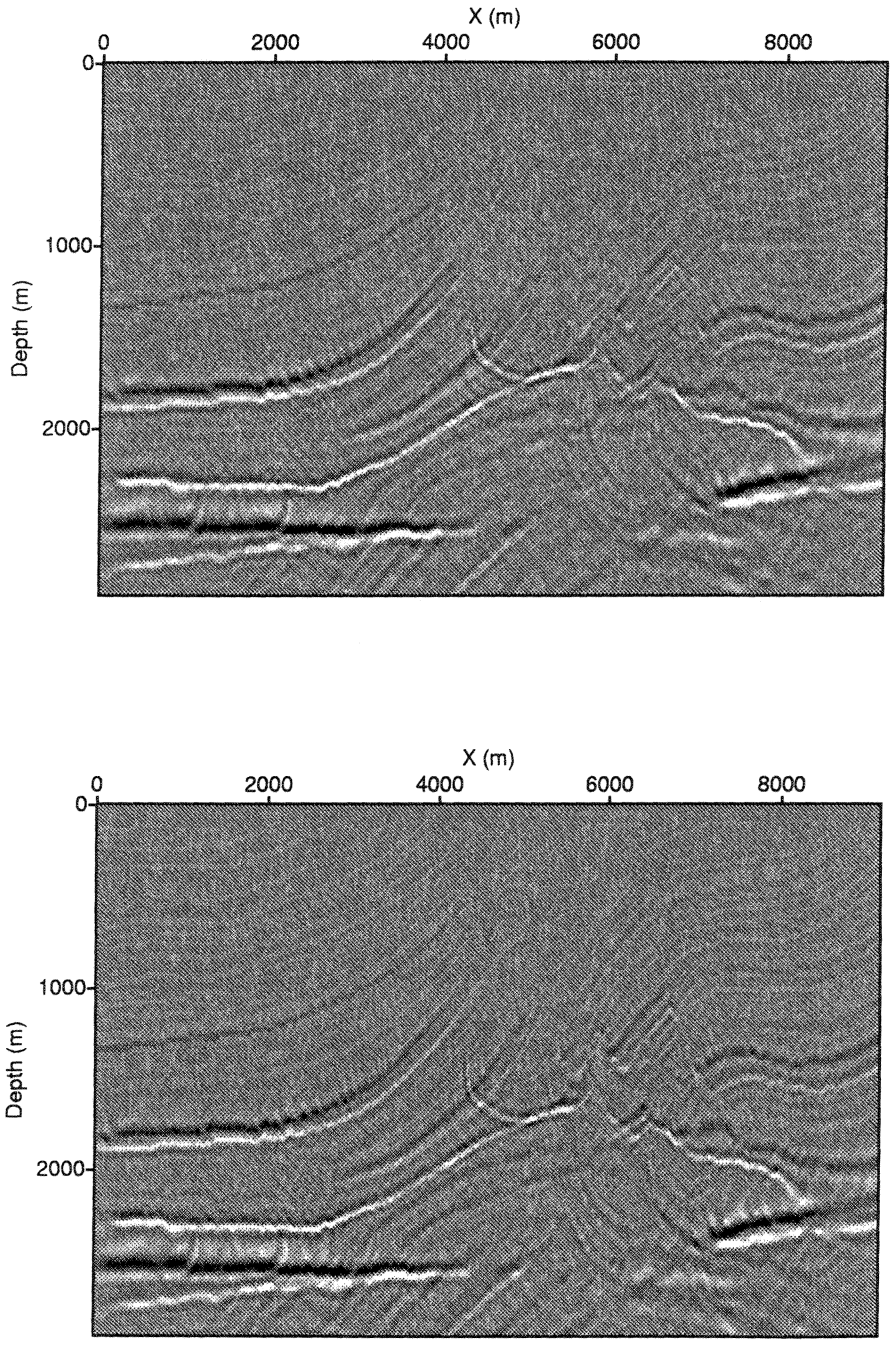


Figure 4.8: Migrated sections using MPM&ITT. Top: Used 20 atoms for MP. Bottom: Used 50 atoms for MP.

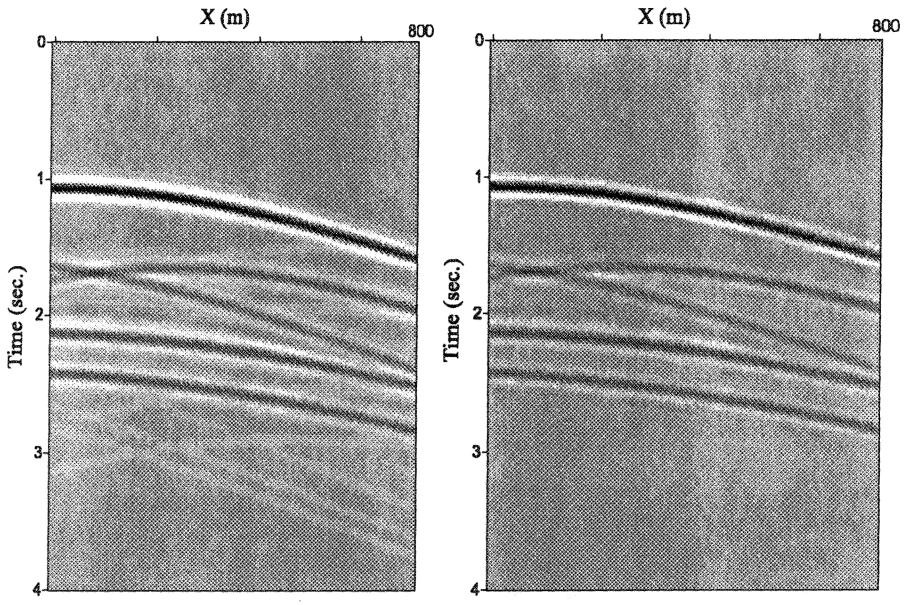


Figure 4.9: Left: One shot gather. Right: MP approximation of the shot gather using 6 atoms.

4.4 Summary

In this chapter, we have discussed the efficiency of Kirchhoff migration on matching pursuit (MP) compressed data. We presented a fast migration approach that is suitable for output based depth migration. In this approach, a seismic trace is decomposed into a linear expansion of adaptive waveforms using a matching pursuit procedure. The Maximum Entropy technique is used to estimate the waveforms that best match the structure of a seismic trace. Then, searching and sorting methods are used to invert the associated travel time tables and to find the positions in the depth domain to where the waveforms will be migrated. Finally, moving those waveforms to a depth image finalizes the migration process.

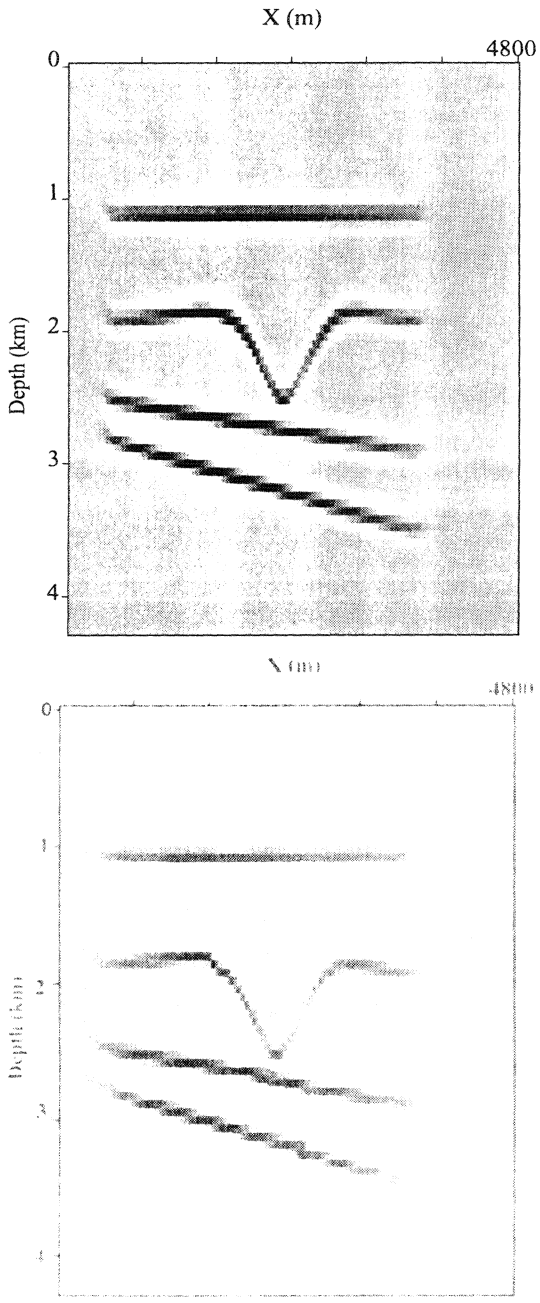


Figure 4.10: Migrated image using the standard Kirchhoff migration (Top) and MPM&ITT (Bottom).

CHAPTER 5

Conclusion and future studies

This thesis examines the problem of migrating seismic data that has been compressed using the wavelet transform and matching pursuit algorithms.

In Chapter 3, the theory of wavelet transform is presented, and some practical aspects are discussed. The WT is used to isolate seismic diffraction in post–stack data, and seismic migration is applied to the isolated energy. The wavelet transform is useful for input based (time migration), where there is an analytical expression for the migration operator.

The proposed method only explores the wavelet transform as a tool to perform seismic data compression. We also realize that WT can be an attractive alternative to solve partial deferential equations.

In Chapter 4, we discuss the efficiency of Kirchhoff migration on matching pursuit (MP) compressed data. We present a new fast migration approach that is suitable for output based depth migration. In this new approach, a seismic trace is decomposed into a linear expansion of adaptive waveforms using a matching pursuit procedure. The waveforms are either Ricker wavelets or estimated wavelets using the Maximum Entropy Method. Then, search and sort methods are used to invert the associated travel time tables and to find the positions in the depth domain to where the waveforms will be migrated. Finally, the seismic migration is done by moving those waveforms in depth. The new approach MPM&ITT is appropriate

for depth migration (output based techniques). The new approach was tested both on synthetic post-stack and pre-stack data. The MPM&ITT, is faster than the ordinary MP migration method and the standard Kirchhoff migration method when the output-based depth migration technique is used. Besides, the new approach is quite flexible since when a few atoms are used, it provides a migrated image of the strongest reflection events present in the data. When we increase the number of atoms, the MP migration yields results which are comparable to the migrated image using the standard Kirchhoff migration method.

We also realize that the MP migration algorithm can be improved by considering the following aspects:

- 2D matching pursuit:

  Seismic data can be further compressed by using a 2D scheme in the MP algorithm. In this case we need to define the domain in which the MP algorithm should operate. i.e., shot-receiver domain or midpoint-offset domain.

- Computation of traveltime tables:

  A algorithm that produces naturally ordered traveltimes may be useful to improve the computational cost of MP migration.

Hopefully, future work on MP migration can yield algorithms that may become standard methods to image seismic data in the oil and gas industry.

Bibliography

[1] Al-Shaykh, O., Miloslavsky, E., Nomura, T., Neff, R. and Zakhor, A., 199 Video compression using matching pursuits: IEEE Transactions on Circui and Systems for Video Technology, 123-143.

[2] Berkhout, A. J., 1980, Seismic migration–imaging of acoustic energy in wav field extropolation: Elsevier Science Pub. Co..

[3] Berkhout, A. J., 1984, Seismic migration: Elsevier, Amsterdam.

[4] Beylkin, G., 1998, Invited lecture at ICM98: Documenta Mathematica, Extra Volume ICM 1998, III, 481–490.

[5] Berryhill, J. R., 1979, Wave–equation datuming: Geophysics, **44**, 1329–1333.

[6] Blackman, R. B., and Tukey, J.W., 1959, The measurement of power spectra from the point of view of communications engineering: Dover Press.

[7] Bouska, J., and Gray, S. H., 1998, Migration in the compressed data domain: CSEG Abstracts, 98, 155–156.

[8] Bosman, C., and Reiter, E., 1993, Seismic data compression using wavelet transforms: 63th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1261–1264.

[9] Burg, J. P., 1967, Maximum entropy spectra analysis: Presented at 37th SEG Meeting, Oklahoma.

[10] Chakraborty, A., and Okaya, D., 1995, Frequency–time decomposition of seismic data using wavelet–based methods: Geophysics, **60**, 1906–1916.

[11] Chen, S. S., and Donoho, D. L., 1996, Atomic decomposition by basis pursuit: Technique report, Standford University.

[12] Chui, C. K., 1992, Wavelet: a tutorial in theory and application: Academic Press.

[13] Claerbout, J. F., and Doherty, S. M., 1972, Downward continuation of moveout corrected seismogram: Geophysics, **37** 741–768.

[14] Claerbout, J. F., 1976, Fundamental of geophysical data processing: McGraw-Hill, Inc.

[15] Claerbout, J. F., 1985, Imaging the earth's interior, Blackwell Scientific Publications.

[16] Claerbout, J. F., and Black, J., 1993, Basic Earth imaging: Standford University (*http : //sepwww.stanford.edu/public/docs/bei/toc_html/index.html*).

[17] Daubechies, I., 1988, Orthonormal bases of compactly supported wavelets: Comm. Pure Appl. Math., **41**, 1988, 906–966.

[18] Daubechies, I., 1990, The wavelet transform, time-frequency localization and signal analysis: IEEE Transactions on Information Theory, **36**, No. 5,

[19] Daubechies, I., 1992, Ten lectures on wavelets: SIAM.

[20] Dessing, F., 1997, A wavelet transform approach to seismic processing: Ph.D. thesis, Delft University of Technology, The Netherlands.

[21] Donoho, P. L., Ergas, R. A., and Villasenor, J. D., 1995, High performance seismic trace compression: 65th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 160-163.

[22] Fyfe, D. J., and Kelamis, P. G., 1992, Removing coherent noise using the linear Radon transform: 54th Mtg. Eur. Assoc. Expl Geophys., Abstracts, 550-551.

[23] Gabor, D., 1946, Theory of communication: Journal of the Institute of Electrical Engineers, London, **93**, 429-457.

[24] Gazdag, J., 1978, Wave-equation migration by phase shift: Geophysics, **43**, 1342-1351.

[25] Goodman, J. W., 1968, Introduction to Fourier optics: McGraw–Hill Book Co., Inc.

[26] Gray, S. H., 1994, Kirchhoff migration using eikonal equation traveltimes: Geophysics, **59**, 810-817.

[27] Gray, S. H., 1998, Speed and accuracy of seismic migration methods, unpublished report.

[28] Hagedoorn, J. G., 1954, A process of seismic reflection interpretation: Geophys. Prospecting, **2**, 85-127.

[29] Hilton, M. L., Jawerth, B. D., and Sengupta, A., 1995, Compressing still and moving images with wavelets: Multimedia Systems, **3**, No. 2.

[30] Hubral, P., 1977, Time migration–some ray theoretiacal aspects: Geophys. Prospecting, **25**, 738-745.

[31] Judson, D. R., Schultz, P. S., and Sherwood, J. W. C., 1980, Depth migration after stack: Geophysics, **45**, 361-375.

[32] Lacoss, R. T., 1971, Data adaptive spectral analysis methods: Geophysics, **36**, 661-675.

[33] Li, X. G., Sacchi, M. D., and Ulrych, T. J., 1994, Wavelet transform in sion: application to 1D tomography, 64th Annual Internat. Mtg., Soc. E Geophys., Expanded Abstracts, 972–975.

[34] Li, X. G., Sacchi, M. D., and Ulrych, T. J., 1996, Wavelet transform invers with a priori scale information: Geophysics, **61**, 1379-1385.

[35] Li, X. G., and Ulrych, T. J., 1996, Multi-scale attribute analysis and tr decomposition: 66th SEG Meeting, Expanded Abstracts, 1634-1637.

[36] Li, X. G., Wang, B., Pann, K., Anderson, J., and Deng, L., 1998, Fast migrati using a matching pursuit algorithm: 68th SEG Annual Internat. Mtg. Soc. Ex Geophys., Expanded abstracts, 1732-1735.

[37] Louis, A. K., Maass, P., and Rieder, A., 1997, Wavelets: theory and applic tions: J. Wiley, New York.

[38] Mallat, S., 1989, A theory for multiresolution signal decomposition: the wavele representation: IEEE Transactions on Pattern Analysis and Machine Intelli gence, **11**, No. 7, 674–693.

[39] Mallat, S., and Zhang, Z.F., 1993, Matching pursuit with time-frequency dic tionaries: IEEE Tansactions on Signal Processing, **41**, 3397-3415.

[40] Meyer, Y., 1989, Orthonormal wavelets: IXth International Congress on Math ematical Physics, 38–47.

[41] Miao, X. G., and Moon, W.M., 1994, Application of the wavelet transform in seismic data processing: 64th Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts, 1461-1464.

[42] Neff, R., Zakhor, A., and Vetterli, M., 1994, Very low bit rate video codin using matching pursuits: in Proceedings of SPIE Conference on Visual Com munication and Image Processing, **2308**, No. 1, 47-60.

[43] Parzen E., 1969, Multiple time series modeling: Academic Press.

[44] Perrier, Valérie and Basdevant, C., 1995, Wavelet spectra compared to Fourie spectra: Journal of Mathematical Physics, **36**, No. 3, 1506–1519.

[45] Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 1992, Numerical recipes in Fortran 77: the art of scientific computing (second edition): Cambridge University Press.

[46] Robinson, E. A., 1967, Predictive decomposition of time series with applications to seismic explorations: Geophysics **32**, 418-484.

[47] Robinson, E. A., and Treitel, S., 1980, Geophysical signal analysis: Prentice-Hall, Inc.

[48] Robinson, J. C. and Robbins, T. R., 1978, Dip–Domain migration of two-dimensional seismic profiles: Geophysics, **43**, 77-93.

[49] Sedgewick, R., 1988, Algorithm, 2nd ed.: Addison–Wesley.

[50] Schneider, W., 1978, Integral formulation for migration in two and three dimensions: Geophysics, **43**, 49-76.

[51] Schultz, P. S., and Sherwood, J. W. C., 1980, Depth migration before stack: Geophysics, **45**, 376-393.

[52] Stolt, R. H., 1978, Migration by Fourier tansform: Geophysics, **43**, 23-48.

[53] Strang, G., and Nguyen, T. 1997, Wavelet and filter banks: Wellesle Cambridge Press.

[54] Strang, G., 1986, Introduction to applied mathematics: Wellesley–Cambridg Press.

[55] Ulrych, T. J. and Bishop,T. N., 1975, Maximum entropy spectral analysis an autoregressive decomposition: Res. Geophys. and Space Phys., **13**, 183-200.

[56] Versteeg, R., 1994, The Marmousi experience: Velocity model determinatio on a synthetic complex data set: The Leading Edge, **13**, No. 9, 927-936.

[57] Vidale, J., 1988, Finite-difference calculation of traveltimes: Bull. Seis. Soc. Am., **78**, 2062–2076.

[58] Wang, B., and Pann, K., 1996, Kirchhoff migration of seismic data compressed by matching pursuit decomposition: 66th Annual International Meeting, SEG, Expanded Abstracts, **96**, 1642-1645.

[59] Yilmaz, Ö., 1987, Seismic data processing: Society of Exploration Geophysicist, Tulsa.