

The double focal transformation

**and its application to
data reconstruction**

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen
op vrijdag 31 januari 2014 om 12.30 uur

door

Hannes KUTSCHA

Diplom in Geophysik,
Westfälische Wilhelmsuniversität Münster, Duitsland,
geboren te Sondershausen, Duitsland.

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

en copromotor:
Dr. ir. D.J. Verschuur

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. A. Gisolf	Technische Universiteit Delft, promotor
Dr. ir. D.J. Verschuur	Technische Universiteit Delft, copromotor
Prof. dr. M. van der Baan	University of Alberta, Canada
Prof. dr. ir. G.J.T. Leus	Technische Universiteit Delft
Prof. dr. ir. C.P.A. Wapenaar	Technische Universiteit Delft
Dr. P. Zwartjes	Shell International, Advisor
Dr. C.X. Tsingas	Saudi Aramco, Advisor
Prof. dr. H.P. Urbach	Technische Universiteit Delft, reservelid

SUPPORT

The research for this thesis was financially supported by the DELPHI Consortium.

ISBN 978-94-6186-272-3

Copyright © 2014 by H.Kutscha

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilised in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the prior permission of the author.

Typeset by the author with the L^AT_EX 2_& Documentation System.

Published by Uitgeverij BOXPress, 's-Hertogenbosch, The Netherlands
Printed by Proefschriftmaken.nl

Contents

1	Introduction	1
1.1	The seismic method	1
1.2	Sampling of the seismic wavefield	2
1.3	Data reconstruction methods	3
1.3.1	Filter-based reconstruction	4
1.3.2	Transformation-based reconstruction	4
1.3.3	Wavefield operator-based reconstruction	6
1.4	The focal reconstruction	7
1.5	Thesis outline	7
2	The model of the double focal transformation	11
2.1	Wavefield extrapolation using operator notation	11
2.2	Extension to three dimensions	19
2.3	Multi-reflector data model	23
2.4	From the WRW model to the concept of double focal transformation	26
2.4.1	The multi-level implementation	26
2.4.2	The single-level implementation	31
3	Implementation of the focal reconstruction	35
3.1	The forward double focal transformation by correlation	35
3.1.1	Manual filtering of aliasing noise	42
3.2	The forward double focal transformation by optimisation	43
3.2.1	Least-squares optimisation	44
3.2.2	Sparseness constrained optimisation	47
3.3	The multi-level implementation	59
4	Examples	63
4.1	Reconstruction of synthetic data of a salt model	63
4.2	Reconstruction of a marine field dataset	68
4.2.1	Reconstruction of the near-offset gap	69

4.2.2	Reconstruction of strongly aliased data	72
4.2.3	Comparison with sparse linear Radon reconstruction	72
5	Focal reconstruction for 3D data	77
5.1	3D data representation	77
5.2	Focal reconstruction for 3D data	78
6	Utilising the linear Radon transformation	83
6.1	The general work flow	84
6.2	Implementation	86
6.3	Results	87
7	Conclusions, Discussion and Recommendations	89
7.1	Conclusions	89
7.2	Discussion and Recommendations	90
7.2.1	The effect of quality and quantity of the utilised focal operators	90
7.2.2	Utilise the linear Radon transformation in combination with the double focal transformation	91
7.2.3	Decreasing computational costs	94
7.2.4	3D versus 2D	95
7.2.5	Iterative implementation of the focal reconstruction	95
7.2.6	Utilising multiple reflections	95
Appendices		99
A	The derivation of the Rayleigh integral	101
A.1	The acoustic wave equation	101
A.2	The Kirchhoff integral	102
A.3	Causality of the Kirchhoff integral	104
A.4	The Rayleigh integrals in a homogeneous medium	107
A.4.1	The Rayleigh integral in 2D space	112
A.5	The Rayleigh integral in inhomogeneous low contrast media	115
A.6	Rayleigh integrals in terms of convolution	116
B	The wave character of the WRW model	119
C	Inversion and Optimisation	123
C.1	The pseudoinverse	123
C.1.1	Basic definitions	123
C.1.2	Utilised equations	124
C.1.3	Derivation of the pseudoinverse	125
C.2	The spgl1 solver	126

C.2.1 The Newton iteration	127
C.2.2 The spectral projected gradient method	128
C.2.3 The projection	128
C.3 The matching pursuit solver	129
D Notation	131
Bibliography	133
Summary	139
Samenvatting	141
CV note	143
Acknowledgements	145

Chapter 1

Introduction

1.1 The seismic method

In order to locate reservoirs that contain oil and gas, or to locate storage possibilities for CO_2 , an image of the subsurface is needed. Such an image can be obtained by seismic imaging, which is a technique that is based on the physics of wavefield propagation.

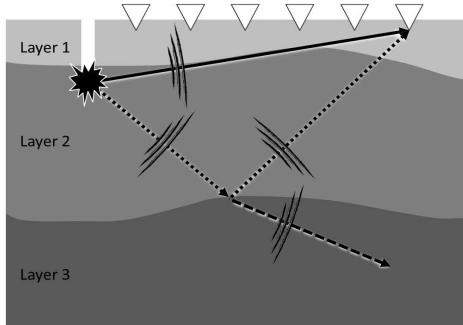


Figure 1.1: Sketch of a simple seismic acquisition. The black star represents the seismic source and the seismic receivers are denoted by white triangles. The different properties (i.e. velocity and density) of the three layers are denoted by different shades of gray.

Figure 1.1 sketches the principle of a simple seismic acquisition. A source, most commonly a seismic vibrator or dynamite on land, or an air gun at sea, excites a wavefield, usually at or close to the surface. In the example shown, dynamite was placed in a borehole. The source wavefield denoted by wavefronts and their corresponding outward pointing normals (i.e. raypaths) propagates through the subsurface. Some part of the wavefield propagates directly (direct wavefield) to

the receivers at the surface (solid line). Another part propagates downwards (dotted line). If it reaches a property discontinuity (here between layer two and three) two phenomena occur: one part of the incident wavefield is reflected (dotted line) and another part is refracted (dashed line). In Figure 1.1 only the direct and the reflected wavefield is propagating upwards to the receivers, but the refracted wavefield might be reflected at a deeper property contrast (i.e. interface between two layers).

The goal of seismic imaging is to use the detected wavefield to determine the structure and the property distribution in the subsurface. In order to increase the amount of information the experiment of Figure 1.1 can be repeated at different locations, or with different set-ups. Subsequently an image of the subsurface is constructed by backwards propagating and focusing the seismic measurements to all depth levels. In that way a representation of the subsurface structure (i.e. layer boundaries) is obtained. This knowledge is commonly used to support drilling decisions. However, in many cases these images can not be directly generated. The reason is that often within a certain area not enough receivers are used, or not enough experiments are carried out. This inadequate data acquisition leads to a coarse sampling of the recorded wavefield. In that case the missing information has to be filled in afterwards by making assumptions on the subsurface structure. The process to generate information of the wavefield at any desired location from coarse measurements is called seismic data reconstruction.

1.2 Sampling of the seismic wavefield

This thesis deals with one of the acquisition based problems: coarse sampling of the seismic wavefield and large data gaps.

The sampling of an incoming wavefield in time is unproblematic, since the recorded signals usually have a frequency range up to 150Hz which can be sampled uniformly and non-aliased with modern equipment. In 3D seismic surveys the wavefield is sampled in four spatial dimensions, providing two lateral degrees of freedom (x and y) for both sources and receivers. The sampling is rarely dense in all four spatial dimensions. The reason for large data gaps can be faulty equipment, inaccessibility or strong noise sources at certain locations such that the measured data become unusable. Also the near-offset range usually can not be detected, because the sensitive detectors would be damaged if placed closely to a seismic source. The most common reason for coarse sampling, however, is the trade-off between geophysical and economical constraints. Often a huge amount of sources and receivers would be necessary to densely cover the whole area of interest in order to avoid the occurrence of aliasing artefacts. In practice this criterion is rarely met, because it would be very time-consuming and expensive to carry out the acquisition. Furthermore, it would require an extensive amount of resources to process the acquired data. Usually, in land seismic acquisition receivers are coarsely sampled in one spatial direction and sources in the other

direction (Figure 1.2a). For marine acquisition the sampling for sources and receivers in the in-line direction is usually sufficient but it is often coarse in the cross-line direction (Figure 1.2b). Another example is the distribution of ocean bottom nodes at the ocean floor, which is relatively expensive. Therefore, they are usually coarsely distributed in both lateral directions, while the sources, which are fired at the water surface, are distributed densely in both lateral directions (Figure 1.2c). For more information on acquisition designs the reader is referred to Vermeer (2002).

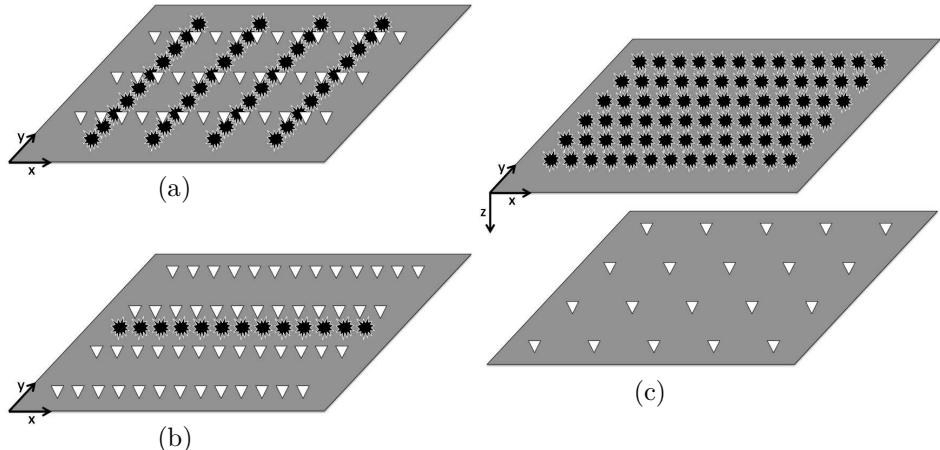


Figure 1.2: Scheme of typical acquisition designs. The black stars represent seismic sources and the white triangles represent seismic receivers. They are usually placed at or close to the surface (indicated by the grey plane). In (a) a typical land acquisition set-up is displayed, where the source sampling is dense in the y -direction and coarse in the x -direction and the receiver sampling is dense in the x -direction and coarse in the y -direction. In (b) a typical marine acquisition set-up is displayed, where sources and receivers are both densely sampled in the x -direction and coarsely sampled in the y -direction. In (c) an acquisition design is displayed, where sources are densely sampled at the water surface (upper grey plane) in both spatial directions and the receivers (ocean bottom nodes) are coarsely sampled at the ocean bottom (lower grey plane).

However, coarse sampling in one or more directions generates aliasing artefacts in subsequent seismic processing steps. This thesis describes a method of filling in information in the data gaps, to achieve a dense and regular sampling.

1.3 Data reconstruction methods

In order to avoid the effects of aliasing and non-uniform sampling in the processing flow, the data is reconstructed on a dense, regular grid. Please note that the following definition is used, to avoid confusion with the terminology: *Regular-*

isation allows to use input data on any kind of grid (regular/irregular/random) to produce output data on another grid (usually regular). However, in this case the input data should not be spatially aliased, such that the Nyquist theorem is obeyed. In *reconstruction* the input data is assumed to be spatially aliased. Hence, the difference of reconstruction compared to regularisation is that the aliasing noise has to be removed to obtain the data at the desired grid. There is the possibility to distinguish further between interpolation and reconstruction. For instance if the input data is part of the output data the term *interpolation* is used, and if the output data could be on a completely different grid than the input data the term *reconstruction* is used. Thus, for *interpolation* the original data is kept in the output result, while new data points are inserted. But these notations are not very clear and there is, to best of our knowledge, no commonly used definition that separates these two terms. Therefore, in this thesis the term interpolation is omitted. In the next three subsections the three groups of reconstruction methods as introduced by Zwartjes (2005) will be described.

1.3.1 Filter-based reconstruction

Filter-based methods reconstruct data from existing measurements by convolution with an interpolating filter in the spatial domain, for instance the sinc function. Please note that this process is analogue to straightforward Fourier regularisation (transforming from the space domain to the wavenumber domain, padding with zeros and transforming back). To regularise non-uniformly sampled data, the data is convolved with e.g. a Gaussian window and the output is placed on a regular grid. An example is the method of generalised convolution (Knutsson and Westin, 1993). A common method for the reconstruction of uniformly sampled, aliased data is the usage of prediction error filters (Claerbout, 1990; Spitz, 1991). Fomel and Claerbout (2001) discuss how to use prediction error filters to fill gaps in uniformly sampled data sets. In general, these methods are not suited for random sampling geometries, but by binning the data to a fine grid the random sampling geometry is replaced by a uniform grid with missing samples. Please note that these methods usually don't perform well for strongly aliased data, very irregularly sampled data or data with large data gaps. As we will show later in section 4.2 the double focal reconstruction, on the other hand, is capable to deal with data gaps up to 200m in field data.

1.3.2 Transformation-based reconstruction

In order to handle the limitations of filter-based reconstruction methods, transformation-based methods aim at utilising the redundancy in seismic data by assuming that the data can be efficiently described in some transform domain. Non-uniformly and/or coarsely sampled data can then be reconstructed in a two-step approach. First, the transform coefficients are derived from the input data (step

one). Once these model parameters are estimated, the densely sampled output data is derived simply by an inverse transform (the second step).

The reconstruction step involves an optimum mapping from the data space to the model space (step one). If this mapping is carried out by simple forward transformation, the model parameters will be distorted (i.e. aliasing noise appears in the model space, due to the missing traces of the coarse input data). By suppressing the aliasing noise (e.g. by muting or suppressing certain areas in the model space) only the signal remains that corresponds to the densely sampled input data.

However, such muting approach presumes that it is actually possible to distinguish the signal from the aliasing noise in the model space, which is most of the time not the case. Therefore, the current state-of-the-art reconstruction methods derive the model space by constrained inversion of the coarse input data. Depending on the chosen transformation the constraint will automatically suppress the aliasing noise in the model space. This is possible because the aliasing noise has different properties in the model space than the signal. Well-known transformations that are used for the transformation-based reconstruction method are the Fourier transformation (Duijndam et al., 1999; Xu et al., 2005; Zwartjes and Gisolf, 2006; Zwartjes and Sacchi, 2007; Schonewille et al., 2009; Trad, 2009), the linear, parabolic and hyperbolic Radon transformation (Kabir and Verschuur, 1995; Herrmann et al., 2000; Trad, 2002; Verschuur et al., 2012) and the curvelet transformation (Hennenfent and Herrmann, 2008; Hennenfent et al., 2010).

In general, such inversion-based approach aims at finding the optimum transform domain parameters such that by the inverse transformation from the model space to the data space - at the original measurement locations - the input data is reproduced. In addition, the involved constraint will select one of the many possible representations in the transform domain, e.g. via a sparseness criterion. Such inversion, thus, aims at minimising the following objective function:

$$J = \|\vec{p}_m - \mathbf{L}^H \vec{x}\|_2 + \lambda F(\vec{x}), \quad (1.1)$$

where the vector \vec{p}_m represents the measured data (stored in one long vector), the vector \vec{x} contains the model domain parameters stored in one long vector and the matrix \mathbf{L} describes the transform from the data space to the model space, such that the $(\cdot)^H$ describes the conjugate transpose operation, i.e. a mapping from the model space to the data space. Function $F(\cdot)$ describes the additional constraint that is included in order to force to a desired and unique solution. Often, this function describes the l_1 norm or another sparseness promoting norm in the transform domain. This approach is followed by a.o. Trad (2003); Zwartjes and Gisolf (2006) and Verschuur and Kutsch (2009).

Minimising the objective function J (equation 1.1) and choosing the additional constraint F as the l_1 norm gives the formulation of the Basis Pursuit Denoise (BPDN) problem in its unconstrained formulation:

$$\underset{\vec{x}}{\text{minimise}} \|\vec{p}_m - \mathbf{L}^H \vec{x}\|_2 + \lambda \|\vec{x}\|_1. \quad (1.2)$$

The BPDN problem in its constrained formulation reads:

$$\underset{\vec{x}}{\text{minimise}} \|\vec{x}\|_1 \text{ subject to } \|\vec{p}_m - \mathbf{L}^H \vec{x}\|_2 \leq \sigma, \quad (1.3)$$

which, for any given λ in equation 1.2, is equivalent to the unconstrained formulation for some (usually unknown a priori) value of σ . The two problems are quite similar, but most specialised numerical algorithms can only solve the unconstrained formulation. For further information to the method of BPDN, see Gill et al. (2011).

Note that once the optimum model parameters \vec{x} have been found, the input data can be reconstructed to any other measurement location by applying the reverse transform with a new operator that now is based on the new locations:

$$\vec{p}_{rec} = \mathbf{L}_{rec}^H \vec{x}, \quad (1.4)$$

where the matrix \mathbf{L}_{rec} is now redefined for the desired output locations of the reconstructed data. Optionally, the original data traces can be merged with the reconstructed data if the geometry allows this.

1.3.3 Wavefield operator-based reconstruction

The transform-based reconstruction methods rely on the chosen transform domain and, thus, assume that the seismic data can be optimally represented by the involved basis functions (Fourier components, plane waves, curvelets etc.). Recent developments also aim at bringing in more information related to actual wave propagation. Such methods are based on the Rayleigh integral, or in a more general case on the Kirchhoff integral. These integrals can be formulated as the convolutions of the data with a wavefield- or propagation operator. In numerical applications these convolutions (continuous integral equations) are replaced by (discrete) summations. These summations rely on the constructive and destructive interference along the summation path. When the sampling is too coarse aliasing artefacts appear due to incomplete destructive interference.

In Nemeth et al. (1999) and Kuehl (2002) an inversion approach was used to suppress the aliasing artefacts with Kirchhoff operators via least-squares migration. The advantage of wavefield operator based methods is that they allow the utilisation of prior information of the subsurface, for the reconstruction process. However, in such cases it is important that the prior information is sufficiently close to the true subsurface. Furthermore, these methods typically are computationally expensive. Examples for data reconstruction with the wavefield continuation method are offset continuation (Bagaini and Spagnolini, 1996), shot continuation (Spagnolini and Opreni, 1996), DMO (Deregowski, 1986) and AMO (Chopra et al., 2009). All these methods have in common that they utilise the wavefield characteristics in a localised manner, where the involved operators are local convolution operators. Thus, they still do not fully exploit the redundancy within the seismic data.

1.4 The focal reconstruction

The double focal transformation, which I am proposing in this thesis, combines the features of wavefield operator-based methods and of transformation-based methods. It describes the propagation in the subsurface with global wavefield operators but in a transformation manner. Hence, the same type of formulation like equation 1.3 is used, except that the transformation matrix now represents wavefield propagation. The propagated data is defined as the transform domain and hence, forward (going from data space to model space) and inverse (going from model space to data space) double focal transformation represent inverse and forward propagation of the seismic data.

Since it is more convenient to understand the propagation from the subsurface to the surface as "forward mode", the inverse double focal transformation is denoted with \mathbf{L} . Logically, the forward double focal transformation is approximated with \mathbf{L}^H .

There are related methods, like the work of Trad (2003), where migration operators are used as a tool to enhance another transformation based reconstruction method. However, in its property to define propagation itself as a transformation the double focal transformation is unique.

A predecessor of the proposed double focal transformation was presented by Berkhouwt and Verschuur (2006), which was called focal transformation, but it involved subsurface operators that described two-way seismic reflection events. The success of this reconstruction method was highly dependent in describing both propagation AND reflection in the subsurface. This put a severe limitation to that method, as for complex subsurface models generating such operators (which basically resembled the major reflection events in the data) was a non-trivial task.

The newly defined double focal transformation only requires the description of rough propagation operators towards strategic depth levels, whereas the actual - difficult to obtain - reflection properties are estimated automatically by the method.

The focal reconstruction can in principle deal with any kind of input sampling (for instance coarse, regular or irregular) and can also output data on any desired grid. However, in the current implementation the original traces are part of the reconstructed data. Furthermore, for the examples in this thesis the case is investigated where data is acquired in principle on a dense, regular grid. Irregular and coarse data are generated by deleting information from the dense data.

1.5 Thesis outline

The objective of this thesis is to reconstruct densely sampled data, from coarsely sampled seismic data making use of prior information e.g. in form of a macro velocity model. The thesis contains the following chapters:

- Chapter 2: *Theory of the double focal transformation.* In this chapter the basic principles of the double focal transformation are derived. First, the Rayleigh integral is formulated as a convolution integral. The discrete version of this convolution integral forms the basis of the WRW model. Modifying the WRW formulation for a more general case the concept of double focal transformation is derived.
- Chapter 3: *Implementation of the focal reconstruction.* In this chapter it is demonstrated how the double focal transformation can be used for seismic data reconstruction. First, a manual filtering in the focal domain is used. Next, reconstruction is achieved by applying a least-squares constraint to the model space. Finally, it is shown that seismic data can be represented sparsely in the focal domain. A sparseness constraint to the model space suppresses aliasing noise and is, therefore, well suited for data reconstruction. Furthermore, the multi-level implementation of the double focal transformation is utilised. It enables a sparser data representation and, therefore, increases the reconstruction capability.
- Chapter 4: *Examples.* This chapter shows the application of the focal reconstruction to two datasets. First, a 2D synthetic dataset is reconstructed. Hereby, the influence of the quality of the focal operator to the reconstruction quality is investigated. Furthermore, a real marine 2D dataset is reconstructed. It contains a large near-offset gap and is additionally spatially aliased. Hereby the performance of the double focal transformation is compared to state of the art linear and parabolic Radon methods.
- Chapter 5: *3D focal reconstruction.* In this chapter the 3D double focal transformation is introduced. In fact, using the operator notation, the basic theory remains the same as for 2D. However, each operator or dataset now becomes 5D instead of 3D. The method is illustrated with the reconstruction of a simple synthetic 3D dataset.
- Chapter 6: *Utilising the linear Radon transformation.* In this chapter the linear Radon transformation is used in combination with the double focal transformation for the reconstruction of a synthetic 2D dataset. It is demonstrated that this step can increase the sparseness of the data representation in the model space and also can reduce the number of virtual depth levels in the multi-level implementation.
- Chapter 7: *Conclusions, Discussion and Recommendations.* In this chapter the key elements of the double focal transformation and its application to seismic data reconstruction are summarised and discussed. Furthermore, general ideas on the applicability of the double focal transformation in different fields are discussed and how different set-ups can improve its efficiency.

- Appendices: In *Appendix A* the Rayleigh integral is derived from Gauss's theorem and later formulated as a convolution integral. This forms the basis for the WRW model and the double focal transformation. In *Appendix B* the wave character of the WRW model is demonstrated on a simple example. *Appendix C* contains the derivation of the pseudo-inverse, which is used for minimum norm inversion. Furthermore, the basic principles of the *spgl1* solver and the matching pursuit solver are introduced as tools for sparse inversion. *Appendix D* gives a brief overview of the notation and abbreviations used within this thesis.

Chapter 2

The model of the double focal transformation

In this chapter it is demonstrated how the double focal transformation can be derived from the basic principles of wavefield extrapolation. For the derivation of the Rayleigh integral and its formulation as a convolution integral, the reader is referred to Appendix A. There, first the Kirchhoff integral is derived, which allows wavefield propagation to any point in an object from measurements taken at the surface all around this object. However, in most cases of field surveys only one-sided illumination and detection is possible. Therefore, the Rayleigh integral has been developed from the Kirchhoff integral. It allows the computation of a wavefield in a point that is separated by an infinitely extended flat surface from the wavefield generating sources. The field has to be measured along this flat plane. Note that the derivation of the Rayleigh integral can be also found in great detail in Gisolf and Verschuur (2010). Finally, the Rayleigh integral is formulated as a convolution integral, which is the basis of the WRW model (Berkhout, 1982) and the double focal transformation.

Please note that throughout the thesis capital letters denote variables in the frequency domain. Variables in the time domain are denoted in lower case. A superscript '+' denotes propagation in the positive z-direction (downwards) and a superscript '-' denotes propagation in the negative z-direction (upwards).

2.1 Wavefield extrapolation using operator notation

The Rayleigh integral is derived in Appendix A. However, for the sake of convenience it is given here again:

$$P(x_A, y_A, z', \omega) = \iint_{-\infty}^{\infty} dx dy W^+(x, x_A, y, y_A, \Delta z, \omega) P(x, y, z_0, \omega). \quad (2.1)$$

It describes the extrapolation of a wavefield P from the point (x, y, z_0) to the point (x_A, y_A, z') . For the notation we will follow Berkhout (1982). In the case that the wavefield of many points at the surface (z_0) is downward extrapolated to many points at the subsurface (z') the Rayleigh integral can be interpreted as the convolution of a wavefield with a propagation operator as is demonstrated in Appendix A section A.6. Note that in the case of an inhomogeneous subsurface the propagation operator is non-stationary, while in the case of a homogeneous subsurface the propagation operator is laterally invariant.

In equation 2.1 $P(x, y, z_0, \omega)$ is the wavefield at a certain point at the surface and $P(x_A, y_A, z', \omega)$ is the wavefield at a certain point in the subsurface. Note that the the downward propagation operator $W^+(x, x_A, y, y_A, \Delta z, \omega)$ can represent the propagation through any media (for instance complex) and can include internal multiple reflections. However, if multiple scattering is omitted and the propagation takes place in a homogeneous medium, the propagation operator can be represented analytically as it is derived in Appendix A in section A.6.

Similarly to the 3D case, the 2D integral is formulated as:

$$P(x_A, z', \omega) = \int_{-\infty}^{\infty} dx W^+(x, x_A, \Delta z, \omega) P(x, z_0, \omega). \quad (2.2)$$

However, in practical situations the wavefields are measured in a discrete fashion, where sources and receivers are placed only at certain x and y positions. Hence, the integrals in equations 2.1 and 2.2 become summations. Please note that in this context k just refers to an index rather than to the physical wave-number. The discrete sampling is defined with: $x_k = k\Delta x$ and $y_l = l\Delta y$, with k ranging from 1 to K and l from 1 to L . The values K and L are supposed to be sufficiently large numbers and Δx and Δy are sufficiently small. In Figure 2.1 propagation on discretised coordinates is illustrated.

For the sake of overview first the 2D case of wavefield extrapolation in terms of matrix multiplications will be derived. Hence, in Figure 2.1 for now y is constant for all points.

The discrete version of equation 2.2 is given by:

$$P(x_{k'}, z', \omega_n) = \sum_{k=1}^K \Delta x W^+(x_{k'}, k, \Delta z, \omega_n) P(x_k, z, \omega_n). \quad (2.3)$$

The propagation operator $W^+(x_{k'}, k, \Delta z, \omega_n)$ describes the propagation of a wavefield in point (x_k, z_0) to a point $(x_{k'}, z')$. The equations are valid for all discrete frequencies ω_n and can be solved independently, which provides a computational advantage. Please note that in equation 2.3 and for the remainder

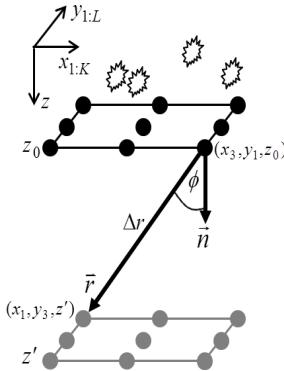


Figure 2.1: Schematic illustration of the downward propagation of the wavefield. The arrow indicates the downward propagation kernel, connecting point (x_3, y_1, z_0) at the surface, to point (x_1, y_3, z') in the subsurface. The angle between the direction of propagation \vec{r} and the downward pointing normal of the surface \vec{n} is denoted by Φ . Wavefield propagation can only be computed from the discrete surface locations (black dots) to the discrete subsurface locations (grey dots).

of this thesis the notation for lateral location is shortened for the sake of convenience from $(x_k, x_{k'}, y_l, y_{l'})$ to $(x_{k',k}, y_{l,l'})$. Equation 2.3 describes a wavefield that is recorded along the line $(x_{1:K}, z_0)$ and is propagated to a point $(x_{k'}, z')$. For the case that in Figure 2.1 the wavefield is only non-zero at location (x_3, z_0) , the summation in equation 2.3 has only one contribution: $P(x_1, z', \omega_n) = \Delta x W^+(x_{3,1}, \Delta z, \omega_n) P(x_3, z_0, \omega_n)$, which is simply a scalar multiplication. However, if all lateral locations at the surface contribute, then the wavefield in all points along $(x_{1:K}, z_0)$ is forward propagated to (x_1, z') and summed (cf. equation 2.3). This summation can also be interpreted as the inner product of two vectors (Figure 2.2). In this notation the frequency dependency is omitted for the sake of convenience.

$$\begin{array}{c|c}
 \boxed{\begin{array}{c} P \\ (x_1, z') \end{array}} & = \quad \boxed{\begin{array}{c|c|c|c}
 W^+ & W^+ & \dots & W^+ \\
 (x_{1,1}, \Delta z) & (x_{1,2}, \Delta z) & \dots & (x_{1,K}, \Delta z)
 \end{array}}
 \end{array}
 \begin{array}{c|c}
 \boxed{\begin{array}{c} P \\ (x_1, z_0) \end{array}} \\
 \boxed{\begin{array}{c} P \\ (x_2, z_0) \end{array}} \\
 \vdots \\
 \boxed{\begin{array}{c} P \\ (x_K, z_0) \end{array}}
 \end{array}$$

Figure 2.2: Summation of equation 2.3 as vector-vector multiplication.

To give the current set-up a practical meaning $P(x_{1:K}, z_0)$ is generated by sources

at the discrete surface locations. Therefore, in Figure 2.3 this vector is denoted as $S(x_{1:K}, z_0)$ with each element of this vector being the monochromatic source wavelet for the corresponding source. If for one source experiment just a single source is fired, only one element of the source vector $S(x_{1:K}, z_0)$ is non-zero. If all sources are fired in the same experiment, the source vector is completely filled. The wavefield at all lateral locations (regardless if they contribute or not) is then downward propagated to laterally distributed subsurface points. In the remainder of this thesis, these points are referred to as "grid-points" (see e.g. the gray dots in Figure 2.1 or Figure 2.5). Previously (Figure 2.2) the wavefield at the surface was propagated to a single grid-point. The propagation to many grid-points (Figure 2.3) is formulated by the propagation matrix $\mathbf{W}^+(z', z_0)$ (see also Berkhouit, 1982). Every k -th row of $\mathbf{W}^+(z', z_0)$ propagates the wavefield from all surface locations $(x_{1:K}, z_0)$ to one grid-point (x_k, z') , and every k -th column propagates one surface point (x_k, z_0) to all grid-points $(x_{1:K}, z')$.

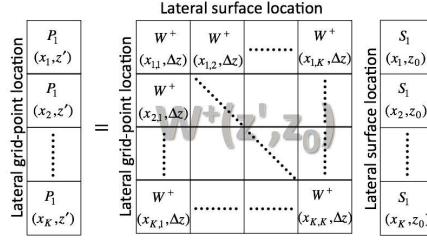


Figure 2.3: The propagation of a source experiment to all (K) grid-point locations via matrix-vector multiplication.

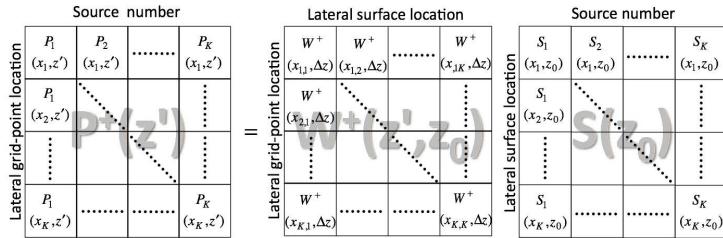


Figure 2.4: Downward propagation of the wavefields of K source experiments by matrix multiplication in the frequency domain. The source matrix consists of K source experiments $S_{1:K}$ (columns). For each experiment sources can be excited at K different lateral locations $x_{1:K}$ (rows).

However, traditional seismic acquisition consists of many source experiments where one source is fired at a time. These source experiments are stored in the so called source matrix, which is notified with the letter \mathbf{S} (see Figure 2.4). The

number of experiments is equal to the number of columns in $\mathbf{S}(z_0)$. Every row represents a lateral location. Please note that the wavefield propagation algorithm works regardless of how many source experiments are considered. However, for the remainder of this thesis a set-up is chosen which corresponds to a fixed spread acquisition. Here a receiver is placed at every possible location $x_{1:K}$. Furthermore, in total K source experiments are carried out, where the source of the k -th experiment is located at x_k . In that case only the main diagonal of $\mathbf{S}(z_0)$ is non-zero. If however, the whole first column of $\mathbf{S}(z_0)$ would be filled with non-zero entries, the corresponding physical set-up is that for this experiment sources at all possible lateral locations $x_{1:K}$ were fired in the same experiment. Please note that there is in fact a scientific field called 'source-blending' or 'simultaneous sources', where sources are not exactly fired at the same time, but so shortly after each other that their wavefields are overlapping. For further information see Berkhouw (2008), Berkhouw et al. (2009), Doulgeris et al. (2010, 2011) or Mahdad et al. (2011).

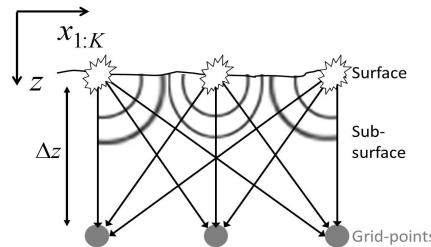


Figure 2.5: Downward propagation of three point-source wavefields to three grid-points.

In Figure 2.5 the propagation of three shot records to a deeper level (z') is illustrated. This propagation can be computed by the matrix multiplication which is denoted in Figure 2.4. Note that one column of propagation matrix $\mathbf{W}^+(z', z_0)$ represents a pressure field at z' due to a unit dipole at z_0 . If the source and grid-point locations are distributed along two parallel lines (2D) or along two plan-parallel planes (3D), then $\mathbf{W}^+(z', z_0)$ forms a Toeplitz matrix, where all elements along any diagonal are the same.

Of course in most field applications both sources and receivers are placed at the surface and the reflected wavefield from the subsurface is detected. As mentioned previously, for the sake of convenience it is assumed that sources and receivers share the same locations. In a homogeneous medium equations 2.1 and 2.2 give an adequate description of sound propagation. However, if a reflecting boundary is reached, two mechanism enter the propagation model: reflection and transmission (Figure 2.6). For the sake of convenience it is assumed that the reflector is at depth z' . The reflectivity matrix $\mathbf{R}(z')$ turns the downward propagating wavefield $\mathbf{P}^+(z') = \mathbf{W}^+(z', z_0)\mathbf{S}(z_0)$ into an upward propagating wavefield $\mathbf{P}(z')^- = \mathbf{R}(z')\mathbf{P}(z')^+$. Again, this process consists of spatial convolutions, which are computed by matrix-multiplications, as is visualised in Figure

2.7. The strength of the upward propagating wavefield depends on the reflectivity of the boundary. If the reflecting boundary is locally reacting, then $\mathbf{R}(z')$ is a diagonal matrix, holding the reflection coefficients for each grid point at z' .

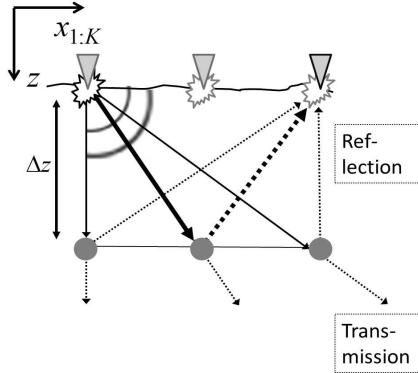


Figure 2.6: At a reflecting boundary transmission and reflection occurs. For the sake of convenience only the ray paths for one source and one receiver (black) is denoted.

In Figure 2.6 for the source firing at x_1 and the receiver at x_3 , the wavefield is reflected at grid-point (x_2, z') indicated by the bold wave-front-normal. For a locally reacting $\mathbf{R}(z')$ no neighbouring grid-points are taken into account for this reflection. If also the off-diagonal elements are non-zero the reflectivity is angle dependent. In the notation of the reflectivity matrix this translates to the case that the wavefield of neighbouring grid-points contributes to the reflection at a certain grid-point. As an example, if $R(x_{1,2})$ is non-zero then for the reflection at grid-point (x_1, z') also the incoming wavefield at grid-point (x_2, z') contributes. The strength of this contribution is scaled by the scalar $R(x_{1,2})$. If $R(x_{2,1})$ is non-zero then the wavefield that leaves grid-point (x_2, z') also gets a contribution from the wavefield that leaves grid-point (x_1, z') . Transforming one row or column of the reflectivity matrix from the spatial domain to the wave-number domain generates the so called 'reflectivity operator', which provides information about the angle-dependency of the reflectivity in the subsurface. For a locally reacting reflector only one entry in a row or column of the reflectivity matrix would be non-zero, which would lead to a flat spectrum in the reflectivity operator and, hence, to an angle-independent reflectivity. For more information about the reflectivity matrix see Berkhouit (1987), de Bruin et al. (1990) and Doulgeris et al. (2009). Please also consult Appendix B for more information about the wave-character of the WRW formulation. In the current model there is only one reflector and hence, the downward transmitted energy is not recorded by the receivers at the surface.

However, the reflected wavefield is then upward propagated to the receivers at the surface. This set-up still complies with the restrictions of the Rayleigh integral.

The volume or half-space into which the up-going wavefield is propagating is above the reflector. The reflector itself acts as a series of secondary sources and these sources are outside the volume. The propagation from z' to z_0 is described by $\mathbf{W}^-(z_0, z')$ (see Figure 2.7), where the “-” denotes negative z direction. Every row of $\mathbf{W}^-(z_0, z')$ propagates the wavefield from all grid-points to one lateral location at the surface and every column propagates the wavefield from one grid-point to all surface locations. In the case of a flat reflector and a constant velocity within the layer through which the wavefield is propagated, $\mathbf{W}^-(z_0, z')$ is the transpose of $\mathbf{W}^+(z', z_0)$. In that case the propagation of a wavefield from the surface to one grid-point has the same effect as the propagation of a wavefield from the sub-surface to one surface-point. However, in practice this is usually not the case. The detection of the up-going wavefield is denoted by $\mathbf{D}(z_0)$. The number of rows represents the number of received experiments. The number of columns represent the different lateral locations. As mentioned previously, here the fixed spread acquisition geometry is used. In that case the detector matrix has the same number of rows as columns and only the main diagonal is non-zero (unit matrix, if the wavelet is not changed). However, the algorithm still works if a different number of receivers is chosen than the number of lateral surface locations. Furthermore, also detectors might be grouped together, such that the actually received signal in the computer is a summation of the signal of many receivers. This is sometimes done to enhance the signal-to-noise ratio. In that case more than only one element per row would be non-zero in $\mathbf{D}(z_0)$. Note that similar to the source-wavelet there is also a receiver-wavelet, which describes the receiver response. In conclusion, the difference of the actual up-going wavefield at the surface and the stored wavefield in the computer, in terms of spectral content or geometry, is expressed by the detector matrix.

Finally, multiplying all matrices in Figure 2.7 describes a complete seismic acquisition in 2D. The result is the so called data-matrix $\mathbf{P}(z_0)$, where each element $P(x_i, x_j, z_0, \omega_n)$ describes the monochromatic wavefield that was generated by source j and received by detector i at the surface. Please note that receivers and sources are only placed at discrete locations at the surface and a limited number of grid-points in the subsurface is used. In Figure 2.6 the bold ray path is the physical one following the Fermat’s principle of shortest travel time. However, obviously not for all source-detector combinations there is a grid-point available for a physical reflection. For some cases the actual reflection would occur laterally in between the grid-points. Also this energy is explained by the WRW model provided that the spatial sampling of these grid-points is dense enough. For more insight see Appendix B, where the wave character of the WRW model is illustrated for a simple example. Finally, it has to be mentioned that although only one reflectivity level is described here, the data matrix can still contain the reflection response from many locations in the subsurface. This can be seen as the summation of individual reflection responses.

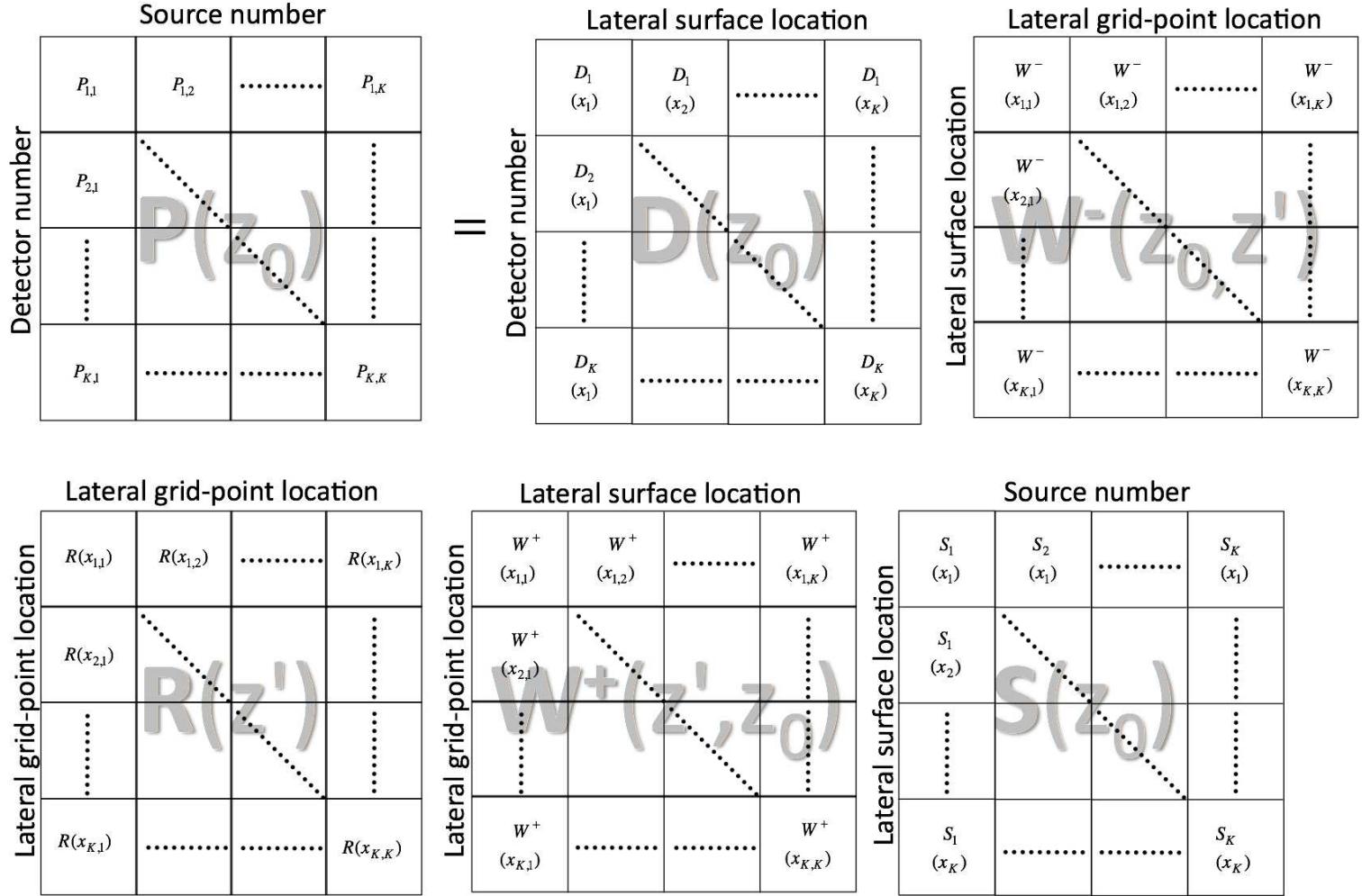


Figure 2.7: The formulation of data creation from a single reflector at z' with the WRW model in 2D:
 $\mathbf{P}(z_0) = \mathbf{D}(z_0)\mathbf{W}^-(z_0, z')\mathbf{R}(z')\mathbf{W}^+(z', z_0)\mathbf{S}(z_0)$.

2.2 Extension to three dimensions

Wavefield extrapolation in 3D can be discretised similarly to the 2D case. The summation corresponding to integral-equation 2.1 can be denoted as:

$$P(x_{k'}, y_{l'}, z', \omega_n) = \sum_{l=1}^L \sum_{k=1}^K W^+(x_{k',k}, y_{l',k}, \Delta z, \omega_n) P(x_k, y_l, z, \omega_n) \Delta x \Delta y, \quad (2.4)$$

which describes the propagation of the wavefield of all discrete points along the 2D plane at z_0 to the grid-point at $(x_{k'}, y_{l'}, z')$ for one temporal frequency. This double summation can be formulated as well as a vector-vector multiplication. For an easier understanding first propagation of a line in the x -direction to one grid-point is formulated (similar to the 2D case) and is displayed in Figure 2.8. Please note that in 3D no line-source-assumption is made (compare equations A.58 and A.59 with A.63 and A.64).

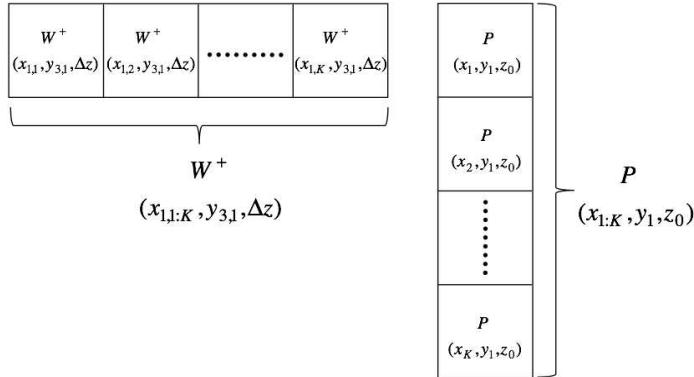


Figure 2.8: 3D downward extrapolation of a 1D line in the x -direction $(x_{1:K}, y_1, z_0)$ to a grid-point at (x_1, y_3, z') in the notation of a vector-vector multiplication.

In Figure 2.1 the set-up of Figure 2.8 corresponds to the propagation and summation of the wavefield from the first row of black dots $(x_{1:3}, y_1, z_0)$ to the gray dot at (x'_1, y'_3, z') . Usually, the wavefield is propagated to another y -coordinate. Hence, all entries of the vector $W(x_{1:K}, y_{3,1}, \Delta z)$ in Figure 2.8 are indicated with $y_{3,1}$ because the wavefield is propagated from coordinate y_1 to y_3 . Only the wavefield at the location y_1 was considered, which is also indicated in the notation of the wavefield vector $P(x_{1:K}, y_1, z_0)$.

In a 3D acquisition of course the wavefield from all y locations contributes to the wavefield in the grid-point. Hence, in Figure 2.1 the wavefield at all black dots (all x - and y -locations) is propagated to one grid-point. The corresponding vector-vector multiplication is displayed in Figure 2.9 and represents equation 2.4.

By comparing Figure 2.8 and 2.9 the reader realises that the first subset of the 3D wavefield propagation vector $W^+(x_{1:1:K}, y_{3:1:L}, \Delta z)$ is actually the 2D case $W^+(x_{1:1:K}, y_{3,1}, \Delta z)$. Hence, the 3D wavefield propagation vector simply consists of a concatenation of 2D vectors each propagating wavefields from a different y -location to the grid-point. Similarly, the 3D wavefield vector $P(x_{1:K}, y_{1:L})$ consists of concatenated subsets for which each the y -coordinate is constant. Hence, each subset represents the wavefield at $x_{1:K}$ for a constant y - coordinate (see also Kinneging et al., 1989).

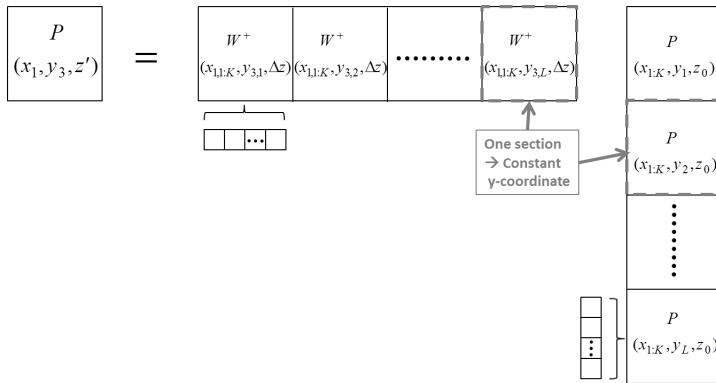


Figure 2.9: 3D downward extrapolation of a 2D plane in the x, y -direction ($x_{1:K}, y_{1:L}, z_0$) to a grid-point at (x_1, y_3, z') in the notation of vector-vector multiplication.

In Figure 2.10 the full 3D data creation corresponding to the WRW model for a single flat reflector at depth z' is visualised. Obviously, the formulation is exactly like in 2D, which makes the translation of a 2D to a 3D implementation very simple. Please note that in this formulation there are K times L possible discrete surface locations at which sources or receivers can be placed. This is denoted by the maximum detector number D_{LK} and the maximum source number S_{LK} . Please also note that by conventionally multiplying the matrices from the right to the left causality of wavefield propagation is obeyed intrinsically: The wavefield is excited by the sources $\mathbf{S}(z_0)$, downward propagated $\mathbf{W}^+(z', z_0)$, reflected $\mathbf{R}(z')$, upward propagated $\mathbf{W}^-(z_0, z')$ and finally received by the detectors $\mathbf{D}(z_0)$. First, the multiplication of the last two matrices is investigated: $\mathbf{W}^+(z', z_0)\mathbf{S}(z_0)$. The vector-vector multiplication of the first row of $\mathbf{W}^+(z', z_0)$ with the first column of $\mathbf{S}(z_0)$ represents the case of Figure 2.9. Just like in 2D the general wavefield at the surface is replaced by sources at the discrete lateral locations at the surface. The first column of $\mathbf{S}(z_0)$ represents, therefore, a single source experiment. Which element of the first row is non-zero determines where the source for this experiment is placed. Again, more than one element being non-zero corresponds to the case that several sources are fired within the recording time of this experiment.

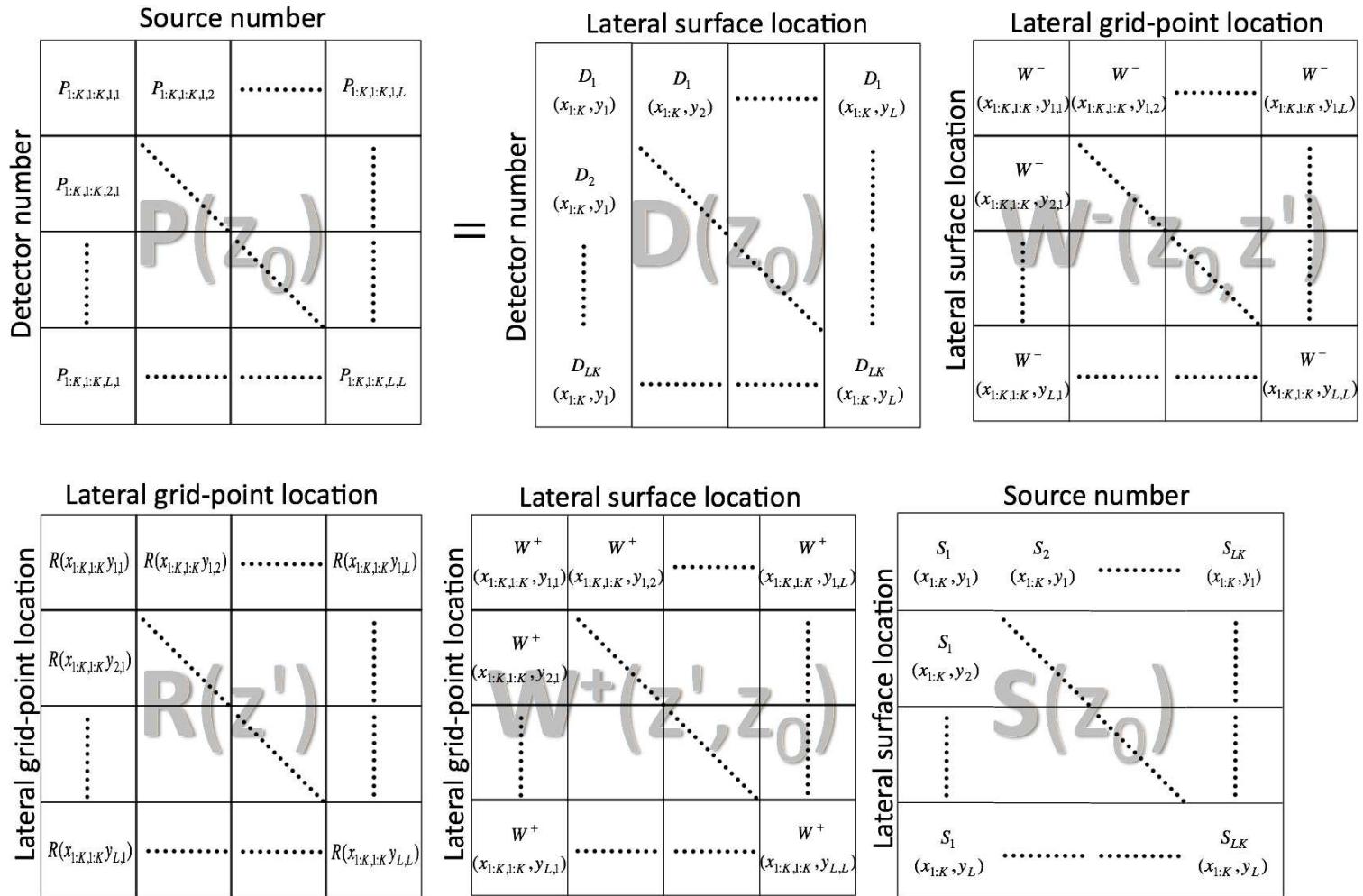


Figure 2.10: The formulation of data creation from a single reflector at z' with the WRW model in 3D: $\mathbf{P}(z_0) = \mathbf{D}(z_0)\mathbf{W}^-(z_0, z')\mathbf{R}(z')\mathbf{W}^+(z', z_0)\mathbf{S}(z_0)$.

The ordinate of the source matrix $\mathbf{S}(z_0)$ represents different lateral locations at the surface. Every section contains all x -locations ($x_{1:K}$) for a constant y -location. The abscissa of $\mathbf{S}(z_0)$ simply represents the number of source experiments. As for the 2D case there is no restriction on how many source experiments are executed in total, or how the sources are distributed along the surface in both lateral directions, or even how many sources per source experiment are fired. However, for the sake of convenience also for the 3D case a fixed spread acquisition geometry is chosen. Hence, receivers are evenly distributed along x - and y -direction and at each receiver-location one source is fired per source-experiment. First, the y -coordinate of the source is constant and the source is placed at different x -locations for each experiment. Then the y -coordinate is changed and the process is repeated (K times L experiments). Such a set-up results in a source matrix $\mathbf{S}(z_0)$ that has only non-zero values along its main diagonal. Please note that for the sake of convenience the number of x - and y -locations is assumed to be the same ($L = K$).

Investigating Figure 2.10 again the reader realises that the propagation matrix $\mathbf{W}^+(z', z_0)$ has the following properties:

- It propagates the wavefield from all surface locations $(x_{1:K}, y_{1:L}, z_0)$ to all grid-points $(x_{1:K}, y_{1:L}, z')$.
- Every row propagates the wavefield from all surface locations to a single grid-point and every column represents the propagation from one surface point to all grid-points.
- Every submatrix contains K rows and K columns, which is denoted by the two x -indices ranging from 1 to K .
- Every submatrix of $\mathbf{W}^+(z', z_0)$ simply represents the 2D wavefield propagation matrix for a fixed y -location of the surface points and the grid-points.
- Along the abscissa of $\mathbf{W}^+(z', z_0)$ every different submatrix represents propagation from a different y -location at the surface. Along the ordinate every different submatrix represents propagation to a different y -location at the grid-point level.

For instance, the submatrix in the first row and second column: $W^+(x_{1:K, 1:K}, y_{1,2})$ describes the propagation of the wavefield from all x -locations with the constant y -location y_2 at the surface to all x -locations in the subsurface with y -location y_1 . The reflectivity matrix in Figure 2.10 transforms a down-going wavefield into an up-going wavefield. In the case of a single angle-independent reflector at depth level z' , the reflectivity $\mathbf{R}(z')$ is a diagonal matrix. Like in the 2D case, $\mathbf{R}(z')$ being a diagonal matrix represents a locally reacting reflector. Also the reflectivity matrix is sub-divided in submatrices. Each submatrix contains a complete 2D reflectivity matrix. Like in the 2D case the angle-dependency is represented by the off-diagonal elements being non-zero. Hence, if the reflector is not reacting

locally in 3D then for the reflection at a certain grid-point also the neighbouring grid-points in x -direction are taken into account. In that case not only the main diagonal of $\mathbf{R}(z')$ is filled with information but also some neighbouring diagonals. However, the same applies in the y -direction. If neighbours in the y -direction contribute to a reflection at a certain grid-point, then diagonals appear in the reflectivity matrix one or two submatrices to the right or below the main diagonal. In more complex data scenarios $\mathbf{R}(z')$ is filled with a couple of narrow bands of diagonals. For instance, if the second submatrix $R(x_{1:K,1:K}, y_{1,2})$ contains non-zero elements, this represents the case that for reflections at y_1 also incoming wavefields at y_2 are contributing as well. And if elements of the submatrix $R(x_{1:K,1:K}, y_{2,1})$ contain non-zero elements, the reflected wavefield at y_2 also gets contributions from wavefields which travel from y_1 . For further information on the reflectivity matrix see de Bruin et al. (1990).

The matrix $\mathbf{W}^-(z_0, z')$ propagates the wavefield from all grid-points to all surface locations. Every row propagates the wavefield of all grid-point locations to one surface location and every column propagates the wavefield from one grid-point to all surface locations.

Similar to the source matrix the detector matrix $\mathbf{D}(z_0)$ is sub-divided in submatrices. Along the abscissa the lateral location changes. Within each submatrix the y -coordinate of a detector is constant and only varying along the x -coordinate. The ordinate of $\mathbf{D}(z_0)$ notifies the receiver number. As for the 2D case if for example one complete row has non-zero values the corresponding physical situation is that the detected wavefield of all receivers is summed and stored in one receiver trace. However, using here the previously explained fixed spread geometry, $\mathbf{D}(z_0)$ is a diagonal matrix.

Finally, each element of the data matrix $\mathbf{P}(z_0)$ is a complex-valued number representing one frequency component of a recorded trace. The element $P_{ij}(z_0)$ denotes the complex valued frequency component of a trace that was generated by a source on location j and received by a detector at location i . Different rows of the data matrix correspond to different receiver gathers, while different columns represent different shot records (source gathers). Any diagonal of this data matrix corresponds to a monochromatic common-offset gather. Due to reciprocity the data matrix is symmetrical along the zero-offset (main-) diagonal. The source wavefield is propagated to all possible depth levels. If the seismic impedance is changing at this level, the reflectivity matrix will not be empty and a reflection occurs. The reflected wavefield is propagated upwards to the receivers.

2.3 Multi-reflector data model

However, until now it was assumed, that there is exactly one reflector in the subsurface to which the data was downward propagated, reflected and upward propagated. In reality there will of course be many more reflectors. Therefore, with the WRW model the reflection responses for many depth levels are summed

together. At every depth level a potential reflector contributes to the generation of the total data. To be more specific, the source wavefield is downward propagated to the first depth level. If there is actually a reflector then a reflection takes place and the reflected wavefield is upward propagated and detected. In the absence of a reflector at this depth level the reflectivity matrix will be empty and the upward propagated wavefield is zero. This is repeated for all M depth levels and the sum of all contributions results in the total reflection data. Note that the formulation of the WRW model for 2D and 3D is exactly the same:

$$\mathbf{P}(z_0) = \mathbf{D}(z_0) \sum_{m=1}^M [\mathbf{W}^-(z_0, m\Delta z) \mathbf{R}(m\Delta z) \mathbf{W}^+(m\Delta z, z_0)] \mathbf{S}(z_0). \quad (2.5)$$

Hereby, the step-size in z direction (Δz) needs to be very small (e.g. 5 meter), because all the summations like equation 2.3 or 2.4 are just approximations of their counterparts in integral form (compare for instance equations 2.2 and 2.3). For large Δz this approximation is not valid any more. Then the WRW model would be ineffective as it would miss reflectivity contrasts in the subsurface. The advantage of arranging seismic data in matrices is that complex operations like extrapolation, multi-shift cross-correlation and convolution (in time) can be expressed by simple matrix multiplications for each frequency. The process of data creation with the WRW model (equation 2.5) is visualised in Figure 2.11a.

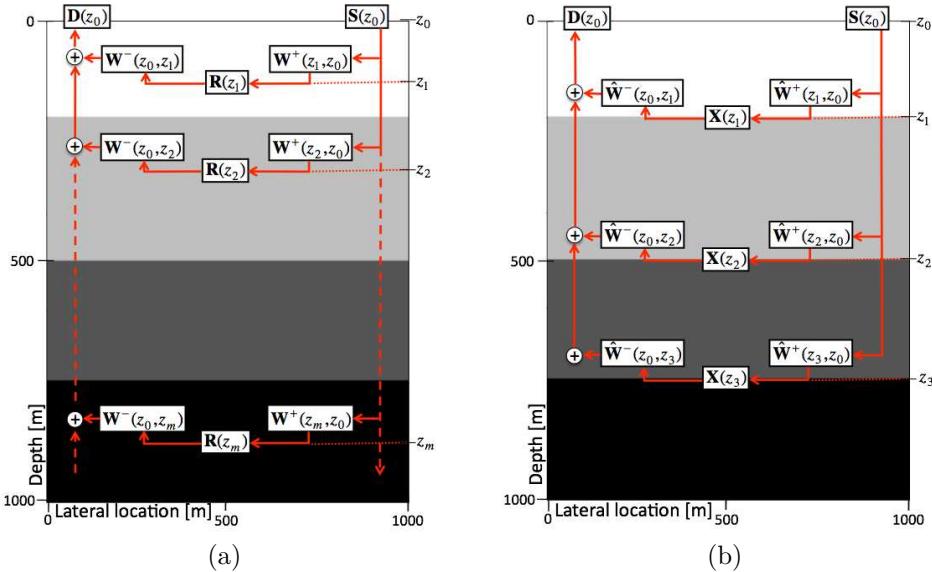


Figure 2.11: Concept for seismic data creation using (a) the WRW model and (b) the double focal transformation. For the double focal transformation three focal operators were used, each perfectly representing the propagation to one of the three layers.

It has to be mentioned that the WRW model contains neither transmission effects nor direct waves or surface waves. On the other hand, everything that is not explained by the propagation operators maps to the reflectivity operator. Hence, the energy is not lost and in principle can be restored.

In short, the WRW model is a completely reversible process that can effectively describe reflection and diffraction events but is less suited to describe direct/surface waves and transmission effects.

In Figure 2.12a a simple 2D subsurface model is displayed. A source is fired at the surface and its wavefield is propagating through the model (denoted by the wave fronts and ray paths) and reflected at the discontinuities in the model. In Figure 2.12b the corresponding 2D shot record is displayed.

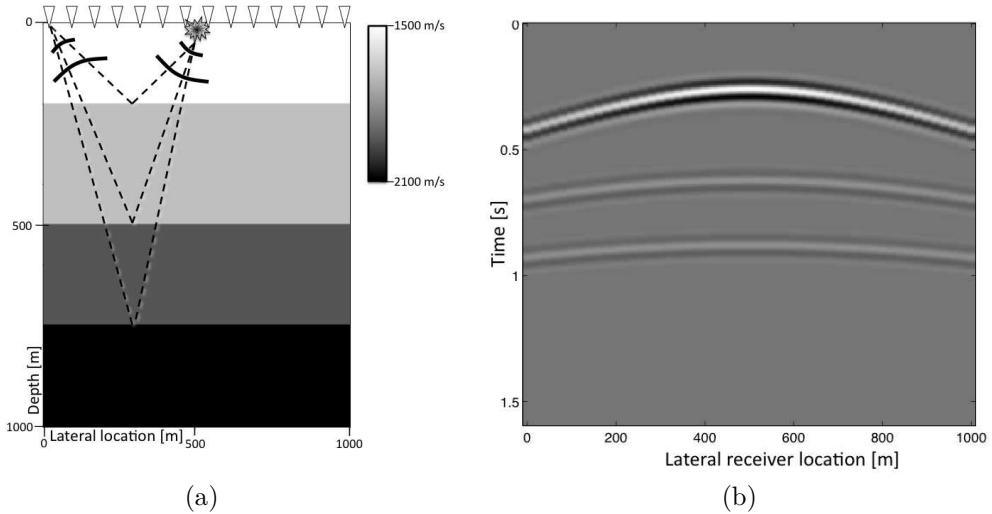


Figure 2.12: Simple three layer seismic model (a). The seismic velocity increases from bright to dark colors. The star denotes a seismic source and the triangles denote seismic receivers. In (b) the corresponding seismic shot record is displayed.

Usually, in seismic experiments more than only one shot is fired. For the investigations in this thesis fixed spread acquisition designs will be used. Hereby, it is assumed that at every receiver location (or very close to it) also a source is fired. For the 2D case the measurements can be stored in a three-dimensional data tensor (Figure 2.13a). Please note that in the example the acquisition consists of 51 source gathers, each containing 51 receivers. However, for the sake of convenience only three source gathers are displayed.

To derive the data matrix for one specific frequency (Figure 2.13b) the data tensor $\mathbf{p}(z_0)$ is transformed to the temporal frequency domain. Each monochromatic subsection (frequency slice) is called a data matrix $\mathbf{P}(z_0)$. This data matrix is of course the same as in Figure 2.7. The data creation of this set-up can be described by the WRW model (equation 2.5).

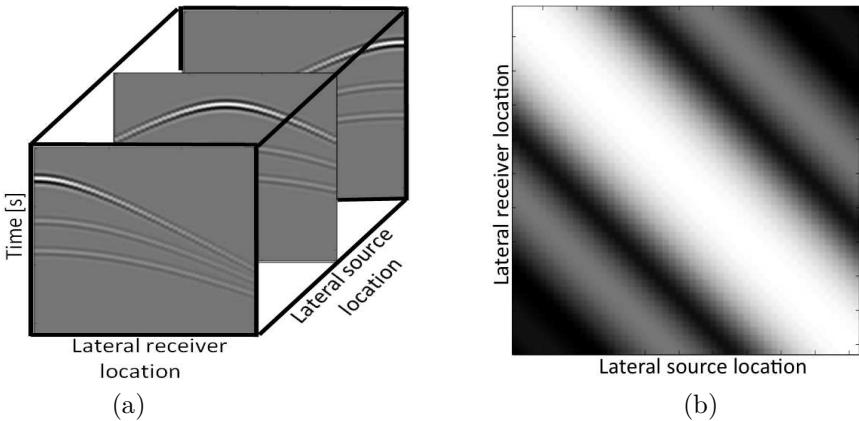


Figure 2.13: The three-dimensional data tensor (a) in the time domain $\mathbf{p}(z_0)$ and the corresponding data matrix $\mathbf{P}(z_0)$ for 15Hz (b).

2.4 From the WRW model to the concept of double focal transformation

2.4.1 The multi-level implementation

The WRW model allows describing the reflectivity in the subsurface on a dense grid of depth levels. However, this is costly. For M depth levels the same number of propagation operators have to be generated. Additionally, in order to derive the true reflectivity for every depth level, quite some knowledge of the subsurface has to be available (e.g. a reliable velocity model). However, reconstruction is always one of the first preprocessing steps. At this stage a detailed velocity model of the subsurface is usually not yet available.

Therefore, the double focal transformation estimates the reflectivity only for a limited number of depth levels. Data creation with the model of double focal transformation is formulated as:

$$\mathbf{P}(z_0) = \hat{\mathbf{D}}(z_0) \sum_{n=1}^N [\hat{\mathbf{W}}^-(z_0, z_n) \mathbf{X}(z_n) \hat{\mathbf{W}}^+(z_n, z_0)] \hat{\mathbf{S}}(z_0) \quad (2.6)$$

and is visualised in Figure 2.11b. The notation is slightly changed. The hatted symbols indicate that only approximated operators are used, i.e. based on an approximated velocity model. Furthermore, instead of using M depth levels with a distance of Δz , only a subset of N depth levels is used. This subset consists of few depth levels distributed over the whole depth range. The N depth levels can also be chosen strategically in order to coincide with prominent reflectors in the data. Please note, since *two* focal operators are used, one propagating energy from the surface to the subsurface and one propagating energy from the subsurface to

the surface, there will be essentially two focusing steps later on. This explains the name *double* focal transformation.

The major difference between equation 2.5 and equation 2.6, or between Figure 2.11a and Figure 2.11b is, that usually $N \ll M$. The consequence of using fewer depth levels and approximate propagation operators $\hat{\mathbf{W}}^+(z_n, z_0)$ is that seismic energy will not necessarily be focused at time zero, which was assumed to be the case in the WRW model. Therefore, the reflectivity matrices $\mathbf{R}(z_m)$ have been replaced by focal domains $\mathbf{X}(z_n)$, which have to be investigated for a time window around time zero. Thus, $\mathbf{X}(z_n)$ represents a frequency slice of a whole inverse or backward propagated data set.

Note that neither the WRW model, nor the concept of double focal transformation, effectively represents direct waves. However, the aim of this thesis is to introduce effective ways for data reconstruction. Linear events in seismic data can already be relatively well reconstructed by several different methods, like Fourier (Zwartjes and Sacchi, 2007; Zwartjes and Gisolf, 2007; Naghizadeh, 2012), linear Radon (Trad et al., 2003; Wang et al., 2010), and curvelet (Hennenfent et al., 2010; Naghizadeh and Sacchi, 2010) methods. More challenging for seismic data reconstruction are complex and strongly curved reflections and diffractions, which are completely covered by the models introduced in equations 2.5 and 2.6 and in Figure 2.11.

Equation 2.6 defines the double focal transformation for the general case, where the source matrix $\hat{\mathbf{S}}(z_0)$ contains amplitude and phase information of the source wavelet and defines the arrangement of sources or source arrays. However, the source wavelet might not be known precisely. Therefore, in the following formulation it will not be separated from the data. The source matrix $\mathbf{S}(z_0)$ is then simply assumed to be an identity matrix and, hence, can be skipped in the formulation. The consequence is, that the source wavelet is mapped to the focal domain. Having a model space that not only contains the subsurface reflectivity and phase factors, but also the wavelet of the data, decreases the sparseness of the data representation. This has no negative consequences for the noise separation, because also the aliasing noise will still contain the source wavelet. Therefore, the relative strength of focusing the data compared to focusing the aliasing noise is not degraded.

Also the detector matrix $\hat{\mathbf{D}}(z_0)$ is set to identity, since only single receivers are used throughout this thesis. The receiver transfer function is also mapped to the focal domain. The simplified version of the model of the double focal transformation is defined as:

$$\mathbf{P}(z_0) = \sum_{n=1}^N [\hat{\mathbf{W}}^-(z_0, z_n) \mathbf{X}(z_n) \hat{\mathbf{W}}^+(z_n, z_0)]. \quad (2.7)$$

This formulation is equivalent to equation 2.6 with the source- and receiver wavelet lumped into the focal domain. The simplification can be justified by the fact that in the case of field data the focal operators represent the true data only

approximately. Therefore, a perfect focusing can not be achieved anyway. Furthermore, the accurate source and detector wavelets are usually unknown, so they also can be skipped in the formulation to save computational cost and memory. Hence, all inaccuracies are mapped to the focal domain, which makes this process reversible and allows a complete restoration of the data. Please note that by skipping the source and receiver matrices, the axis labels of the data changes from (receiver gather, source gather) to (lateral receiver location, lateral source location).

In comparison:

- The WRW model allows to retrieve the reflectivity for every chosen depth level based on a reliable velocity model. The reflectivity for every point in the subsurface includes angle dependency but has no wavelet. The data is represented by M reflectivity matrices.
- The double focal transformation allows to retrieve focal domains. These focal domains contain the backwards or inverse propagated data to strategic depth levels. The data is represented by N focal domains, each having in principle the same size as the data. The quality of the focusing depends on the velocity model from which the focal operators were generated. The wavelet is not removed from the data.

In practical situations a precise velocity model of the subsurface is rarely available. However, for the purpose of demonstration it is assumed that in this special case a very precise velocity model is known (Figure 2.11b). Therefore, the three strategic depth levels in the velocity model coincide with the three true reflectors. Hence, in this particular case each reflector is focused in the corresponding focal domain $\mathbf{X}(z_n)$ at time zero.

In Figure 2.14a the focal domain corresponding to the first virtual depth level is displayed. All three focal domains look very similar as can be observed in Figure 2.16. To emphasise the capability of the focal transformation for sparse data representation, note that the whole set of hyperbolic reflection events corresponding to the first reflector in Figure 2.13a is focused to an event along a very localised area around the diagonal in the focal domain around time zero (denoted by the red line in Figure 2.14).

Equation 2.7 is visualised for the dominant frequency (15Hz) in Figure 2.15. The propagation of the data from the virtual depth levels to the surface can be computed by simple matrix multiplications with the approximate propagation operators in the frequency domain. The fading out of the focal domains in the corners is related to edge effects. This leads to a decrease in focusing of the data, which reduces the reconstruction capability at the outer locations of the fixed spread acquisition.

The time-equivalent to equation 2.7 is visualised in Figure 2.16, where the matrix multiplications of Figure 2.15 translate to multi-shift convolutions. Note that multi-shift convolution is an operation closely related to a 2D convolution.

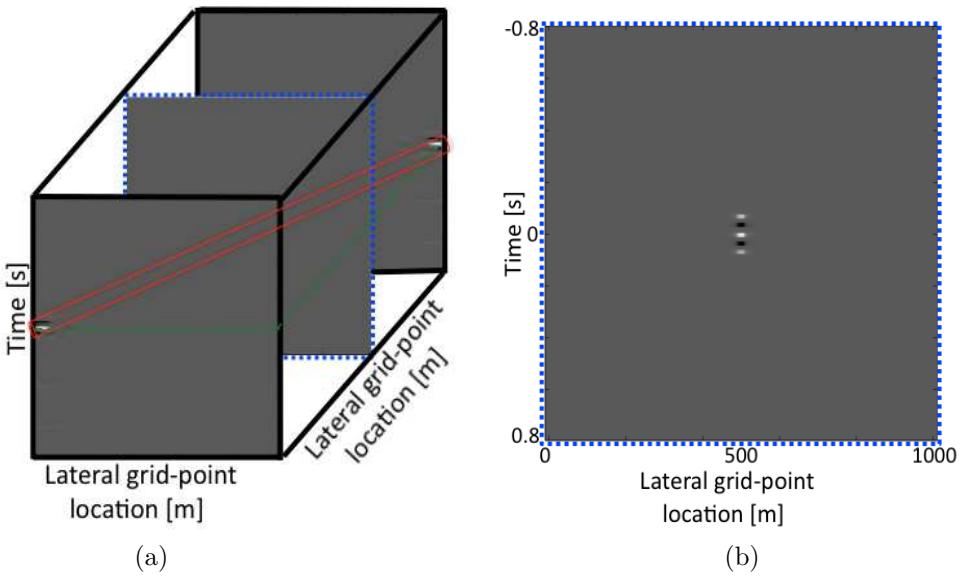


Figure 2.14: (a) The focal domain corresponding to the first strategic depth level as a 3D tensor. (b) The middle section of the focal domain in (a).

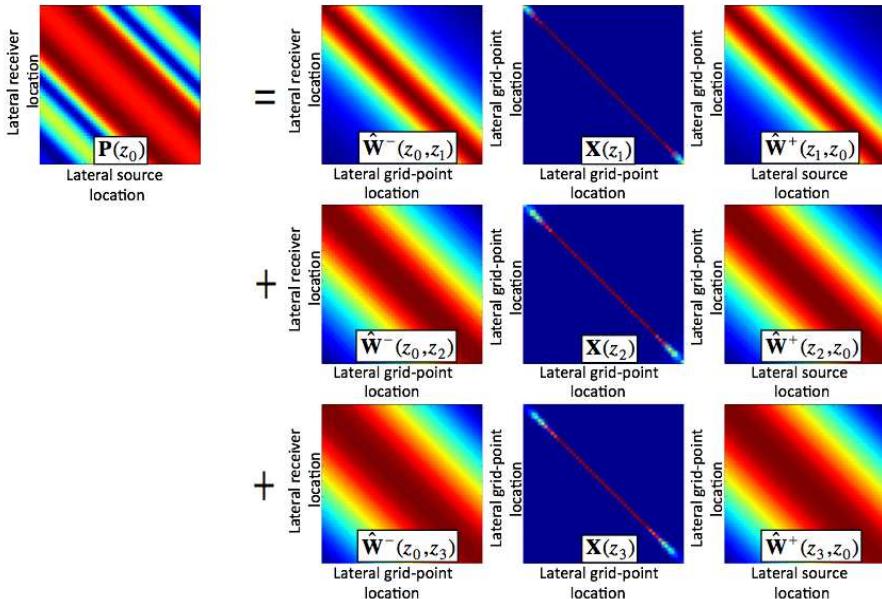


Figure 2.15: Graphical representation of equation 2.7 for the dominant frequency (15Hz) for the hypothetical case that the true velocity model of the subsurface is known and the subsurface only contains three reflections, where multiples are neglected.

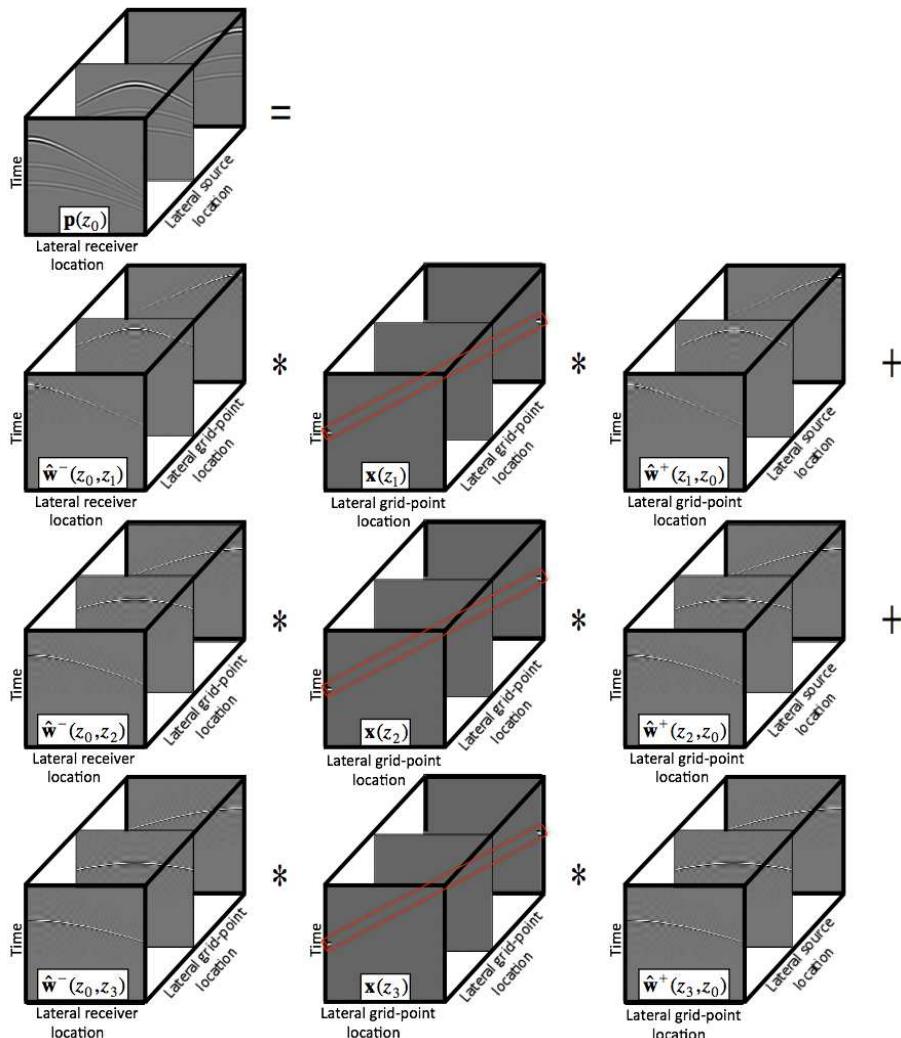


Figure 2.16: Creation of seismic data with the concept of double focal transformation. The stars denote multi-shift-convolutions.

2.4.2 The single-level implementation

A special case of the double focal transformation is the single-layer implementation. The model of data creation is visualised in Figure 2.17.

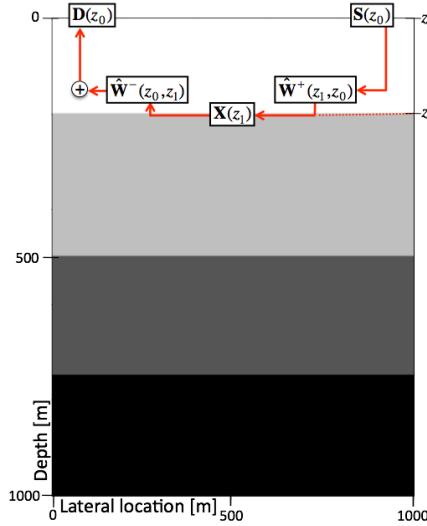


Figure 2.17: Concept for seismic data creation using the double focal transformation for one virtual depth level.

Here the data is focused to one depth level only. Consequently, only one reflection event is compressed. This is visualised in Figure 2.18. The other two reflectors are still spatially extended. Basically one could understand this set-up as a seismic acquisition inversely propagated to the first reflector. Hence, sources and receivers are virtually placed at the first reflector. For the first reflector the sources are fired, immediately reflected and received, which leads to a reflection event at time zero in Figure 2.18.

The corresponding equation in the frequency domain is formulated as:

$$\mathbf{P}(z_0) = \hat{\mathbf{W}}^-(z_0, z_1) \mathbf{X}(z_1) \hat{\mathbf{W}}^+(z_1, z_0) \quad (2.8)$$

and is visualised in Figure 2.19.

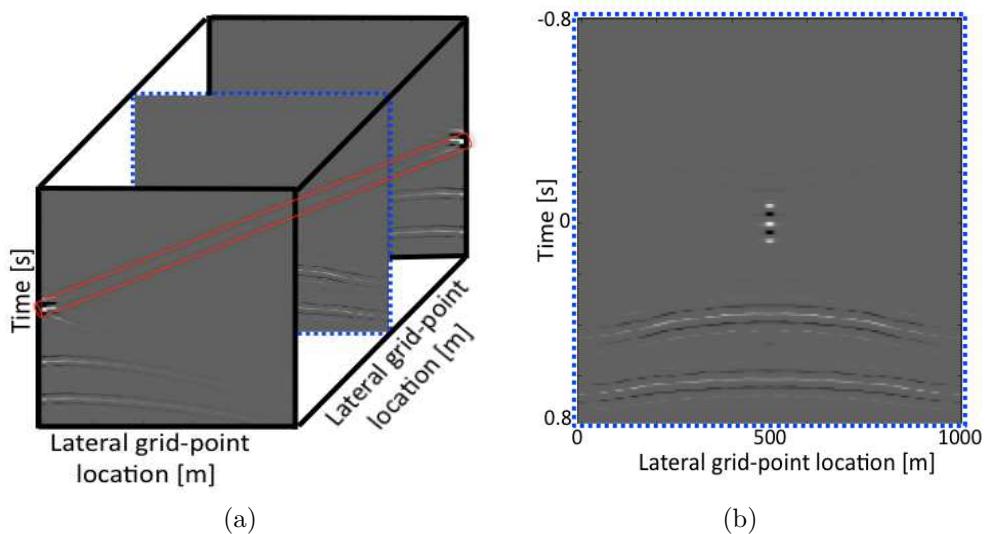


Figure 2.18: (a) The focal domain corresponding to the first strategic depth level as a 3D tensor for the single-level double focal transformation. (b) The middle section of the focal domain in (a).

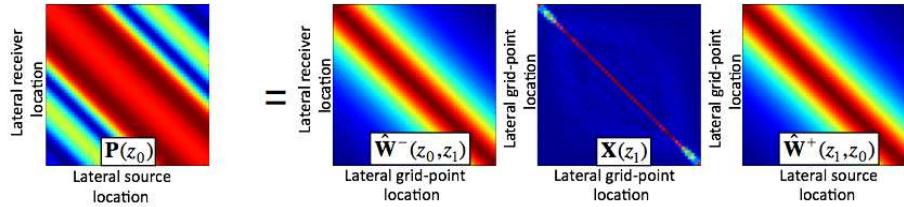


Figure 2.19: Graphical expression to equation 2.8 for the dominant frequency (15Hz) for the hypothetical case that the true velocity model of the first layer is known. The main diagonal of $\mathbf{X}(z_1)$ is relatively strong, because of the strong focusing of the first reflector. However, the off-diagonal elements are not equal to zero but are just relatively small, because they represent the spread-out energy of the second and third event in the focal domain.

Equation 2.8 can be solved for the focal domain as is depicted in Chapter 3. Note that the monochromatic focal domain matrix $\mathbf{X}(z_1)$ in Figure 2.18 has also all off-diagonal elements being non-zero. The time domain expression is visualised in Figure 2.20. Again, matrix multiplications in the frequency domain correspond to multi-shift convolutions in the time domain.

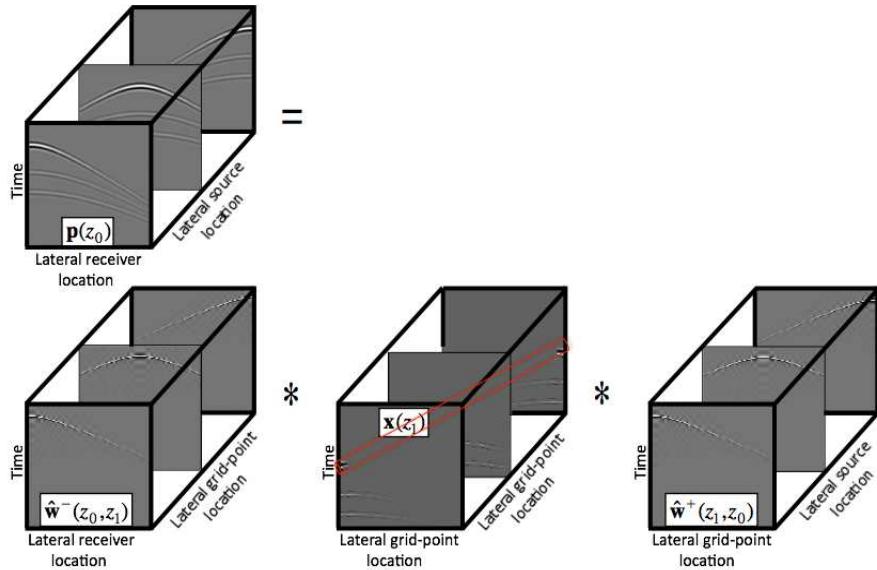


Figure 2.20: Time domain representation related to equation 2.8 and Figure 2.19.

The data is focused to a lesser extent than in the multi-level case. For the multi-level case three sets of hyperbolic reflection events could be represented by three linear events in the focal domain. In the single-level implementation only the first reflection event is compressed. However, equation 2.8 has fewer factors, which reduces the computational cost and the ambiguity of the solution.

Chapter 3

Implementation of the focal reconstruction

In this chapter several ways are illustrated to utilise the double focal transformation as a tool for seismic data reconstruction.

3.1 The forward double focal transformation by correlation

This section builds seamlessly on the results the previous chapter. Therefore, the reader is kindly ask to view Figure 2.20 again, which visualises the data creation with the double focal transformation on a 2D data example. The corresponding equation reads:

$$\mathbf{p}(z_0) = \mathbf{w}(z_0, z_1) * \mathbf{x}(z_1) * \mathbf{w}(z_1, z_0), \quad (3.1)$$

where the $*$ represent multi-shift convolutions of three-dimensional tensors. Note that the implementation for 2D and 3D data is similar.

The convolution in space and time of the focal domain with the focal operators can be reformulated as a matrix vector equation:

$$\vec{p}(z_0) = \mathbf{L}(z_0, z_1) \vec{x}(z_1). \quad (3.2)$$

For the sake of convenience the depth-indication will be skipped from the formulation. Until notified otherwise, \vec{p} is the data at the surface z_0 and \vec{x} the focal domain corresponding to depth level z_1 . Equation 3.2 then becomes:

$$\vec{p} = \mathbf{L}\vec{x}. \quad (3.3)$$

Many transformation-based reconstruction methods can be represented by this kind of equation such as Fourier transformation (Zwartjes and Sacchi, 2007),

Radon transformation (Kabir and Verschuur, 1995) or curvelet transformation (Hennenfent et al., 2010). In this expression \vec{p} is the vectorised data space, \mathbf{L} is the transformation matrix and \vec{x} is the model space. Note that \vec{p} contains the complete pre-stack dataset. All traces of all shot records are concatenated and stored as a long vector (vectorised). \vec{x} is the focal domain, also stored as a long vector. Both data and model space are in the space-time domain. \mathbf{L} resembles the forward propagation of the data. In principle \mathbf{L} denotes a huge convolution matrix. For the case of a 2D fixed spread acquisition, it has the dimensions $nt * nrcv * nsrcc, nt * ng * ng$. The number of time samples is denoted by nt , the number of receivers and sources is denoted by $nrcv$ and $nsrcc$ respectively, and the number of grid-points is denoted by ng . For a 3D fixed spread acquisition it has the dimensions $nt * nrcv_x * nsrcc_x * nrcv_y * nsrcc_y, nt * ng_x * ng_x * ng_y * ng_y$, where $()_x$ and $()_y$ denote coordinates in the x and y direction. In practice this matrix can only be generated explicitly for small datasets. Therefore, \mathbf{L} is applied as a function, which imitates the behaviour of an explicit transformation matrix. This function consists of the following steps:

1. Resort the focal domain from a long vector \vec{x} to a 3D tensor like in Figure 2.18a.
2. Transform this tensor from the time to the frequency domain.
3. Multiply every frequency slice of the focal domain with its corresponding frequency slices of the focal operator to derive the data matrix, as was visualised in Figure 2.19.
4. Transform the data from the frequency domain to the time domain to obtain the 3D data tensor.
5. Finally, vectorising the data tensor yields \vec{p} .

Since convolution in the time domain is equivalent to multiplication in the frequency domain, applying the previously explained function to the vectorised focal domain \vec{x} provides exactly the same result as applying the explicit convolution matrix \mathbf{L} .

In order to estimate the focal domain vector $\hat{\vec{x}}$, first the focusing process is defined as:

$$\hat{\vec{x}} = \alpha \mathbf{L}^H \vec{p}, \quad (3.4)$$

where $(\cdot)^H$ denotes the adjoint (i.e. the conjugate-transpose) of a matrix and $(\hat{\cdot})$ denotes that the result is not necessarily the true representation but only an estimation. Equation 3.4 also represents the first step in retrieving the focal domain iteratively by inversion. Similarly to the convolution matrix \mathbf{L} , \mathbf{L}^H becomes a huge correlation matrix, which is again imitated by an implicit function. Note that the application of the focal operators can strongly effect the scaling of the result.

Therefore, the scaling factor α is used to ensure that the reconstructed data:

$$\vec{p}_{out} = \mathbf{L}(\alpha \mathbf{L}^H \vec{p}_{in}), \quad (3.5)$$

is matching the measured traces \vec{p}_{in} in an optimum way. The scalar α can, for instance, be chosen in a way that the strongest value in the reconstructed data matches the strongest value in the measured data. Another possibility is to choose α in a way that the average amplitude of the reconstructed data matches the average of the measured traces.

Transforming from the data space to the model space with matrix \mathbf{L}^H introduces correlation artefacts in the focal domain. In Figure 3.1 four representative pictures are displayed.

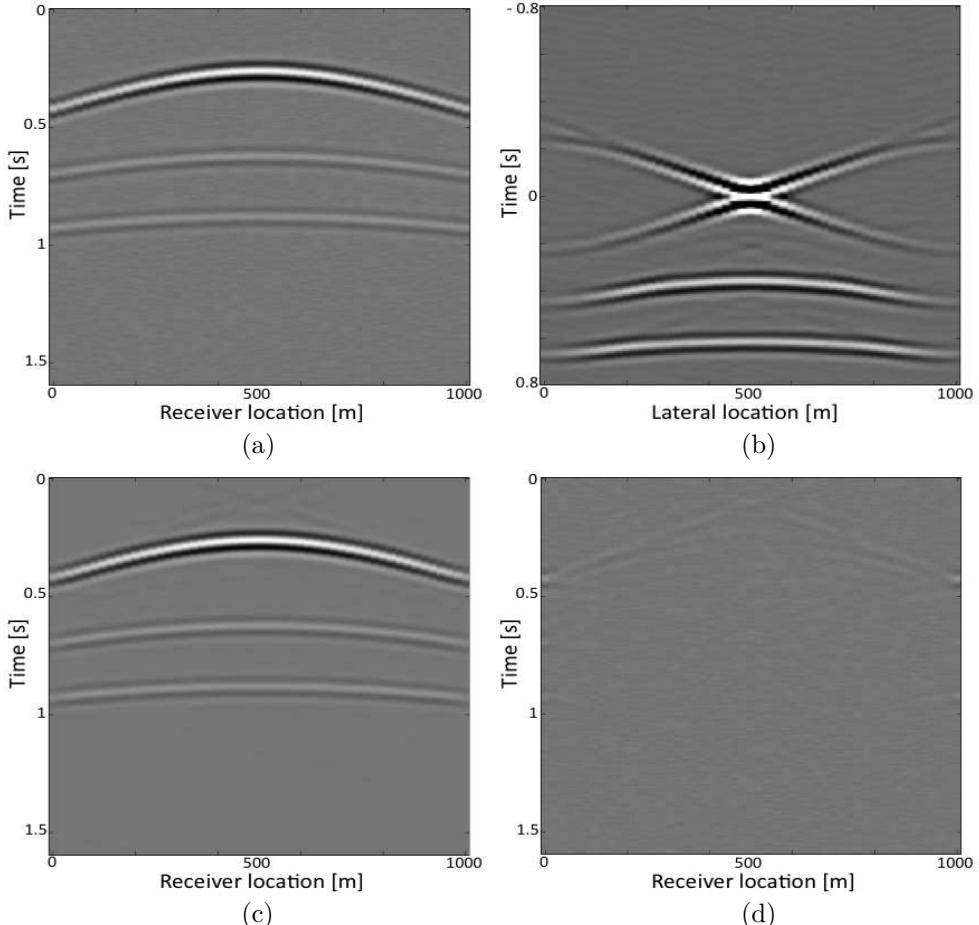


Figure 3.1: The middle section of the input data (a), the cross-correlation focal domain (b), the reconstructed data (c) and the difference (d).

In Figure 3.1a one shot record of the input data is visualised. This shot record corresponds to the one in Figure 2.12b at page 25, but here white noise with a maximum amplitude of five percent of the maximum amplitude of the data has been added to the data. In Figure 3.1b one slice through the backward propagated data is displayed, which is the output of equation 3.4. Note that instead of the desired band-limited spike around time zero the first reflection is represented by the so-called 'correlation cross'. By comparing Figure 3.1a and 3.1b it becomes obvious, that this is not the desired sparse data representation. However, transforming this focal domain back to the data space with equation 3.3 delivers a good reconstruction result as can be obtained in Figure 3.1c. The difference plot in Figure 3.1d consists of random noise and correlation artefacts. Random noise can not be mapped by the focal transformation and, therefore, remains in the residual. Note that the difference plots in this thesis only show the data residual at the known traces.

Until now it was assumed that the input data is dense. In that case the data vector \vec{p} has the same length as the model vector \vec{x} . Hence, the number of rows of the correlation matrix L^H is equal to the number of its columns. This is visualised in Figure 3.2a.

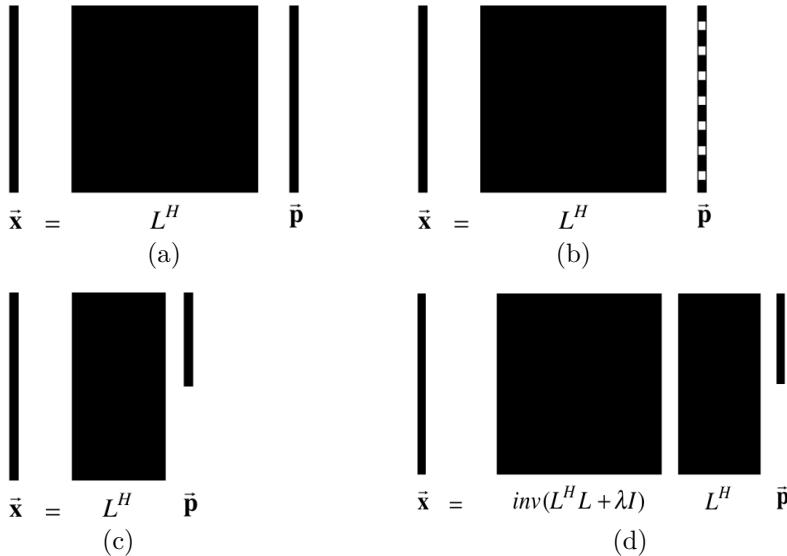


Figure 3.2: Schematic representation of the correlation type forward double focal transformation for dense input data (a) and for coarse input data (b and c). In (d) the inversion type (pseudoinverse) of forward double focal transformation is used.

In the case of missing traces there are two choices. The first choice is visualised in Figure 3.2b. There, the missing traces are set to zero but kept in the data vector (denoted by the white parts). Another possibility is displayed in Figure

3.2c. There, the empty traces are simply ignored and not taken into account and hence, the corresponding columns of \mathbf{L}^H are also ignored. By skipping the empty traces in \vec{p} and the corresponding columns in \mathbf{L}^H the equation becomes under-determined. For the correlation type forward double focal transformation both formulations (b or c) lead to the same focal domain. However, it is essential to avoid any assumptions about the data at the unknown locations (e.g. setting them to zero) when the focal domain is derived by stabilised inversion (section 3.2, equation 3.9, Figure 3.2d). Independent of the implementation of the double focal transformation, the focal domain is always defined on a dense sampling. The reconstruction step is either applied by manipulating the focal domain *after* computation (e.g. thresholding or muting, explained in section 3.1.1) or *during* the mapping of data to the focal domain (mathematical constraints, explained in section 3.2.2). The final mapping from the densely sampled focal domain to the densely sampled data space (defocusing), however, is always a well-posed process, using a square convolution matrix \mathbf{L} .

An example of coarse input data is visualised in Figure 3.3a where two out of three traces have been removed from the measurements. The corresponding focal domain is represented in Figure 3.3b. The only difference to Figure 3.1b is that additional aliasing noise occurs. It appears as a stripe-pattern around zero time (marked by the red triangles). This noise appears due to the missing traces in the data domain. The resulting reconstructed dataset is visualised in Figure 3.3c and the data misfit is visualised in Figure 3.3d. The residual for the known traces is small, but the reconstruction in the data gaps is insufficient. Since the data were just transformed to the focal domain and back to the data domain, the obvious question is: why is there any reconstruction at all? In Figure 3.4a the aliasing noise in the coarse input data is clearly visible in the wave-number frequency domain. When the coarse data is correlated with the focal operator, all energy outside the seismic bandwidth of the focal operator (Figure 3.4b) is multiplied by zero and therefore, automatically muted. Effectively, focal transformation also acts as a bandpass-filter, which leads to aliasing noise suppression and, therefore, to reconstruction. However, a closer look at Figure 3.4c shows, that within the seismic bandwidth there is still aliasing noise overlapping with the data content. As a result, the reconstruction is insufficient in the gaps and artefacts appear in the reconstructed section in Figure 3.3c. Comparing Figure 3.4c and 3.3b, it can be observed that the aliasing noise is much clearer separated from the data content in the focal domain, than in the wave-number frequency domain. The double focal transformation clearly separates aliasing noise that was generated by regular undersampling.

Here, a comparison with the Fourier transformation is interesting, since regular undersampling is the worst-case scenario for Fourier reconstruction. This becomes more clear by investigating Figure 3.5 where the blue sine-curve represents the true signal that is acquired coarsely in a regular fashion in (a) and in an irregular fashion in (b). For regular sampling at least two samples per wave-length are required to describe the continuous signal accurately (Nyquist theorem). If this

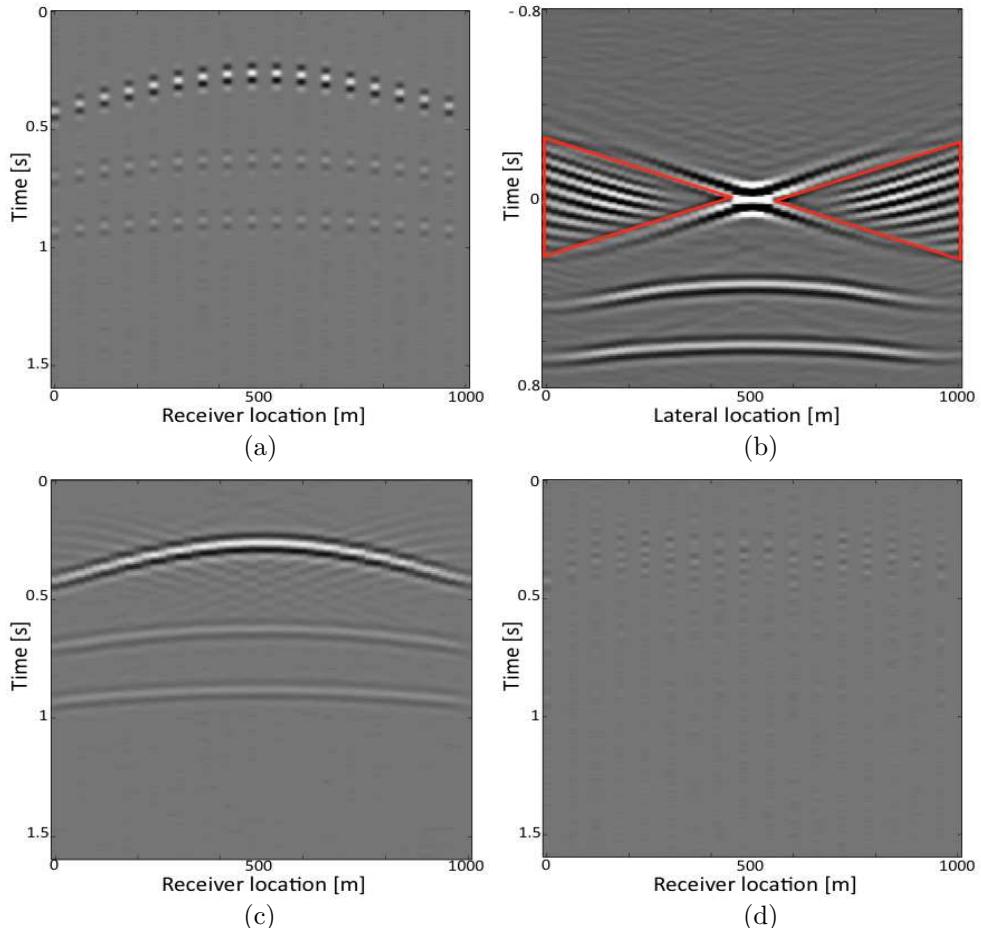


Figure 3.3: The middle section of the coarse input data (a), the cross-correlation focal domain (b), the reconstructed data (c) and the difference between input and reconstructed data (d).

criterion is not full-filled also lower frequency signals fit the measurements (green curve in Figure 3.5a). For the case that the same number of samples is distributed in an irregular fashion (Figure 3.5b), no lower frequency component fits the measurements. In both cases higher frequency signals can be fitted through the samples as well, which is indicated for the irregular sampling case by the green curve in Figure 3.5b. However, this noise is often already outside the Nyquist frequency and, therefore, doesn't disturb the true signal recovery. Even though this principle is illustrated here for a very simple example, it represents a general property of the Fourier transformation and is also valid for higher-dimensional, real measurements.

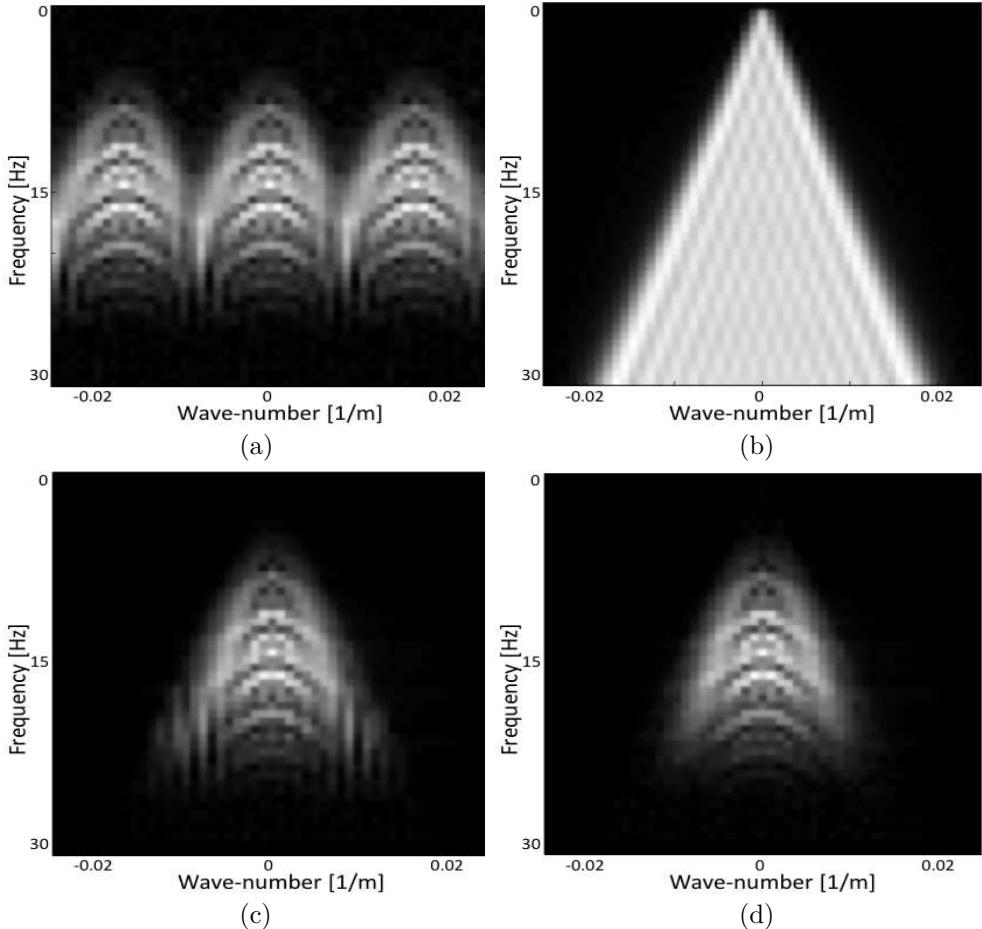


Figure 3.4: FK-domain representations of the middle section of: The coarse input data (a), the focal operator (b), the reconstructed data due to band limitation (c) and the reconstructed data due to manual muting in the focal domain (d).

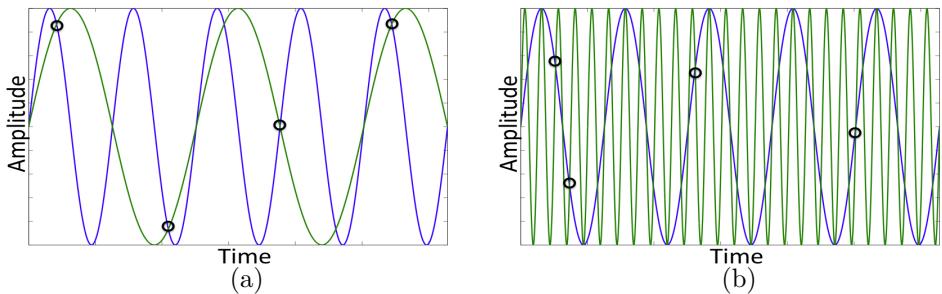


Figure 3.5: In both plots the blue curve represents the original signal. In (a) the coarse regular sampling (black circles) allows the appearance of low frequency noise (green). In (b) due to the irregular sampling only higher frequency noise appears (green).

3.1.1 Manual filtering of aliasing noise

Based on the preceding observations, a simple de-aliasing filter can be generated based on the velocity profile of the focal operator. This filter is used to mute the noise area in the focal domain, which is indicated by red triangles in Figure 3.3b. The result is the suppression of energy in this area as is shown in Figure 3.6b. Transforming this muted focal domain to the data space leads to a good reconstruction result with only very few artefacts and residuals, as can be observed in Figure 3.6c and 3.6d. Transforming the reconstructed data to the wave-number frequency domain (Figure 3.4d) confirms the successful elimination of aliasing noise. This route was also proposed by Berkhouwt and Verschuur (2005).

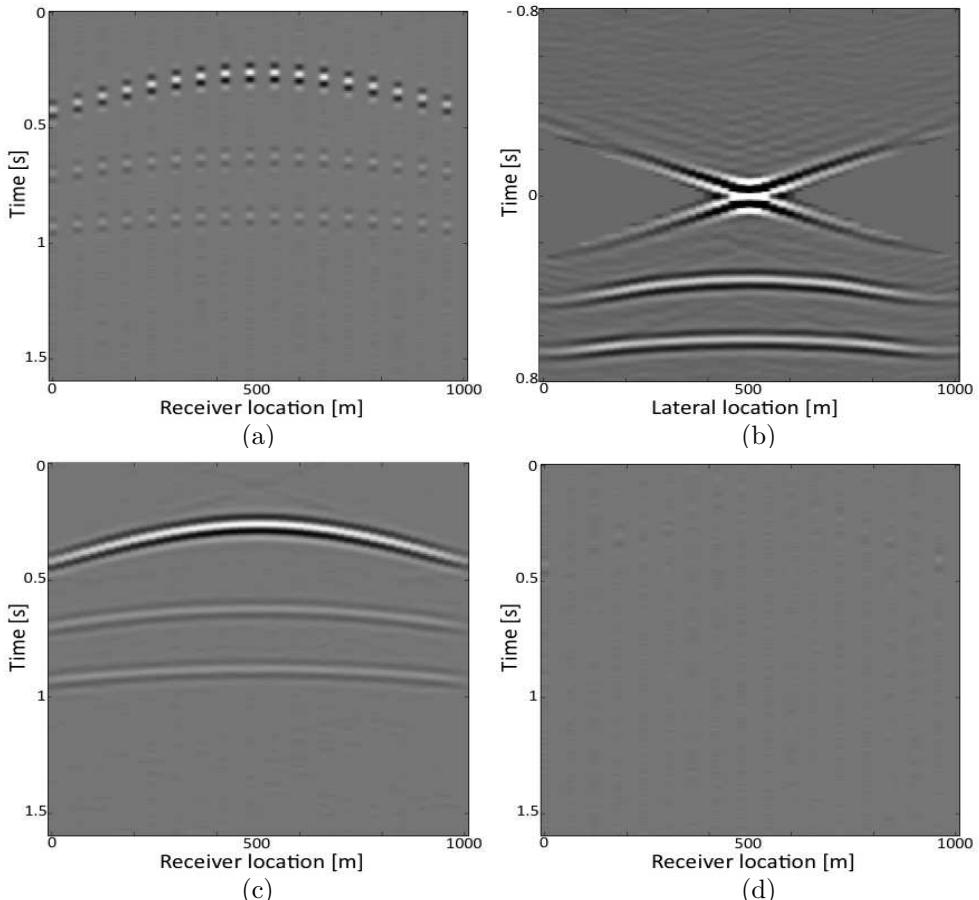


Figure 3.6: The middle section of the coarse input data (a), the cross-correlation focal domain, where the noise area was manually muted (b), the reconstructed data (c) and the difference between input and reconstructed data (d).

To avoid confusion the following definitions are used and explained:

- Like in most other transformations the *forward* double focal transformation maps the data space (data) to the model space (focal domain) and the *inverse* double focal transformation maps the model space (focal domain) back to the data space (data).
- The inverse double focal transformation (equation 3.3) always transforms a focal domain, which is defined on a dense grid, to the reconstructed data, which is also defined on a dense grid. Therefore, convolution operator \mathbf{L} is a square matrix. Note that this will change for the multi-level implementation (section 3.3).
- The *forward* double focal transformation can be implemented in two ways. The correlation focal domain can be derived by equation 3.4 and, for the case of missing traces, \mathbf{L}^H is not necessarily a square matrix. Alternatively, the focal domain can be derived by optimisation schemes. These schemes essentially solve an *inversion* problem. This is not to be confused with the definition of the *inverse* double focal transformation.

3.2 The forward double focal transformation by optimisation

The reconstruction result in Figure 3.6c looks already promising, but it can be further improved. For instance, there are still correlation artefacts just above the apex of the first reflection event. Also the residual at the known traces in Figure 3.6d consists not yet of white noise only, but remnants of the first reflector are visible. The reason for this is, that the transformation matrix \mathbf{L}^H is not exactly a scaled version of \mathbf{L}^{-1} , like it is in the case of the Fourier transformation. Using the correlation type forward transformation \mathbf{L}^H instead of the exact inversion type \mathbf{L}^{-1} consequently leads to an imprecise mapping of data and, therefore, to artefacts and residuals.

A better result can be achieved by redefining the forward double focal transformation as an inverse problem. Thus, instead of using the correlation operator \mathbf{L}^H the true inverse \mathbf{L}^{-1} is used for the forward double focal transformation:

$$\vec{x} = \mathbf{L}^{-1} \vec{p}. \quad (3.6)$$

There are two possibilities to address this inversion problem. The first one (similar to the one explained for the correlation type implementation) is to consider the empty traces as information and just set them to zero as is indicated in Figure 3.2b. In that case data and model vector have the same length. Consequently, \mathbf{L} is a square matrix and can be inverted. The aliasing noise, which is not filtered by the seismic bandwidth of the focal operator, maps to the focal domain. There it can be muted. After inverse double focal transformation the reconstructed data can be obtained. On the other hand, just from an intuitive standpoint, assuming

that the unknown traces are zero and afterwards repairing them doesn't seem to be very effective. A better option would be to assume nothing about the empty traces. Hence, the empty traces and corresponding columns in \mathbf{L}^H are deleted, as is indicated in Figure 3.2c. Consequently \mathbf{L} becomes a non-square matrix, which is not invertible. Therefore, focal transformation will be defined as a constrained optimisation problem.

3.2.1 Least-squares optimisation

Redefining focal transformation as an optimisation problem can be achieved by minimising the following cost function J :

$$J = \|\vec{p} - \mathbf{L}\vec{x}\|_2. \quad (3.7)$$

The l_2 norm of the residual ($\vec{r} = \vec{p} - \mathbf{L}\vec{x}$) is defined as:

$$\|\vec{r}\|_2 = \sum_i r_i^2.$$

The objective of equation 3.7 is to find a focal domain \vec{x} , which minimises the l_2 norm of the data-misfit \vec{r} . The solution to equation 3.7 is the least squares inverse:

$$\hat{\vec{x}} = (\mathbf{L}^H \mathbf{L})^{-1} \mathbf{L}^H \vec{p}. \quad (3.8)$$

The new matrix $(\mathbf{L}^H \mathbf{L})$ is by definition a square matrix, but not necessarily well-conditioned. The inversion of this matrix results in unstable solutions. Commonly, the unit matrix \mathbf{I} is added to the matrix (equation 3.9). With the scalar λ the strength of the stabilisation is regulated. Note that $\hat{\vec{x}}$ is the solution to the inversion problem and might be different depending on the definition of this inversion (for instance, it is dependent on the magnitude of λ):

$$\hat{\vec{x}} = (\mathbf{L}^H \mathbf{L} + \lambda \mathbf{I})^{-1} \mathbf{L}^H \vec{p}. \quad (3.9)$$

At first glance this stabilisation step looks like an arbitrary trick to make the matrix $(\mathbf{L}^H \mathbf{L})$ invertible. However, equation 3.9 is not any longer the solution to equation 3.7 but to another optimisation problem, which minimises:

$$J = \|\vec{p} - \mathbf{L}\vec{x}\|_2 + \lambda \|\vec{x}\|_2. \quad (3.10)$$

For the derivation of this function for complex values see Appendix C. In addition to the objective of the minimum data-misfit (first term in equation 3.10) the l_2 norm of the solution space (focal domain) is minimised as well (second term). With parameter λ the balance between the two terms is regulated. This is the so-called 'minimum energy' assumption. Hence, in equation 3.10 a focal domain is searched that provides the minimum data misfit \vec{r} and at the same time has

the least energy. The solution to this problem only depends on the choice of λ . A schematic representation of equation 3.9 was already visualised in Figure 3.2d.

Please note that all results using the least-squares constraint are generated using the LSQR algorithm. The term LSQR stands for Least SQuares Regression and the software uses an iterative (conjugate-gradient-like) method. For further information see Paige and Saunders (1982). From Figure 3.7d it becomes obvious that the approach introduced above leads to a perfect data-misfit for the known traces, where only white noise is visible. Compared to the correlation result in Figure 3.3d this is a desired improvement. Also the reconstruction in the gaps Figure 3.7c is improved compared to Figure 3.3c.

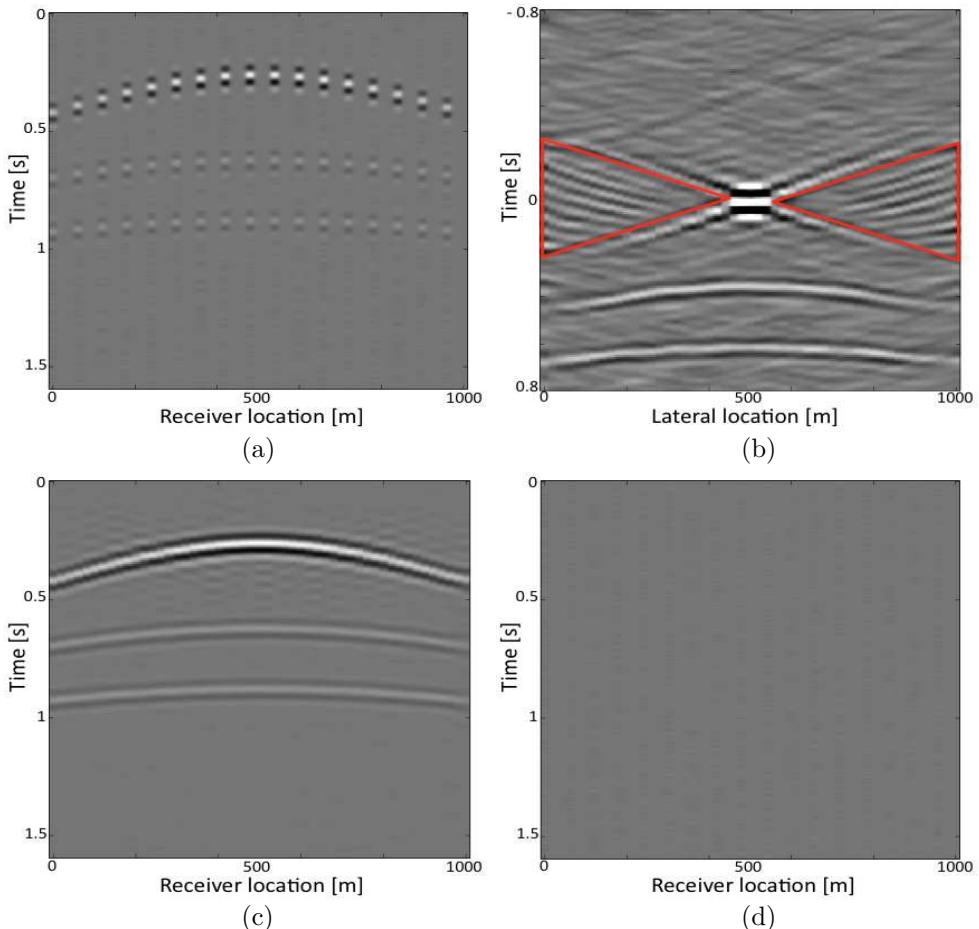


Figure 3.7: The middle section of the coarse input data (a), the least-squares constrained focal domain (b), the reconstructed data (c) and the difference between input and reconstructed data (d).

The corresponding focal domain in Figure 3.7b has the same key features as its correlation counterpart in Figure 3.3b. However, it is filled everywhere with weak energy. This is a well-known feature of the least-squares constraint to the model space (second term in equation 3.10). If, for example, two focal domains provide the same data misfit then the least-squares constraint picks the one where the energy is as equally distributed in the focal domain as possible. Hence, many weak values in the focal domain are preferred to a few strong ones.

At this point it is important to remember that the major difference of the focal reconstruction to most other reconstruction methods is that it uses wave propagation, a physical process, to transform the data. Therefore, the correlation type focal domain looks very much like a backward propagated data set. Nevertheless, Figure 3.7b is the result of an optimisation process (equation 3.10). Therefore, energy in the focal domain is distributed such that the objective function J is minimised. As a result, the estimated focal domains do not always have a physical equivalent. In the following, different strategies and optimisation schemes are introduced.

Penalising aliasing noise by model space weighting

Even though the reconstruction in Figure 3.7c is improved compared to the one in Figure 3.3c, still aliasing noise is visible in the least-squares focal domain in Figure 3.7b. The least-squares constraint stabilises the solution, but does not suppress the occurrence of aliasing noise. Again this area could be muted. However, a weighting function provides a more flexible implementation. For instance for more complex data scenarios, noise and data area might partly overlap. By strictly setting the noise area to zero the data might be harmed. Utilising the weighting matrix $\mathbf{\Lambda}_x$, on the other hand, provides the possibility to multiply every element in the focal domain with a scalar during the inversion process. In this way certain areas can be emphasised and others suppressed.

The optimisation problem is then defined as:

$$J = \|\vec{p} - \mathbf{L}\mathbf{\Lambda}_x\vec{x}\|_2 + \lambda\|\mathbf{\Lambda}_x\vec{x}\|_2, \quad (3.11)$$

where $\mathbf{\Lambda}_x$ denotes a diagonal matrix. Small values on the diagonal penalise energy at the corresponding locations in the model vector.

For the following reconstruction example, the weighting in the noise areas was chosen to be zero as can be observed in Figure 3.8b. This strict weighting was performed, because in this situation the aliasing noise clearly maps to a different area in the focal domain, than the data. In more complex situations it will be less certain where the noise maps to. In that case the weighting function provides the flexibility to use different shapes with different amplitudes (e.g. 0, 0.5, 1 or anything in between). The result is a very good reconstruction (Figure 3.8c) with only small artefacts. The data-misfit consists mainly of white noise, which is hardly visible in Figure 3.8d, because of its weak amplitude.

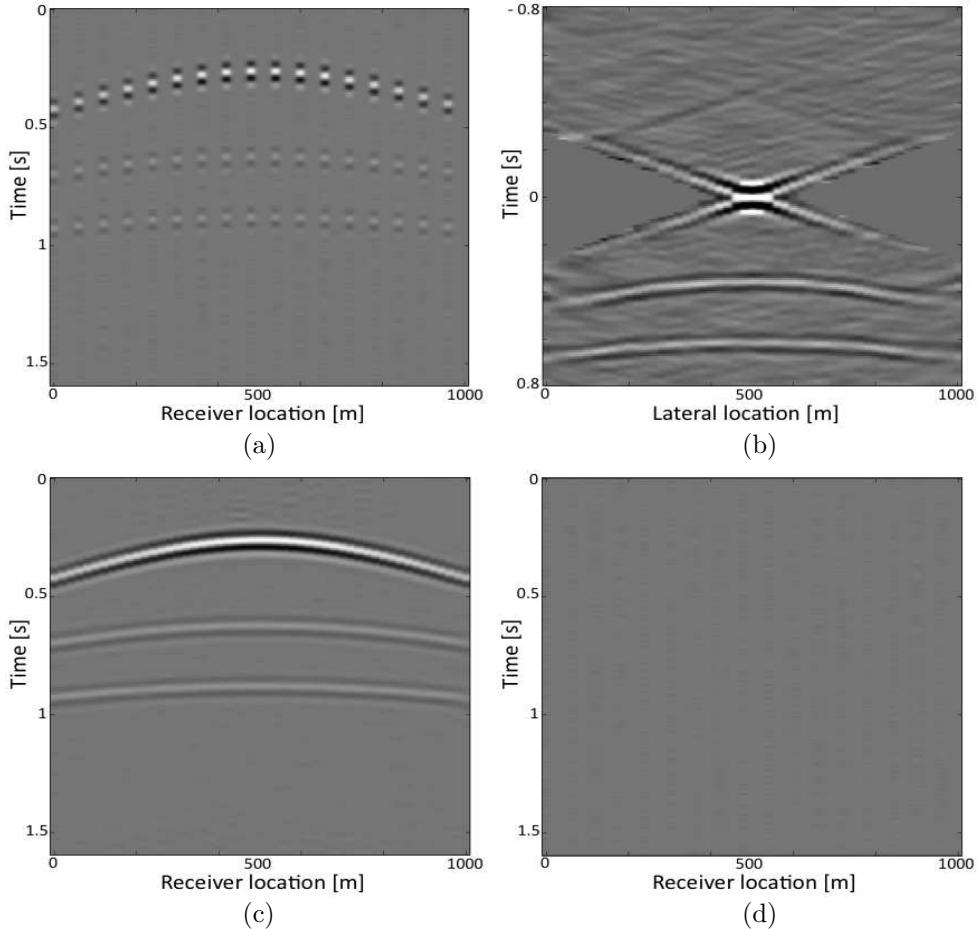


Figure 3.8: The middle section of the coarse input data (a), the least-squares constrained focal domain using model weighting (b), the reconstructed data (c) and the difference between input and reconstructed data (d).

3.2.2 Sparseness constrained optimisation

The previously introduced reconstruction method works very well if the aliasing noise area is clearly separated from the data in the focal domain. However, this is not always the case. In Figure 3.9a the input data are irregularly sampled. It is clearly visible in Figure 3.9b that aliasing noise and signal content are not fully separated any more. The same holds for more complex data sets. If data and noise are overlapping in the focal domain it is impossible to define a model weight where the noise can be entirely suppressed without harming the data.

However, there is still the other property which can be exploited: while the data

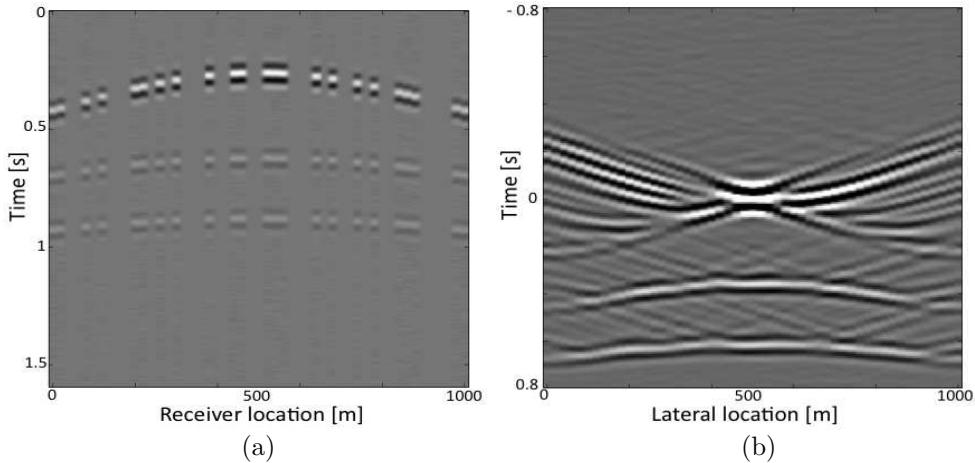


Figure 3.9: The middle section of irregularly sampled data (a) and the corresponding (cross-correlation type) focal domain (b).

content is focused, aliasing noise spreads out. Therefore, the data is represented relatively sparsely, while the aliasing noise is not. The logical step is to replace the least-squares constraint in equation 3.10 by a sparseness constraint. In that way a focal domain is looked for, which contains as few as possible non-zero elements, while still maintaining a good data misfit. This can be accomplished by choosing non-quadratic penalty functions, like the Cauchy norm, or the l_0 norm, to constrain the optimisation (see equations 3.12a,b and Zwartjes and Gisolf, 2007). However, these constraints greatly increase the non-linearity of the inversion problem. On the other hand, choosing a l_1 constraint (equation 3.12c) leads to much more stable inversion results with better convergence properties. Furthermore, under certain assumptions the solution with the smallest l_1 norm is the sparsest solution to a problem, see Donoho (2004).

For the sake of convenience four different norms are listed below, where equation 3.12d represents the l_2 norm:

$$\|\vec{x}\|_0 = \sum_i 1 : x_i \neq 0 , \quad (l_0) \quad (3.12a)$$

$$\text{Cauchy norm} = \sum_i \frac{1}{2} \ln(1 + x_i^2) , \quad (\text{Cauchy}) \quad (3.12b)$$

$$\|\vec{x}\|_1 = \sum_i |x_i| , \quad (l_1) \quad (3.12c)$$

$$\|\vec{x}\|_2 = \sqrt{\sum_i x_i^2} . \quad (l_2) \quad (3.12d)$$

Hence, adopting the l_1 norm, in the following a focal domain with the minimum absolute value is searched that still provides a minimum data misfit.

The corresponding optimisation problem can be formulated as the so-called "Basis Pursuit De-noise" problem (van den Berg and Friedlander, 2008a):

$$\underset{\vec{x}}{\text{minimise}} \quad \|\vec{x}\|_1 \text{ subject to } \|\vec{p} - \mathbf{L}\vec{x}\|_2 \leq \sigma. \quad (3.13)$$

In equation 3.13 the l_1 norm of the focal domain is minimised while keeping the l_2 norm of the data misfit smaller than the scalar σ , which is most commonly the noise level of the input data. All results in this thesis using the l_1 constraint (equation 3.13) will be computed with the spgl1 solver, which stands for Spectral Projected Gradient - l_1 . It is a MATLAB solver for large-scale l_1 norm regularised least-squares problems. It requires relatively many iterations to converge, but convergence to the correct solution is always guaranteed. For further information see Appendix C and van den Berg and Friedlander (2008a). The Basis Pursuit De-noise problem is closely related to the Lasso problem:

$$\underset{\vec{x}}{\text{minimise}} \quad \|\vec{p} - \mathbf{L}\vec{x}\|_2 \text{ subject to } \|\vec{x}\|_1 \leq \tau, \quad (3.14)$$

which can be optimised in an easier manner. In equation 3.14 the sparseness of the solution is controlled by the parameter τ , while the l_2 norm of the data residual is minimised. For certain values of σ and τ the optimal points of these two problems coincide. Therefore, the spgl1 solver optimises a number of Lasso problems and resets the sparseness constraint τ inbetween (Newton update), to converge to the same solution as solving the Basis Pursuit Denoise problem (equation 3.13) would do.

In Figure 3.10b the focal domain is derived from 200 iterations of the spgl1 solver, from the coarse input data (Figure 3.10a). The corresponding reconstruction result (Figure 3.10c) is perfect and the data-misfit (Figure 3.10d) is within the noise level. The data were reconstructed by simply assuming that the focal domain must be sparse in the l_1 sense. There are no assumptions made on where the aliasing noise would eventually map to. Thus, additional weighting in the focal domain is not required. This makes the l_1 optimisation a very flexible implementation, which is suitable for different sampling cases and more complex data as well. Note the strong similarity of the focal domain to the one related to the fully sampled input data in Figure 2.18 on page 32.

In Figure 3.11a the performance of the spgl1 solver is visualised. The horizontal axis denotes the l_1 norm of the focal domain and the vertical axis denotes the l_2 norm of the data-misfit. The solver starts with an initial τ value. At this point the focal domain is almost empty and, hence, $\|\vec{x}\|_1$ is small. Consequently, the corresponding data residual is high, because it needs a certain amount of energy in the model space to explain the energy in the data space. Next, a Newton iteration is performed, which allows for a higher absolute value in the solution space. This allows the data misfit to reduce. Afterwards, the Lasso problem for the current τ

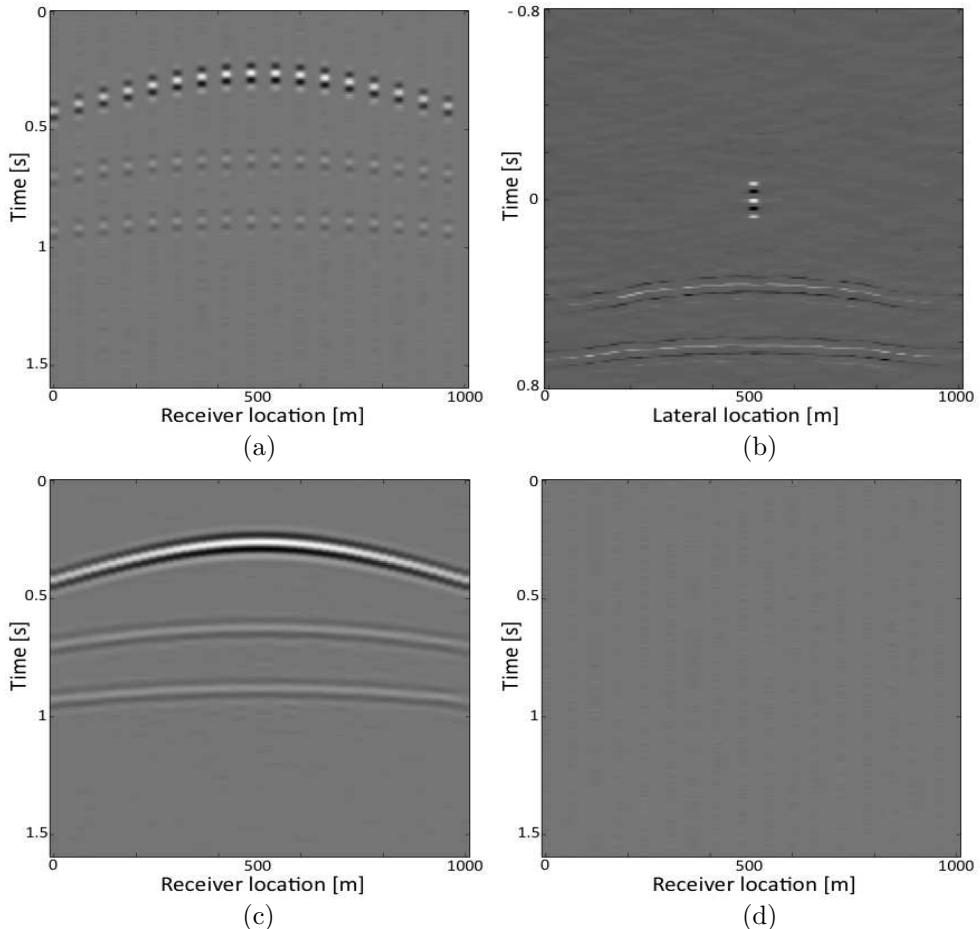


Figure 3.10: The middle section of the coarse input data (a), the l_1 constrained focal domain after 200 *spgl1* iterations (b), the reconstructed data (c) and the difference between input and reconstructed data (d).

is optimised, which lowers the data misfit without increasing the absolute value of the model space. This can be understood as a kind of redistribution of the energy in the focal domain to achieve a better data-fit. At a certain point, however, more energy in the focal domain is needed to explain all the events in the data space. Then again a Newton iteration resets the τ , and so on, until the required data-misfit or the maximum number of iterations is reached.

In Figure 3.11b the performance of the *spgl1* solver is compared to the LSQR solver and a Matching Pursuit implementation. Basically the LSQR solver minimises the data misfit at the known traces in only a couple of iterations but generates a very large absolute value in the solution space. As is known from Figure 3.7

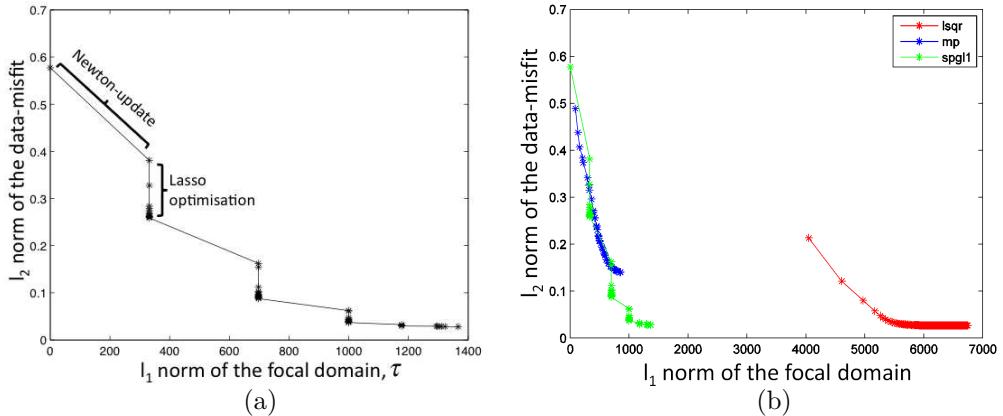


Figure 3.11: The *spgl1* solver performance curve for 200 iterations (a) for deriving the focal domain (Figure 3.10b) from the coarse input data (Figure 3.10a). In (b) the *spgl1* performance curve (green) is compared to the LSQR solver, which solves equation 3.10 (red) and a Matching Pursuit algorithm (blue).

this leads to a poor reconstruction result, because the least-squares assumption is not suitable for the focal reconstruction. Remember that the LSQR solver aims for a minimum l_2 norm of the solution, which usually leads to a smearing of energy (many little values are preferred to a few strong ones) in the model space. Measuring the l_1 norm of such a solution will result in a relatively large value, as can be observed by the red curve in Figure 3.11b. The Matching Pursuit solver (see Appendix C) initially converges very quickly, and gives for the same data residual as *spgl1* an even sparser focal domain. However, the convergence rate drastically slows down for a higher number of iterations and it appears that it can't reach the low data-misfit of the *spgl1* or the LSQR solver. Hence, Matching Pursuit is well suited for an initial reconstruction result, but less useful for a full data reconstruction process, which covers all the low amplitude details in the data.

As was explained before, deriving the focal domain is performed by a constrained optimisation. The least-squares constraint maps the first reflector to a correlation cross in Figure 3.7b, because it has a smaller l_2 norm than a band-limited spike. Also the aliasing noise is spread across a relatively large area in the focal domain and, therefore, has a small l_2 norm as well. To minimise the l_1 norm in equation 3.13, on the other hand, maps the first reflector to a band-limited spike (Figure 3.10b), because this has a smaller absolute value than a correlation cross. Since the aliasing noise does not contribute to a smaller data-residue, but has a high absolute value, it is automatically suppressed. Remember that both focal domains (the l_2 constrained one and the l_1 constrained one) have the same corresponding data-residual. Hence, for a minimum data-misfit *at the known traces* it doesn't matter if the first reflector is represented by a correlation cross, or a band-limited

spike (compare Figure 3.7d and 3.10d), but for an adequate reconstruction *in the gaps* a sparse data representation is essential.

Data weighting

When interpreting Figure 3.11a in a physical way, a general behaviour of the spgl1 solver can be observed. First, it puts some energy in the focal domain to explain the strongest events in the data space (e.g. the strongest reflectors). In the current example this corresponds to a strong spike in the focal domain around time zero (Figure 3.10b). With this strong spike the whole first reflector can be explained. Since the first reflector has a higher amplitude than the second and third one, it is, from a data-misfit point of view, more important. Subsequently much more unfocused energy is needed (second and third event in the focal domain) to only explain the weaker second and third reflector. In conclusion, the spgl1 solver prefers strong events over weak ones. In Figure 3.12a and b the focal domain and the corresponding reconstructed data are displayed for a maximum of 20 spgl1 iterations. As expected, only the first reflector is reconstructed.

To avoid the emphasis of reconstruction of strong events at early iterations of the inversion process, a data weighting can be applied. In order to do that, first, a smoothed, densely sampled version of the coarse input data (absolute value) is generated (Figure 3.12c) and used to precondition the transformation matrix \mathbf{L} . One possibility is to weight the coarse input data previous to the inversion:

$$\vec{p}_w = \mathbf{\Lambda}_p \vec{p}, \quad (3.15)$$

where \vec{p} is the coarse input data and $\mathbf{\Lambda}_p$ is a diagonal matrix with $((\vec{p}_s)^k + \epsilon)^{-1}$ on the main diagonal. Hereby, \vec{p}_s is the smoothed data (absolute value) at the known trace locations, ϵ is a stabilisation-variable (small positive number) to avoid division by zero and \vec{p}_w is the coarse, weighted data (Figure 3.12d). The integer k regulates the influence of the weighting.

In the case that $k = 1$, the events in the weighted data would have roughly everywhere the same amplitude. For $k = 2$, the amplitude distribution of the input data would be inverted (i.e. strong events become weak and weak events become strong). Optionally, a scalar can be applied to equation 3.15, to scale the maximum amplitude of the weighted data to the input data. However, this has no influence on the performance. Subsequently a focal domain can be looked for with the help of equation 3.13. Since the amplitude of the second and third reflector was increased relative to the first one, they will be reconstructed at earlier iterations than in the un-weighted reconstruction case.

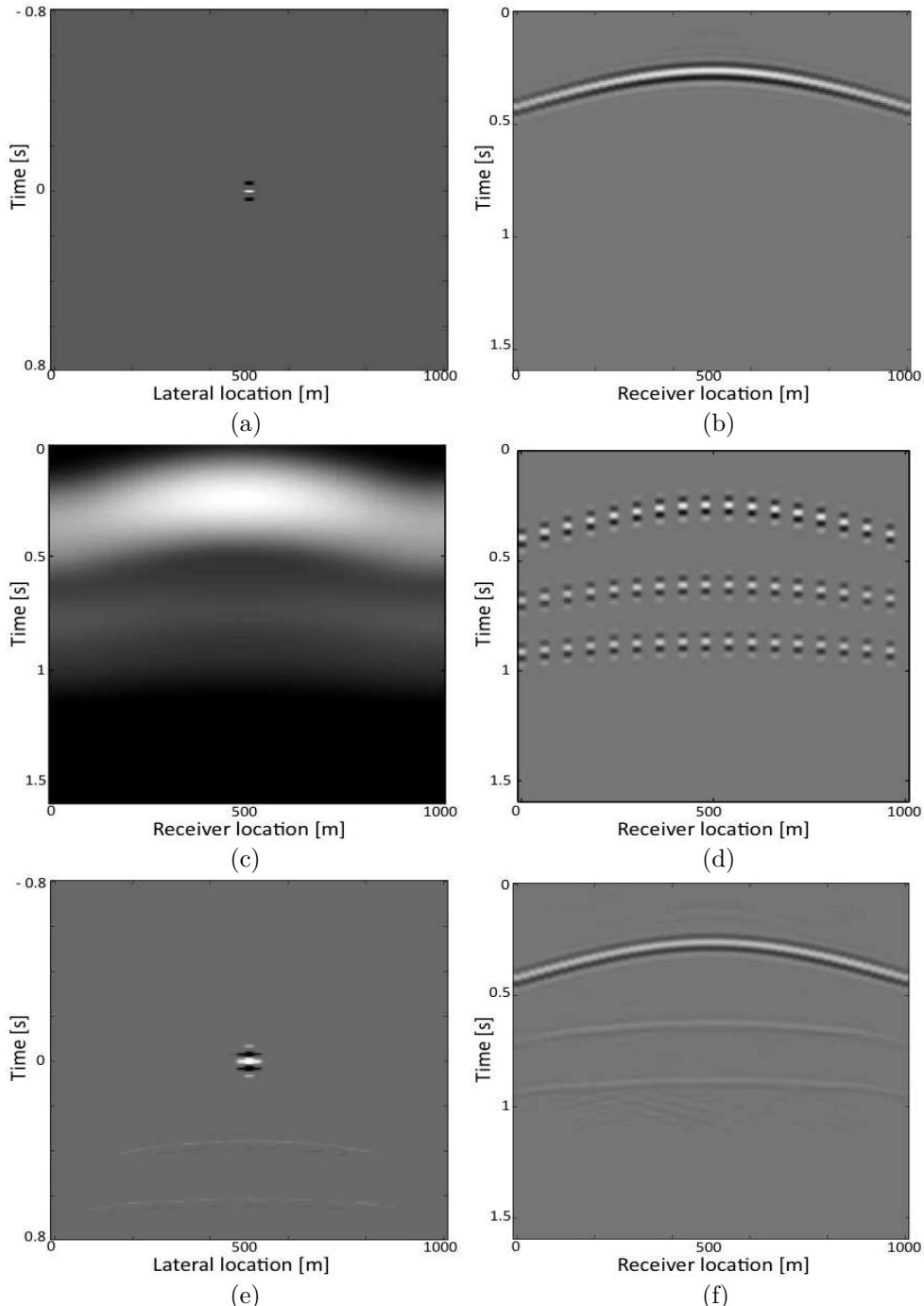


Figure 3.12: Middle section of the focal domain after 20 *spgl1* iterations (a) and corresponding reconstructed data (b). In (c) the middle section of the smoothed version of the dense data is displayed and in (d) a weighted version ($k = 1$) of the coarse data in optimisation problem equation 3.17. In (e) and (f) focal domain and corresponding reconstructed data are displayed using data weighting with $k = 2$ for 20 *spgl1* iterations.

Afterwards, the weighting of the reconstructed data can be cancelled by simply applying the weighting matrix inversely to obtain the reconstructed data with the original amplitude distribution:

$$\hat{\vec{p}} = (\mathbf{\Lambda}_p)^{-1} \hat{\vec{p}}_w. \quad (3.16)$$

Note that the weighing matrix in equation 3.15 relates the coarse original data and the coarse weighted data. Therefore, it has a smaller dimension than the inverse weighting matrix in equation 3.16, which relates the dense, reconstructed, weighted data and the dense reconstructed data with original amplitude distribution.

For the given simple data example a decent reconstruction result would be achieved in that way. However, in a more realistic case reflection events are crossing. A data weighting is essentially an amplitude normalisation of the data. But to normalise crossing events generates non-physical behaviour, which creates artefacts in the focal domain and, therefore, degrades the reconstruction quality.

A more generally applicable way to avoid the preference for reconstructing strong reflectors, is to use the true data as input and include the data weighting in the transformation. The data-weighted Basis Pursuit Denoise problem reads:

$$\underset{\vec{x}}{\text{minimise}} \quad ||\vec{x}||_1 \quad \text{subject to} \quad ||\vec{p} - (\mathbf{\Lambda}_p)^{-1} \mathbf{L} \vec{x}||_2 \leq \sigma. \quad (3.17)$$

As explained before, in equation 3.15 the choice ($k = 1$) would equalise the amplitude of all events in the input data and, therefore, give them an equal contribution to the data-misfit in equation 3.17. However, still the first reflector can be represented by a band-limited spike in the focal domain, while reflector two and three have residual spatial phase in the focal domain and are, thus, less sparsely represented. Therefore, after 20 spgl1 iterations still only the first reflector would be addressed.

By choosing $k = 2$ the amplitude distribution in the data is basically inverted. Hence, not taking into account weak reflectors for the representation in the focal domain leads to a higher data-misfit than for strong reflectors. The focal domain derived with $k = 2$ is visualised in Figure 3.12e and the corresponding reconstruction result in Figure 3.12f. It can be observed that with a data weighting already after 20 spgl1 iterations all three reflectors are included. However, this is not yet a perfect reconstruction result. The total data misfit of the weighted optimisation is roughly the same as for the genuine optimisation. The main achievement is that weak reflectors are taken into account at an earlier stage of the reconstruction process. This might be an advantage for expensive computations, where high iteration numbers need to be avoided. Another advantage is that residual data misfits in the strong reflectors will not hamper the correct estimation of weaker target reflectors.

The influence of imprecise focal operators

Until now it was assumed that the perfect velocity model is available. Of course, in a real data situation this is never the case. Therefore, the influence of an inaccurate velocity model is investigated here. In Figure 3.13a a model is displayed, in which the true first reflector is dipping from 250m depth on the right hand side to 350m depth on the left hand side. This leads to different input data as before, as displayed in Figure 3.13b. However, for the generation of the focal operator, still a plane reflector at 200m depth is assumed (indicated by the red line). The focal domain generated with the erroneous focal operator via 200 spgl1 iterations is displayed in Figure 3.13c and d. Since the virtual reflector was assumed too shallow, the first reflector is not mapped at zero time. Instead, it is under-corrected (at positive time) and not completely focused (remaining spatial phase). Furthermore, from the left to the right the depth-difference between the true reflector and the assumed virtual depth level decreases. Consequently, in the first section of the focal domain tensor in Figure 3.13c there is a higher time-error than the last section. Even though the focusing in the focal domain is not perfect, still a decent aliasing noise suppression is achieved, which leads to the desired data reconstruction and a small data residual (Figure 3.13e,f).

Please note that the robustness of the double focal transformation against inaccuracies in the focal operators is not only utilised for seismic data reconstruction (Kutscha et al., 2010), but also for primary estimation (Lopez and Verschuur, 2013a) and deblending (Doulgeris et al., 2012).

The observation that a good aliasing noise suppression is achievable even with an imprecise focal operators is actually not surprising if one considers that also reflector two and three are properly reconstructed. Since only one virtual depth level is used to explain three reflectors, the corresponding focal operator is inaccurate for at least two of the three reflectors, by default.

In Figure 3.14a another model is shown where the first layer has a velocity of 1600m/s. The corresponding coarse input data is displayed in Figure 3.14b. The focal operator was generated based on a velocity model with the first layer having a velocity of 1500m/s. Using the imprecise focal operator leads to the focal domain depicted in Figure 3.14c and d. The first reflector is over-corrected to negative times and not completely focused. However, the reconstruction quality is still sufficient as can be observed in Figure 3.14e and f.

In general, the quality of the focal reconstruction is almost not influenced by structural and velocity errors of the macro model on which the focal operators are based, if these errors are within a certain range. This property is colloquially referred to as "the robustness of the focal reconstruction". Since the same focal operators (eventually imprecise) are used to derive the focal domain (e.g. by equation 3.13) and to transform back to the data space (equation 3.3) the data is always matched at the original traces. However, the sparseness of the data representation in the focal domain influences the reconstruction quality.

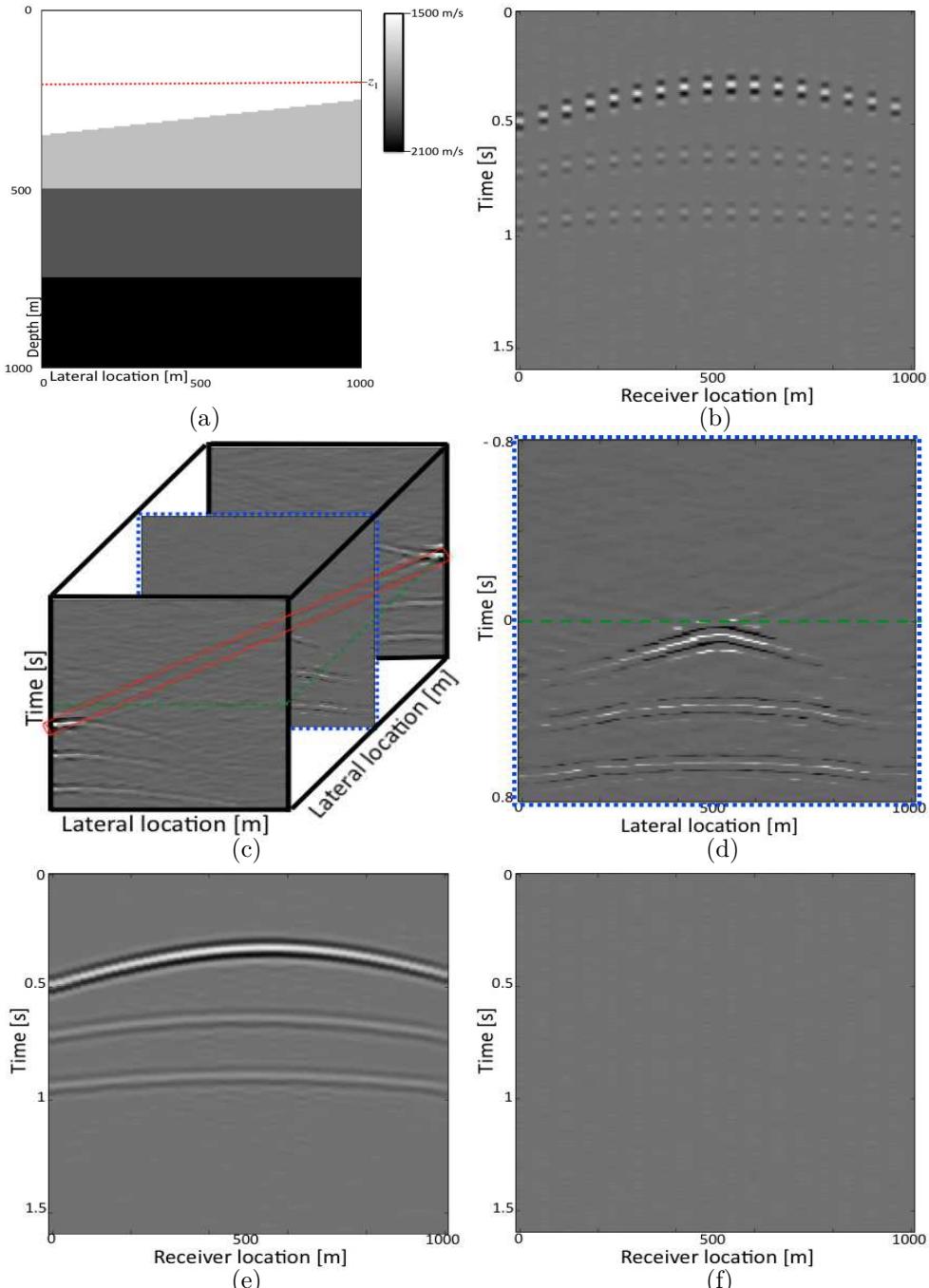


Figure 3.13: Subsurface model (a), where the first reflector is dipping from 250m depth to 350m depth from right to left. The corresponding coarse input data is displayed in (b). For the focal reconstruction a focal operator was used that was generated from a virtual plain reflector at 200m depth. The corresponding focal domain is displayed in (c) as a tensor. The middle section of (c) is displayed in (d). The dashed green line indicates zero time. The reconstructed data is visualised in (e) and the difference to the measured traces in (f).

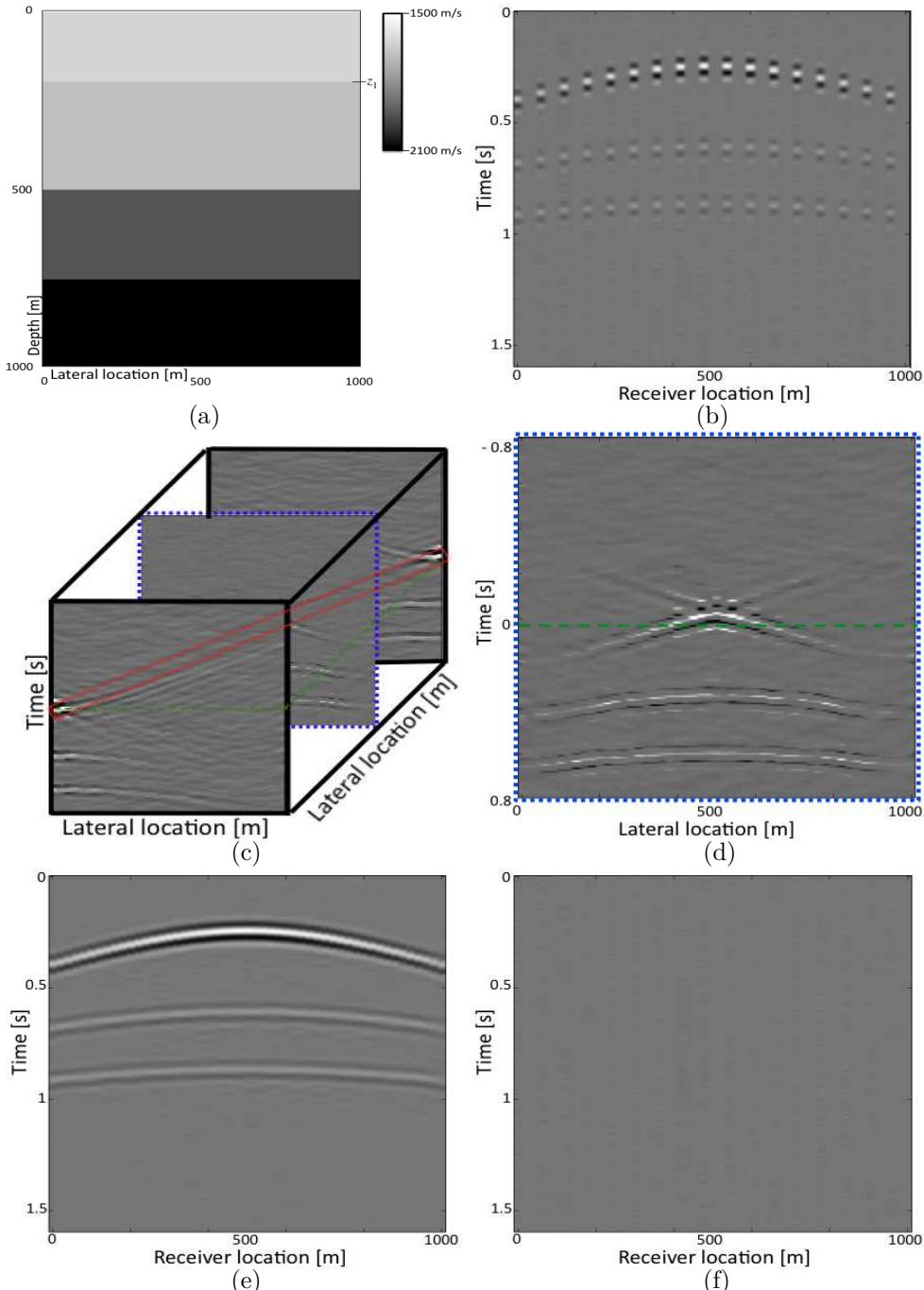


Figure 3.14: The first layer in the model in (a) has a velocity of 1600 m/s. The corresponding coarse input data is displayed in (b). A focal operator was generated based on the assumption that the first layer has a velocity of 1500 m/s. The corresponding focal domain is displayed in (c) as a tensor and in (d) the middle section of (c) is displayed. The dashed green line indicates zero time. The reconstructed data is visualised in (e) and the difference to the measured traces in (f).

In Figure 3.15 a quantitative evaluation of the previous statement is depicted. Here a plot of the reconstruction error for dense and coarse data as a function of velocity error in the focal operator is displayed. In the case of dense input data (Figure 3.15a) the following can be observed: If the focal operator is based on a velocity model with too low velocities, the reconstruction is almost not influenced, even by velocity errors up to -500m/s . The focal domain then contains residual phase, which is corrected for by applying the erroneous focal operator also for the inverse double focal transformation. However, if the focal operator is based on too high velocities the data mismatch increases more prominently. Also in this case the data are not completely focused in the focal domain, which should be corrected for by the inverse double focal transformation. However, a focal operator with a too high velocity acts as a velocity filter on the data. The larger the velocity error, the more the data is harmed. The focal operator automatically filters all energy outside its spatial bandwidth meaning that high angles are lost. It is interesting to realise that this was a desired effect for reconstruction by just forward and inverse transforming the data in section 3.1 on page 41. It is, hence, suggested to always prefer focal operators based a slightly too low velocity rather than focal operators based on slightly too high velocities.

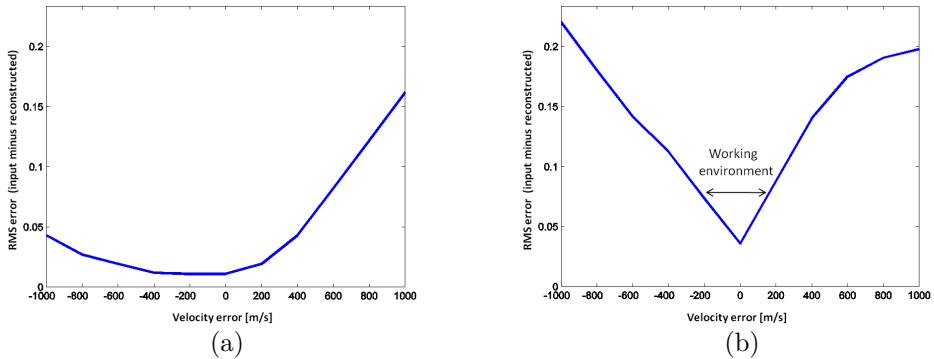


Figure 3.15: Plots of the data mismatch as a function of the velocity-error of the focal operator. The horizontal axis depicts the velocity error of the focal operator compared to the true velocity model. The vertical axis denotes the relative data mismatch of the focal-reconstructed data to the input data. In (a) the input data is dense and in (b) the input data is coarse. In both cases the densely sampled traces are taken into account to compute the data-mismatch.

The influence of missing traces to the robustness of the focal reconstruction

For real coarse data there is no possibility to evaluate the reconstruction in the data gaps objectively. Therefore, only the measured traces are taken into account for the computation of the data-misfit. In that case the data-misfit dependency

of velocity-error of the focal operator is very similar to the case of dense input data (Figure 3.15a).

On the other hand, since this is synthetic data, the dense input data is actually available for comparison. In Figure 3.15b the reconstructed data is compared at all traces, even though the utilised input data was coarse. The same is of course possible if densely sampled field data is available, which can be undersampled afterwards to test the reconstruction algorithm. However, it can be observed that the data mismatch increases almost equally for both, focal operators based on too high velocities and based on too low velocities. Since in both cases the data isn't properly focused, the assumption that the data is sparsely represented in the focal domain is not fully correct any more. That decreases the capability to distinguish the weak dispersed aliasing noise from strong focused signal energy in the focal domain, and, therefore, decreases the reconstruction quality. For the current example a velocity error of roughly $-200m/s$ to $+200m/s$ is tolerable for decent reconstruction results. This range is, of course, data-dependent.

Please note that in the remainder of this section again an accurate macro model is assumed, from which focal operators are generated.

3.3 The multi-level implementation

In the next step the limits of the current implementation are explored. Only one out of five receivers is active (Figure 3.16a). After 200 spgl1 iterations the focal domain in Figure 3.16b is derived. The measured traces are almost perfectly reconstructed, as can be observed in Figure 3.16d. However, the reconstruction in the gaps is insufficient for the second and third reflector. The amplitude of the reflectors is not laterally consistent and fades out in the data gaps. This is also represented by abrupt amplitude changes of the second and third event in the focal domain. This effect increases for even larger data gaps. To solve this issue the optimisation problem could be adapted. For instance a constraint for lateral consistency could be implemented.

The incapability of the focal reconstruction to handle the larger data-gaps in Figure 3.16 is due to the fact that the data representation for the second and third reflector is not sparse enough for a proper aliasing-noise suppression by the sparseness constraint. For a sparser data-representation the multi-level implementation of the double focal transformation is developed. With this implementation the loop is closed to the original definition of the data generation with the focal transformation (equation 2.7 in section 2.4 at page 27). Instead of only one, three virtual depth levels are chosen from which focal operators are generated. For the sake of convenience in this example the focal operators have the correct depth and velocity (i.e. the NMO velocity of the overburden). Therefore, every reflector can be focused to a band-limited spike in the corresponding focal domain.

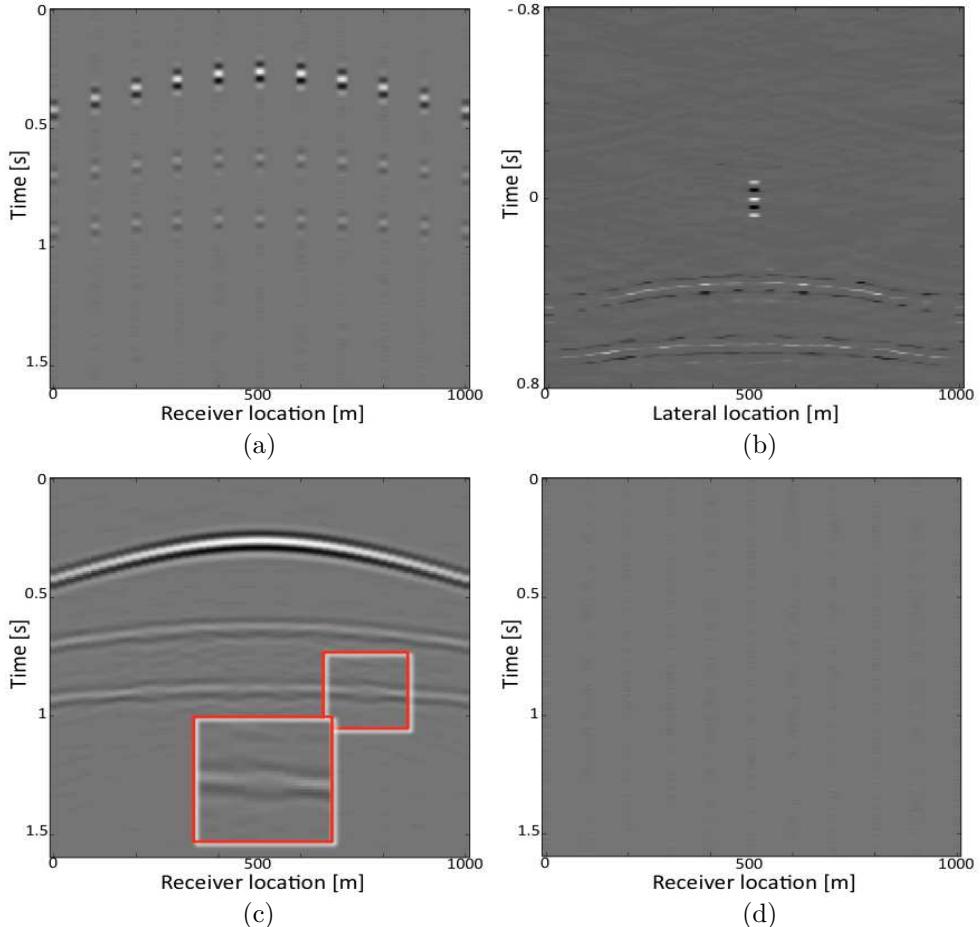


Figure 3.16: The middle section of the coarse input data for 100m receiver sampling (a). The corresponding focal domain (b), reconstructed data (c) and difference plot (d). In (c) an exemplary area where the reconstruction is insufficient is magnified (red box).

The optimisation to derive the focal domains reads:

$$\underset{\vec{x}_{\{1 \dots N\}}}{\text{minimise}} \sum_{n=1}^N \|\vec{x}_n\|_1 \quad \text{subject to } \|\vec{p} - \sum_{n=1}^N (\mathbf{L}_n \vec{x}_n)\|_2 \leq \sigma. \quad (3.18)$$

Please note that this formulation was already suggested in Kutsch and Verschuur (2012a). In the examples shown the number of virtual depth levels N equals 3. While the data-mismatch has to be lower than σ , the sum of the absolute value of the focal domains for each depth level is minimised. The solver basically distributes energy in the N focal domains in a way that, when they are propagated

to the surface with their corresponding focal operators and summed, they match the input data at the measured traces (second term of equation 3.18). In addition these focal domains must have a small absolute value (first term of equation 3.18). We consider again the data with large gaps from Figure 3.16a, being repeated in Figure 3.17a. We utilise the multi-level double focal transformation for the data reconstruction. As can be observed in Figure 3.17, the three reflectors are indeed focused to band limited spikes in their corresponding focal domains. In that way a physical interpretation is possible (please note the strong similarity to the focal domains in Figure 2.16 on page 30). On the other hand, for more complex data examples, or inaccurate focal operators, the energy in the focal domain can be distributed differently and in a way that doesn't allow any more a physical interpretation. After all, the focal domains are the result of an optimisation scheme.

However, the most important observation is that the coarse data-set in Figure 3.17a can be perfectly reconstructed after 200 spgl1 iterations solving equation 3.18. The reconstruction in Figure 3.17e is perfect with laterally consistent reflectors and the data residual is within the noise level (Figure 3.17f).

The reason for this successful reconstruction is that the input data is represented by three band-limited spikes in Figure 3.17, which is much sparser than in the single-level implementation in Figure 3.10 on page 50. Therefore, the sparseness constraint can more effectively suppress the spread-out, weak-amplitude aliasing noise in the focal domains.

In a real data case, assuming that for *every* true reflector the *perfect* focal operator would be available, the data could be represented by band-limited spikes in their corresponding focal domains, which allows data reconstruction for very large data gaps. On the other hand, the multi-level implementation increases the memory usage and computational cost. Hence, for real data examples with complex geology, focal operators are generated only for a couple of strategic depth levels. Mostly, these levels correspond to strong reflectors in the input data, as will be explained in chapter 4.

Please note that in the original version of the focal transformation (see Berkhou and Verschuur, 2006) subsurface operators were used that described two-way seismic reflection events, hence, both, propagation and reflection. In that respect it was very important to provide suited operators that represent the subsurface properly. Because these operators were difficult to be generated accurately enough, the usability of this method was limited. The new concept of double focal transformation, on the other hand, only provides the propagation operators between surface and subsurface and leaves the determination of the focal domain (extended reflectivity) to the inversion scheme. This makes it more stable and less vulnerable to errors in the propagation operators as demonstrated in Figure 3.15 at page 58.

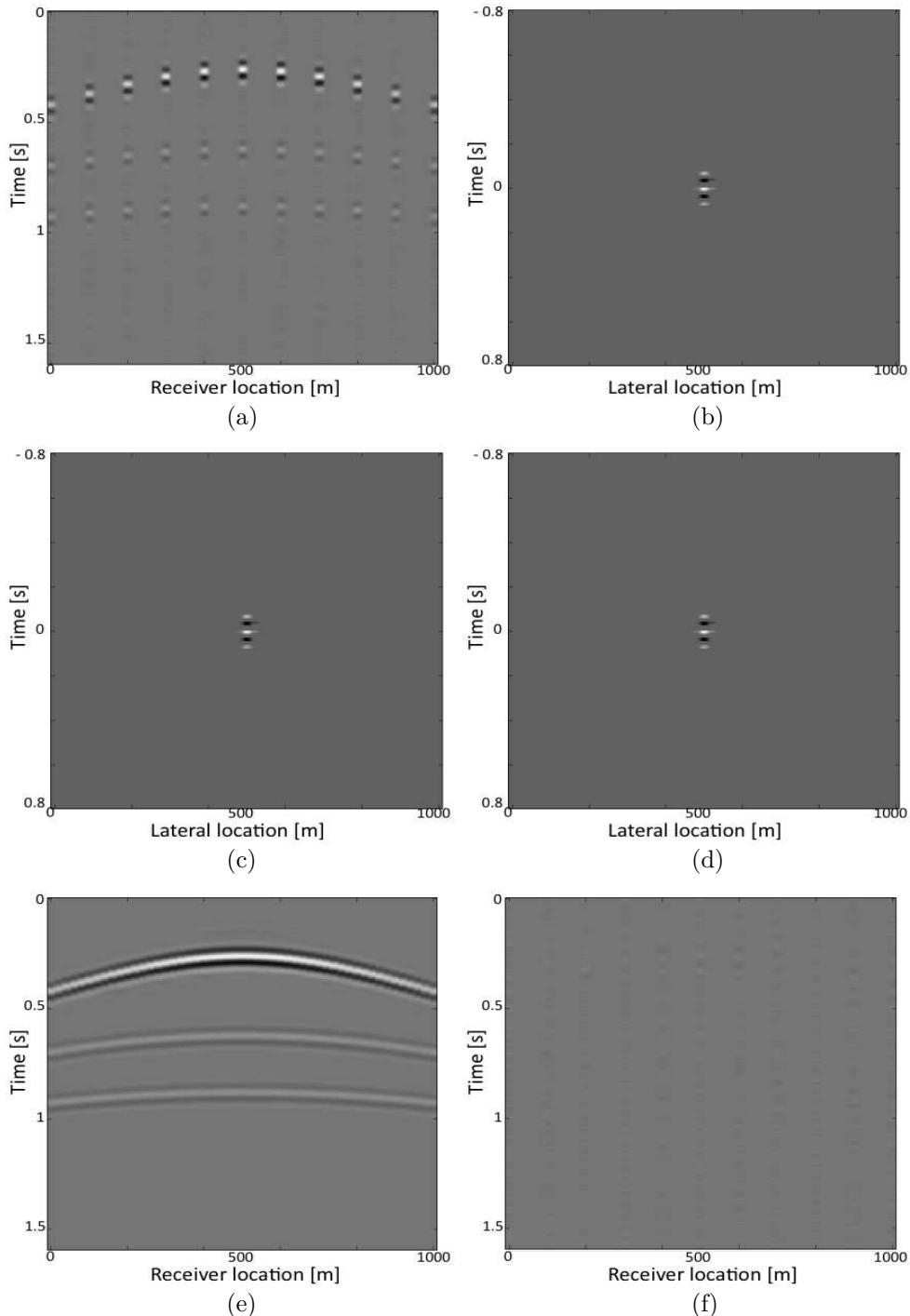


Figure 3.17: The middle section of the coarse input data with 100m receiver sampling (a). The middle sections of the corresponding three focal domains (b,c,d). The reconstructed data (e) and the difference plot (f).

Chapter 4

Examples

In this chapter the double focal transformation is used for the reconstruction of a 2D synthetic dataset and of a 2D marine field data set. Furthermore, the performance of the method on the field data is compared to the reconstruction results that were derived with state-of-the-art parabolic and linear Radon methods.

4.1 Reconstruction of synthetic data of a salt model

In this section the capabilities of the double focal reconstruction are demonstrated for more complex synthetic data. The subsurface model from which the synthetic data was generated is visualised in Figure 4.1 with velocities ranging from $1500m/s$ to $3500m/s$. Along the surface 101 receivers were distributed with a spatial sampling of $30m$. In this experiment every receiver also acts as a source, resulting in a fixed spread multi-gather acquisition of 101 source gathers with each 101 receivers. The temporal sampling is $8ms$ and the maximum frequency is $30Hz$. Please note that the reconstruction capabilities of the double focal transformation were extensively studied for this dataset in Kutscha and Verschuur (2012b).

In Figure 4.2a a representative shot record of the dense input data with source location at $1500m$ is visualised. Note that for the sake of a more realistic scenario, white noise was added with an amplitude of five percent of the maximum data amplitude. Furthermore, for the remainder of this section all data plots have been clipped to 20 percent of the maximum data amplitude in order to make the weak events better visible for the reader. As mentioned, this is a rather complex data scenario with strong and weak amplitudes, interfering events, locally varying dips and diffractions. Two properties are highlighted in Figure 4.2a. In the box "A" the reflection event is not continuing in the dense input data and in box "B" a very weak diffraction event is visible. To generate a situation where the input data is spatially aliased by a factor of two, every second receiver was removed.

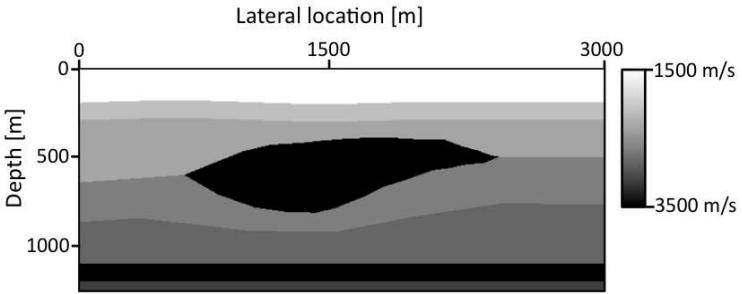


Figure 4.1: The simplified subsurface model of a salt structure. The black shape in the middle of the model represents a high-velocity salt inclusion.

Investigating the coarse input data in Figure 4.2b, it is very difficult to guess if there might be an non-continuous reflection event in "A" or a weak diffraction event in "B".

In Figure 4.3a the middle section of the focal domain is shown, which was derived from the dense input data after 200 spgl1 iterations. Here, a focal operator was used that was generated in an homogeneous velocity model of 1500m/s, from a depth-level of approximately the first reflector. Obviously, the first reflector is not compressed to a band-limited spike in the focal domain, but a spatially extended cross is visible in Figure 4.3a. This can be explained by the fact that the synthetic data were modelled with a finite difference algorithm that includes angle-dependent amplitude behaviour. Hence, the input data have different amplitudes for different reflection angles. For instance for high offsets the reflections can become post-critical and the reflection amplitude increases significantly. On the other hand, the focal operator has a decreasing amplitude with offset because its amplitude behaviour only takes geometrical spreading into account. In short, the cross in the focal domain appears because of different AVO (Amplitude Versus Offset) behaviour of the input data and the focal operators. A physical way to understand this cross in the focal domain, is to imagine that the focal operator has to be shifted along the cross and summed to generated the true first reflection event. In that way the AVO difference is compensated for. The greater the AVO-difference between data and focal operator the more spatially extended the cross in the focal domain will be. This cross is, of course, not sparse and, therefore, limits the capabilities of the focal reconstruction. In practice always a trade-off between sparseness and proper AVO representation is obtained.

A logical question is, if it would help the method of focal reconstruction to make the focal operators more complicated. One possibility is to adjust the AVO behaviour of the focal operators in a way to represent the input data better. This was investigated extensively but the discussion of this goes beyond the scope of this thesis. As a principal conclusion it can be stated that neither the usage of

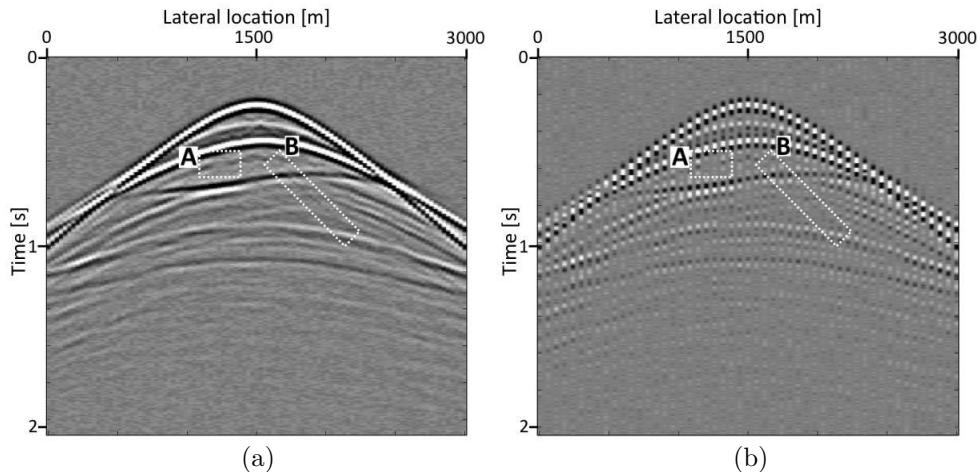


Figure 4.2: Synthetic shot record for a source at 1500m for the case of dense input data (a) and coarse input data (b). "A" highlights an in-continues reflection event and "B" a weak diffraction event.

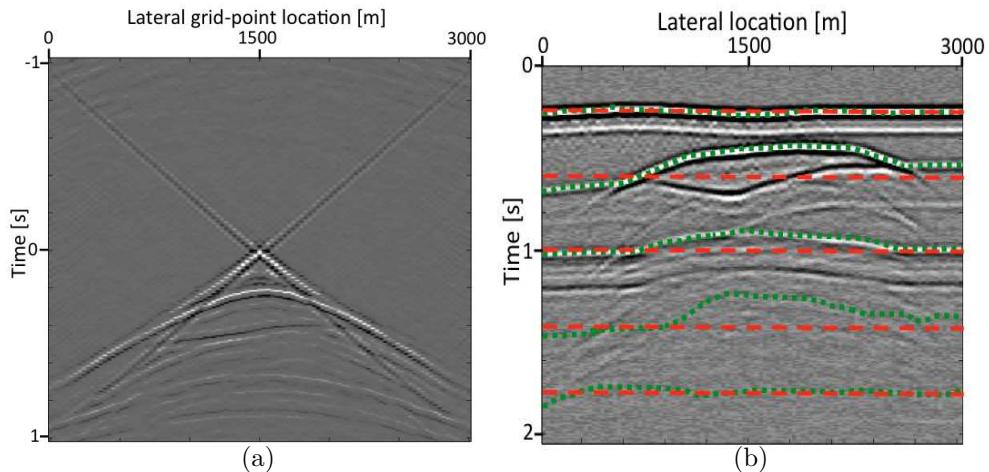


Figure 4.3: The focal domain (a) derived from the dense input data (Figure 4.2a) using a single focal operator corresponding to the depth level of the first reflector. In (b) the zero-offset section of the input data is visualised. The red and green dashed lines correspond to depth levels from which focal operators are generated utilising NMO velocities of the corresponding depth levels.

focal operators with more complex AVO behaviour nor the cancelling of the AVO behaviour of the input data by a data weighting (section 3.2.2 at page 52) did generally improve the reconstruction capability of the double focal transformation. The reason is that for real data often different events and different areas in the data have different AVO behaviour, which is difficult to account for with a limited number of focal operators. A second possibility is to generate focal operators based on an inhomogeneous velocity model. The theoretical framework to this idea is introduced in section A.5. However, experiments have shown that these more complex focal operators did not generally improve the reconstruction result for complex data, where the number of focal operators is much smaller than the number of reflectors in the subsurface. As a third option the depth levels from which the focal operators are computed are chosen to be laterally variant in depth to represent the true reflector locations better. For the sake of illustration, this is demonstrated in the following. The coarse input data (Figure 4.2b) will be reconstructed using the multi-level implementation of the focal reconstruction (section 3.3) for five depth levels.

In the first case these five focal operators were generated from flat depth levels. This is represented by the red dashed lines in the zero-offset representation of the input data in Figure 4.3b. It is clearly visible that the true subsurface structure is much more complex than just flat layers as is assumed by the focal operators. For instance, the true reflector is actually below the second depth level at a lateral location ranging from 0 to 400 meters. Further to the right the one reflector actually splits in two because of the salt-inclusion. On the other hand, the depth level of the second focal operator simply continues through this structure. Finally, from 2800 to 3000 meters of lateral location the actual reflector is above the assumed depth level.

In the second case a trade off is chosen, such that the depth of the focal operators is following the main reflectors and also the whole depth range is covered. This is indicated by the green dashed lines in Figure 4.3b.

The reconstruction result for both cases is visualised in Figure 4.4.

In both cases a proper data reconstruction is accomplished. For instance the non-continuous reflection event (marked by box "A") and the weak diffraction event (marked by box "B") have been preserved. Furthermore, the white noise is suppressed as it cannot be sparsely represented in the double focal domain. As a conclusion it can be stated that the more complicated focal operators did not improve the reconstruction result. Actually, the reconstruction in Figure 4.4b is slightly worse than in Figure 4.4a. Especially the first reflector at a lateral location between 2500 and 3000 meters is reconstructed less well in Figure 4.4b compared to Figure 4.4a. The data residual basically consists only of white noise, apart from one area in Figure 4.5b, where the first reflector was not completely reconstructed. However, keep in mind that the amplitude is clipped, such that also weak residuals look stronger than they actually are. A slight adjustment of the depth of the first reflector (change the first green dashed line in Figure 4.3b) could solve this issue. It has also to be mentioned that following the different depth

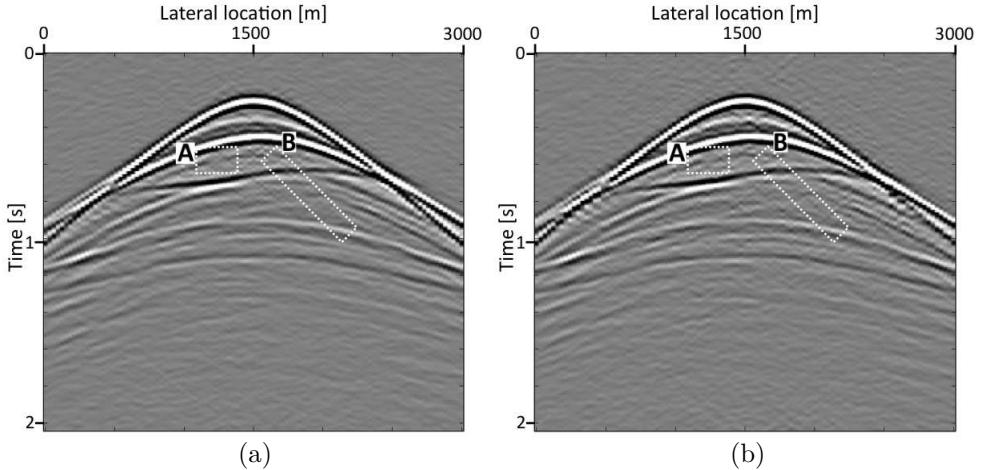


Figure 4.4: (a) Reconstruction result of the coarse input data (Figure 4.2b) using five focal operators based on five flat depth levels as indicated by the red dashed lines in Figure 4.3b. In (b) five focal operators were used based on varying depth levels as indicated by the green dashed lines in Figure 4.3b.

levels of the reflectors, the NMO velocity corresponding to the focal operator has not been adjusted. For example, when the depth level of the second focal operator follows the top of the salt, the corresponding NMO velocity should have been decreased. A more precise implementation could improve the capability of complex focal operators.

As a general conclusion of this section it can be stated that the usage of more complex focal operators, in the way that was explained above, does not generally improve the reconstruction result. Naturally, the combination of all three possibilities mentioned would allow an even more accurate representation of the data by the focal operators, but only for the parts of the data to which they are corresponding. However, in some cases very precise prior knowledge (e.g. for the case of time laps seismic processing) is available. Theoretically, it would be possible to use a very high amount of very precise focal operators to account for all parts of the data. Unfortunately, with the current implementation of the double focal transformation this would be impractical due to the high computational cost. Note that using a high amount of very precise focal operators would ensure a focusing of all energy in the centres of the corresponding focal domains. In that case only the central part of these focal domains would have to be computed, which would decrease the computational effort. This might be an interesting topic for further research.

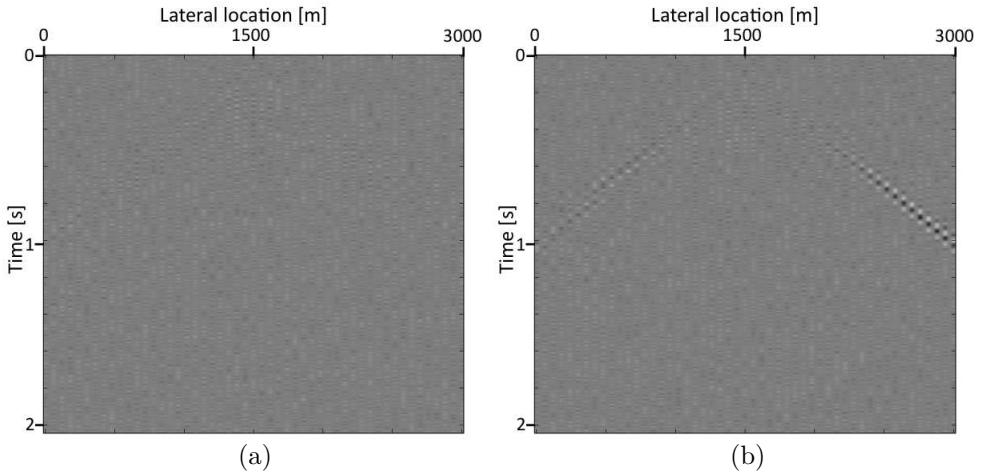


Figure 4.5: The difference plots (a and b) with respect to the dense true data corresponding to the reconstruction results of Figure 4.4 (a and b). The plots are generated at the same amplitude scale as Figure 4.4.

4.2 Reconstruction of a marine field dataset

In the following the double focal transformation is utilised for the reconstruction of a 2D marine data set from the Gulf of Mexico. The double focal transformation is principally defined for all geometries and sampling scenarios. However, currently it is implemented for fixed spread geometries based on a regular sub-grid only. The usage of fixed spread geometries is best suited for the formulation of the double focal transformation as matrix multiplications (see section 2.1 and equation 2.6 on page 26). This implementation is not only very intuitive but also efficient and fast when programs, such as MATLAB, are used that are specialised for fast matrix multiplications.

It is, of course, possible to translate any acquisition design to a fixed spread acquisition geometry. In Figure 4.6 this is visualised for the case of a conventional marine data acquisition. In Figure 4.6a the moving marine vessel is depicted that consecutively excites four wavefields that are detected by the corresponding receivers. This moving spread design is represented by the black dots in the data matrix in Figure 4.6b. Using reciprocity (exchanging source and receiver position leads to the same measurement) the grey dots are filled with information (Fokkema and van den Berg, 1993). By selecting only the data within the dashed box, the marine data can be treated as a fixed spread acquisition. Doing this changes the (moving spread) near-offset extrapolation problem to a (fixed spread) gap reconstruction problem. It is also possible to pad zeros for the missing traces and use the whole data matrix in Figure 4.6b. However, it is important not to take the padded traces into account as information. In that way the data consistency

is ensured in the inversion process. Note that in the latter case the data will also be extrapolated for the missing far-offsets and the missing information at the beginning of the seismic line.

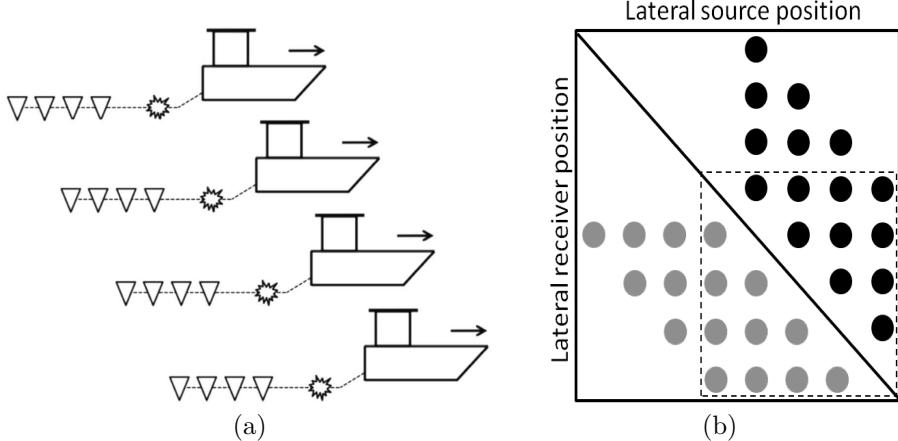


Figure 4.6: A marine vessel (a) moves from left to right and generates four shot-records. In (b) these shot records are displayed in the data matrix (black dots). Using reciprocity the data corresponding to the gray dots can be generated. The dashed line bounds a fixed spread situation. The main diagonal represents the location of zero-offset data, which are usually missing.

It has to be mentioned that in practice acquisition designs often are rather inefficiently represented by a stationary geometry. They would require very large data matrices that are very sparsely populated. In that case the matrix multiplications (summing every complete row with every complete column) in equation 2.6 can be replaced by summations of only the relevant elements (e.g. by for-loops). For the field data example we have chosen the option of data-sub-matrix selection (dashed line in Figure 4.6b). The fixed spread acquisition that was derived in that way for this example consists of 101 source gathers with each 101 receivers. It is sampled in the time domain with $\Delta t = 6ms$ and the spatial sampling is $\Delta x = 26.67m$. The maximum frequency is 60Hz.

4.2.1 Reconstruction of the near-offset gap

In Figure 4.7a a typical receiver gather of the data set is displayed. The near-offset gap has a range of 186m. Note that throughout this section the amplitude in the data plots has been clipped to 20 percent of the maximum of the input data for a better visibility of the weak events. In Figure 4.7b the gap was reconstructed by a conventional approach: the Common MidPoint (CMP) sorted data was corrected using Normal Move-Out (NMO) velocities. Afterwards, it was reconstructed with a parabolic Radon transform (Kabir and Verschuur, 1995) and inversely NMO corrected.

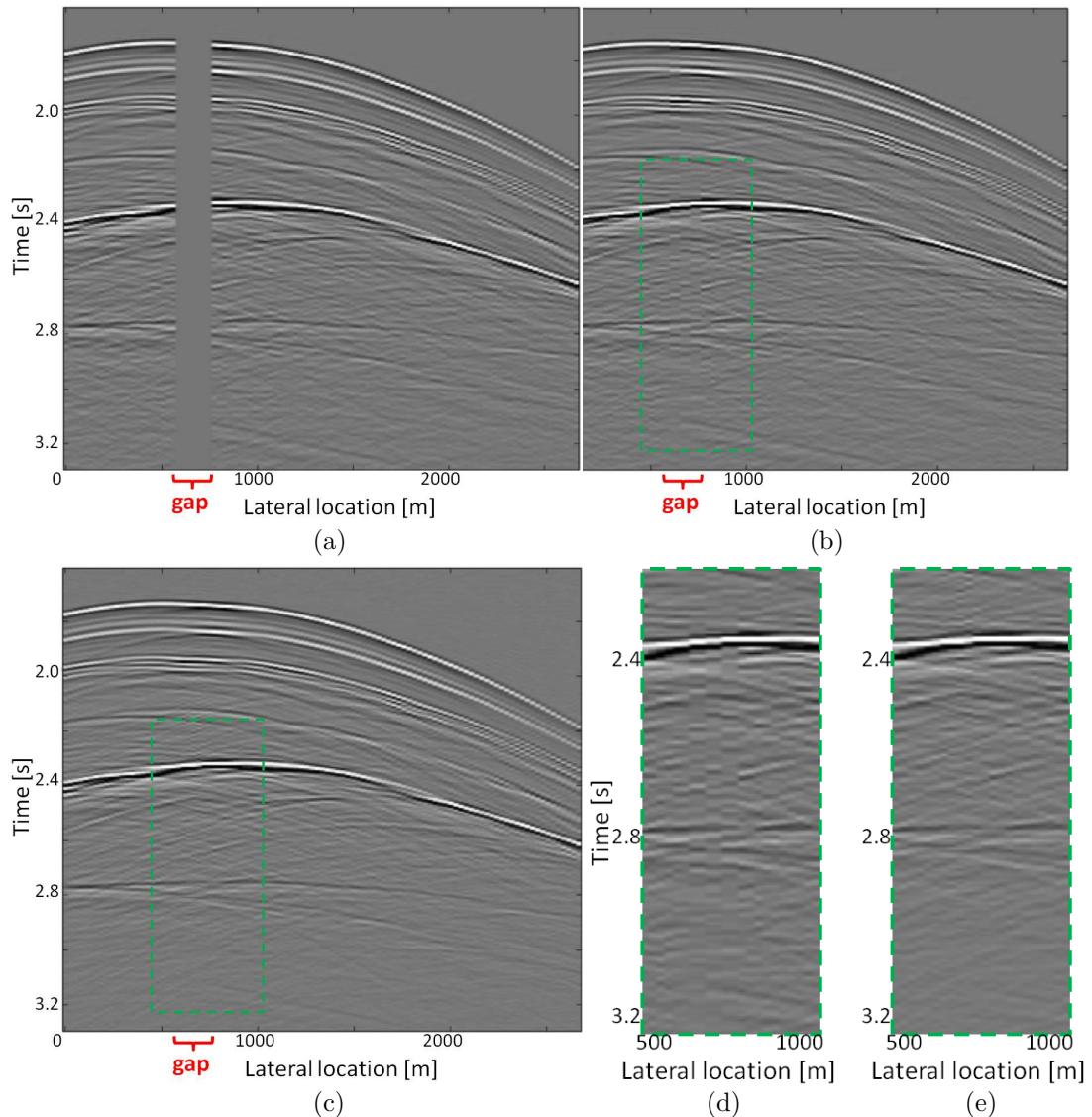


Figure 4.7: Typical receiver gather from a marine acquisition of the Gulf of Mexico. In (a) the near-offsets are missing. In (b) they were reconstructed by the parabolic Radon interpolation method. In (c) they were reconstructed using the double focal transformation. In (d) and (e) the green marked areas of (b) and (c) are magnified for a better comparison.

In Figure 4.7c the data was reconstructed with the multi-level double focal transformation (see equation 3.18 at page 60). In order to accomplish the best trade-off between sparse data representation and computational cost, four depth levels were used. From each level a focal operator was generated based on the corresponding NMO velocity. These focal operators were used for the multi-level implementation of the focal reconstruction. As done in Verschuur and Marhfoul (2005), the choice of these levels depends on the strongest reflection events in the data. However, the depth levels only have to be roughly in the range of strong reflectors and don't need to follow the varying depths exactly, as was demonstrated in section 4.1.

The focal reconstruction result in Figure 4.7(c) is more consistent and less blocky than the result in Figure 4.7(b) where the parabolic Radon method was used. Especially, in the time window from 2.5 to 2.7 seconds the energy of the diffraction events is uncorrelated in the parabolic Radon result in Figure 4.7(d), whereas this energy is distributed more consistently in the focal reconstruction result in Figure 4.7(e). The improved reconstruction is mainly due to the fact that for the focal reconstruction the whole 2D dataset was considered at once. In that way the redundancy of the data could be utilised. In contrast, for Figure 4.7(b) it is assumed that the data consists of certain parabolas in the NMO-corrected CMP domain. The summation along these paths leads to eventually sparsely represented energy in the model space. However, the reconstruction is performed for every CMP gather independently. In that way the data-redundancy is not fully utilised. Furthermore, if the data is rather complex, the parabolic assumption fails and the representation in the parabolic Radon space becomes less sparse. Often a smearing of energy in the model space is visible, which decreases the capability to suppress the aliasing noise. Finally, in the case of shallow water the stretch effects from the NMO correction distort the signal, yielding suboptimal reconstruction results with the parabolic Radon transformation (Verschuur, 2006).

In contrast, the double focal transformation just focusses energy in the focal domain. This is also true for less accurate focal operators, as demonstrated in section 4.1. In that case instead of a band-limited spike a little event containing the residual spatial phase is generated in the focal domain. However, aliasing noise is mapped to another area and is relatively spread out. Hence, as long as an undersampled reflector or diffractor is focussed by the forward double focal transformation, it will be reconstructed properly.

Please note that the result in Figure 4.7c was derived from the input data in Figure 4.7a using 200 spgl1 iterations. At this stage not yet all details are fully reconstructed. Especially the weak amplitude details for later times are not yet completely covered in Figure 4.7c. This is due to the fact that strong events are more important in order to obey the minimum data misfit criteria (second term in equation 3.18 on page 60). This behaviour was already investigated in section 3.2.2 at page 52. Allowing more spgl1 iterations can further improve the reconstruction result. The use of data-weighting was investigated. As expected, the weak events were reconstructed earlier compared to the implementation without data-weighting. However, for higher iteration numbers the total reconstruction

quality of the implementation without data weighting was superior to the implementation with data-weighting.

4.2.2 Reconstruction of strongly aliased data

To further test the reconstruction algorithm, a sampling operator was applied to the input data, deleting two out of three sources. The new source spacing is $80m$ and the largest data gap is $213.4m$. In Figure 4.8 the data matrix for this situation is visualised.

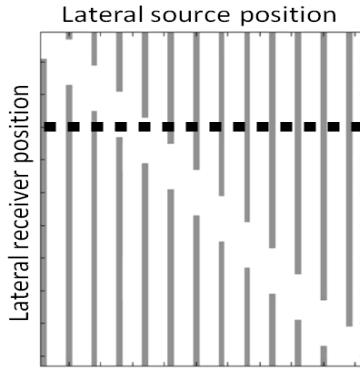


Figure 4.8: The data matrix representing the situation of the Gulf of Mexico data set with additionally missing sources. The grey columns indicate that the receiver sampling is dense and that the sources are coarsely sampled. Additionally, the near-offsets are missing. The black dashed row corresponds to the receiver gather displayed in Figure 4.9.

In Figure 4.9a the undersampled receiver gather is visualised in the space-time and in Figure 4.9c in the wavenumber-frequency domain. Strong spatial aliasing is clearly visible. After focal reconstruction using the same set-up as before the aliasing noise is suppressed (Figure 4.9d) and the data is properly reconstructed (Figure 4.9b).

4.2.3 Comparison with sparse linear Radon reconstruction

In order to evaluate the capabilities of the double focal transformation for data reconstruction, it is compared to a state-of-the-art, sparse, linear Radon reconstruction method, as described by Trad et al. (2003) and Verschuur et al. (2012). Two versions are considered: first a 2D reconstruction, that takes one gather $\mathbf{p}(x_{src}, t)$ and optimally represents this in the linear Radon space $\mathbf{x}_r(p_{xsrc}, \tau)$, where p_{xsrc} is the horizontal ray parameter and τ the intercept time, as was done in Trad et al. (2003). By including an additional sparseness constraint in this linear Radon space, the method tries to decompose the input data in 2D local plane

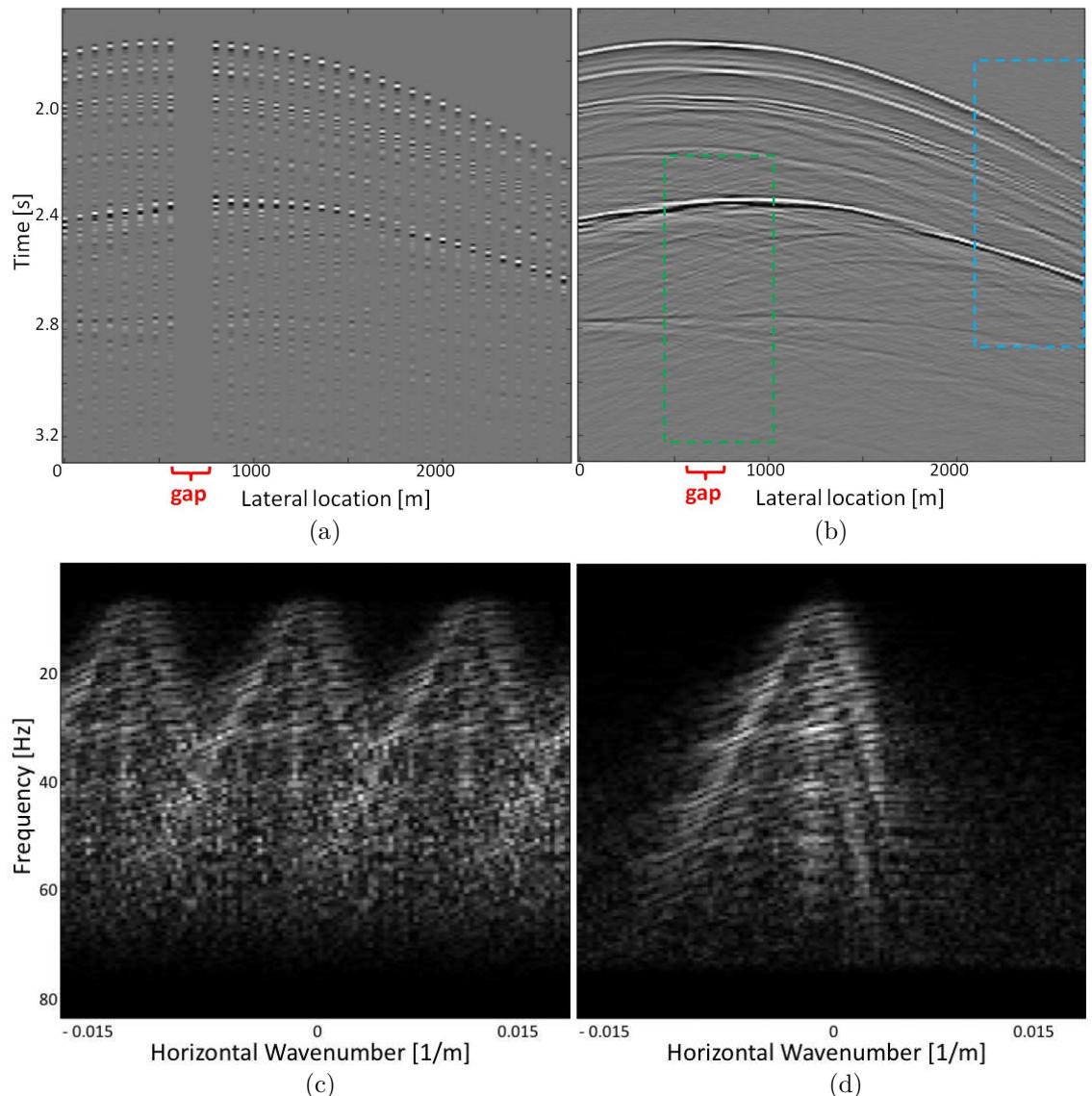


Figure 4.9: Strongly spatially aliased receiver gather in the space-time (a) and wavenumber-frequency domain (c). Reconstructed receiver gather using multi-level focal reconstruction in the space-time (b) and wavenumber-frequency domain (d).

waves. With an extended version of such reconstruction, the three dimensional data $\mathbf{p}(x_{rcv}, x_{src}, t)$ can also be represented in the double linear Radon domain $\mathbf{x}_r(p_{xrcv}, p_{xsrc}, \tau)$ where p_{xrcv} and p_{xsrc} represent the horizontal ray parameters related to the receiver and source coordinate, respectively, and τ is the intercept time. In Verschuur et al. (2012) such higher-dimensional sparse Radon transform is described. Again, an additional sparseness constraint in the Radon space will enforce a representation of the input data with the least amount of plane wave components. To improve the linear plane-wave assumption, such transform is usually done in local space-time windows, such that within each window the data can be assumed to consist of a limited set of plane wave components.

It has to be mentioned that during the generation of the reconstruction results with this method it became clear that it performs better when the Cauchy norm is used for sparse inversion instead of a l_1 norm (see equation 3.12 at page 48). First, the same receiver gather as displayed in Figure 4.9a was used as input for the 2D linear Radon reconstruction method. The result after 50 iterations is displayed in Figure 4.10a. It can be observed that the coarse sampling is handled rather well. Also linear events are reconstructed satisfactorily. However, the reconstruction in the large near-offset gap is poor and diffractions are hardly represented at all. Furthermore, reconstruction artefacts are visible, especially within an offset range up to 1000m.

In the next step, five receiver gathers were reconstructed simultaneously with the 3D linear Radon transformation. This dataset has now three coordinates (receiver, source, time). In Figure 4.10b the reconstruction result for the same receiver gather as in Figure 4.10a is displayed for comparison. It is clearly visible that the extension to a higher dimension (receiver coordinate) improved the quality of the reconstruction result as it makes more use of the redundancy of the data. There are much less artefacts and the reflectors are slightly sharper. However, still the near offset gap is not reconstructed satisfactorily and diffractions are hardly present. This can be easily explained. The strength of a sparse reconstruction method lies in its capability to compress seismic signal in the transform domain, whereas aliasing noise spreads out. The sparseness constraint suppresses the aliasing noise, and thus, leads to data reconstruction. This is certainly the case for the mid/far offset range of this dataset, where events are mostly linear and, therefore, can be represented efficiently in the linear Radon domain. Actually, the reconstruction result of the linear Radon transformation looks sharper and more consistent in the mid/far offset range (see Figure 4.10f) than the focal reconstruction result (see Figure 4.10e) in which at least the first reflector is still slightly aliased. Events with strong curvature, on the other hand, cannot be represented sparsely in the linear Radon domain, which prevents the successful suppression of aliasing noise through the sparseness constraint to the model space. This leads to a worse reconstruction of the curved strong reflector at around 2.4 seconds in the linear Radon result (Figure 4.10d) compared to the focal reconstruction result (Figure 4.10c). Also diffractions are hardly visible in Figure 4.10d whereas they are properly contained in Figure 4.10c. The linear event at around 2.8 seconds,

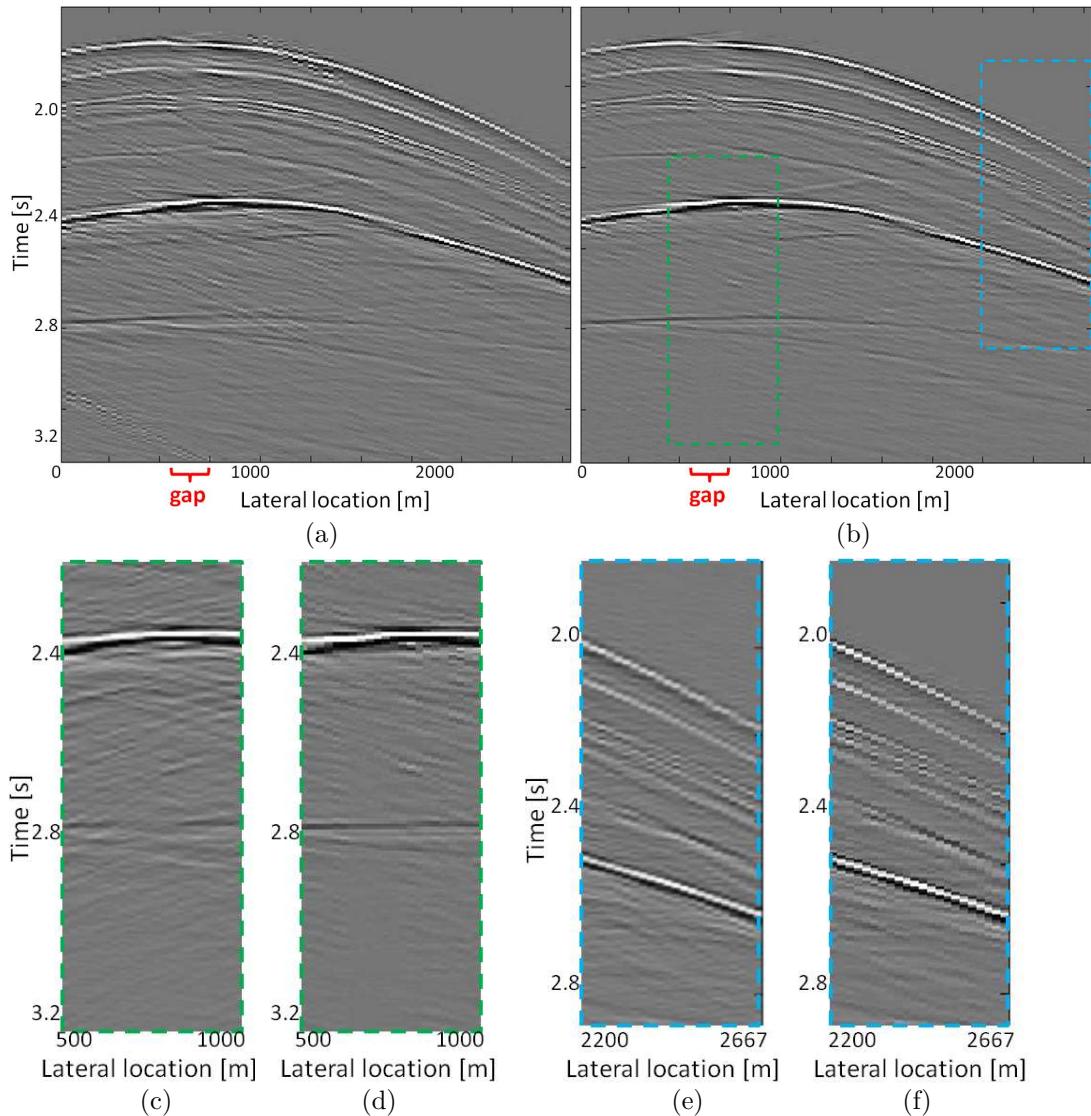


Figure 4.10: Reconstruction result utilising the sparse linear Radon transformation for one receiver gather in (a) and for five receiver gathers simultaneously in (b). The reconstruction results of the double focal transformation (Figure 4.9b) are displayed magnified for the near offset range (c) and the mid/far offset range (e). The reconstruction results of the multi-gather linear Radon transformation (b) are displayed magnified for the near offset range (d) and the mid/far offset range (f).

on the other hand, is clearly better reconstructed by the linear Radon method. By comparing the strengths and weaknesses of both methods for different parts of the data it becomes obvious that a combination of both methods into a hybrid one might be a very interesting field of study and will be discussed in the conclusions in section 7.2.2.

For the best linear Radon result, the data was subdivided in localised windows in the lateral direction. In that way globally curved events appear localised stepwise linear or close to linear. In that way the seismic data can be represented sparser, and hence, better reconstruction results can be achieved. In fact, for the result in Figure 4.10a and 4.10b the data was subdivided into two windows of 1500 meters (with a certain overlap) in the lateral direction, as this provided the best data reconstruction. Taking smaller windows improved the overall reconstruction slightly but the reconstruction in the near offset gap could then not be ensured any more. Obviously, the minimum window size must be a couple of times larger than the largest gap in the data. Furthermore, taking too small data windows can lead to the reconstruction of inconsistent events or events with discontinuities. It has to be mentioned that there are alternative ways to make use of the linear Radon domain for seismic data reconstruction. For instance it can be applied to NMO corrected CMP gathers (as was also done in Verschuur et al., 2012) or to COVTs (Common Offset Vector Tiles, see Cary, 1999), where the seismic events are generally represented more linear than in receiver gathers. However, to fully exploit such domains the earth is assumed to be laterally smooth, whereas the focal domain is aiming at reconstructing data from more heterogeneous structures.

Chapter 5

Focal reconstruction for 3D data

In this chapter we will - based on a simple numerical example - discuss the extension of the focal reconstruction method to the full 3D data case.

5.1 3D data representation

Usually, in seismic applications sources and receivers are distributed along the x, y -plane. The ideal case is depicted in Figure 5.1a, with sources and receivers densely sampled in both spatial directions. For this specific example, a model with only one horizontal reflector is considered, where the actual synthetic fixed spread acquisition design consists of 15 by 15 receivers and 15 by 15 sources with a spatial sampling of 20m in both lateral directions, a temporal sampling of 8ms and a maximum frequency of 30Hz. The corresponding monochromatic data matrix is depicted in Figure 5.1b. For the sake of a more realistic scenario, random noise has been added with a maximum amplitude of five percent of the maximum amplitude of the input data. Please note that even though the physical space in which the synthetic measurements are performed is three dimensional, (x,y,z) the resulting data have five coordinates (receiver-x; receiver-y; source-x; source-y; and time). Thus, reconstruction for such data volume is usually referred to as ‘5D reconstruction’ (Trad, 2009).

The formulation of the data matrix as it was introduced in section 2.1 is displayed in Figure 5.2a. Please see also equation 2.5 on page 24 and Figure 2.10 on page 21. Every cell of $\mathbf{P}(z_0)$ in Figure 5.2a consists of a complete 2D subset of 15 source gathers containing each 15 detectors with constant detector-y and source-y location. In Figure 5.1b the 15 by 15 cells are clearly visible, which represent 225 2D subsets. As mentioned in section 2.1, the main diagonal of $\mathbf{P}(z_0)$ represents a 3D zero-offset measurement. Bright colors in Figure 5.1b represent strong amplitudes and dark colors weak ones. The larger the offset in x - or y -direction of a certain trace, the weaker the amplitude, which is denoted by darker colours further away

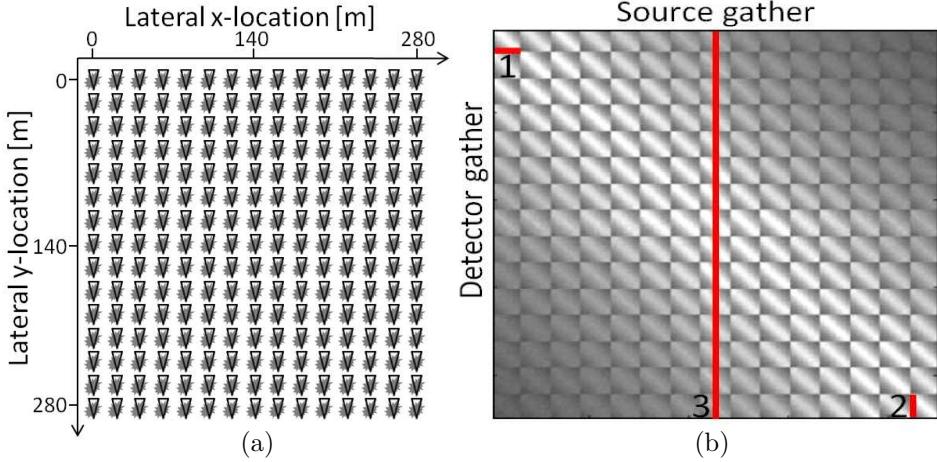


Figure 5.1: 3D fixed spread acquisition design (a) where grey stars represent sources and black triangles represent detectors, which are distributed along the surface. In (b) the corresponding monochromatic data matrix is displayed for the case of one reflector in the subsurface.

from the main diagonal of the data matrix and also within each submatrix, which leads to its block-Toeplitz structure of the data matrix.

Comparing Figure 5.1a and 5.1b and paying attention to the red lines in Figure 5.1b we define:

- Line-segment (1) represents a 2D detector gather. The sources of this gather correspond to the grey stars at the locations ($x = 0m : 280m$, $y = 0m$) in Figure 5.1a. The detector corresponds to the black triangle at the location ($x = 280m$, $y = 0m$).
- Line-segment (2) represents a 2D source gather with the detectors distributed at the locations ($x = 0m : 280m$, $y = 280m$) and the source placed at ($x = 0m$, $y = 280m$) in Figure 5.1a.
- Line (3) represents a 3D shot record including all detectors (all x - and y -locations in an areal distribution) and a single source at the middle ($x = 140m$, $y = 140m$) in Figure 5.1a.

5.2 Focal reconstruction for 3D data

In the next step, 7 out of 15 sources in x -direction and 7 out of 15 receivers in y -direction are removed from the dense input data, such that only 28 percent of the original traces remain. The resulting situation is typical for cross-spread

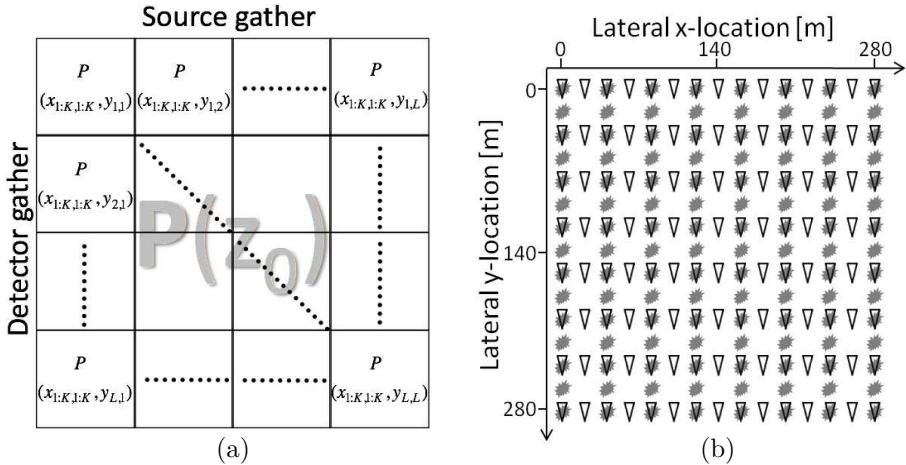


Figure 5.2: (a) The schematic formulation of the data matrix $\mathbf{P}(z_0)$. Every cell consists of a 2D subset with constant detector- y and source- y location. In (b) the fixed spread acquisition design is depicted with missing sources in the x -direction and missing receivers in the y -direction.

land acquisitions (Vermeer, 2002) and is displayed in Figure 5.2b. The corresponding coarse data matrix is visualised in Figure 5.3a. The red line in this figure corresponds to a monochromatic detector gather, where the detector is placed at $(x = 140\text{m}, y = 0\text{m})$ in Figure 5.2b and the 7 sources are excited along the positions $(x = 0\text{m}:280\text{m}, y = 140\text{m})$. The time-equivalent of this coarse monochromatic shot record is displayed in Figure 5.3c. Please note that the red line just indicates the position of a 2D subset of the 3D measurement that is chosen to evaluate the reconstruction quality. However, here the whole 3D dataset is reconstructed at once. If for instance in Figure 5.3a the red line would have been placed one cell lower (y location of the detectors changes to 20m), then all displayed input traces in Figure 5.3c would be empty and still a decent reconstruction could be accomplished, which wouldn't be possible with a 2D implementation. However, it is difficult to compare an empty input gather with a reconstructed gather. Therefore, this display option was neglected.

When applying the focal reconstruction in 3D, exactly the same formulation can be used as in the 2D case. In the current case the single level implementation is used. To derive a sparse focal domain from the coarse 3D input data, the basis pursuit denoise problem is solved as introduced in section 3.2.2 (see equation 3.13 on page 49). For the sake of convenience it is again formulated here:

$$\underset{\vec{x}}{\text{minimise}} \quad \|\vec{x}\|_1 \quad \text{subject to} \quad \|\vec{p} - \mathbf{L}\vec{x}\|_2 \leq \sigma, \quad (5.1)$$

where \vec{x} is the vectorised focal domain, \vec{p} the vectorised data and L the inverse

double focal transformation matrix for the 3D case. The formulation of the implicit function \mathbf{L} in 3D is equivalent to the one in 2D as it was laid out in the work flow in section 3.1 on page 36.

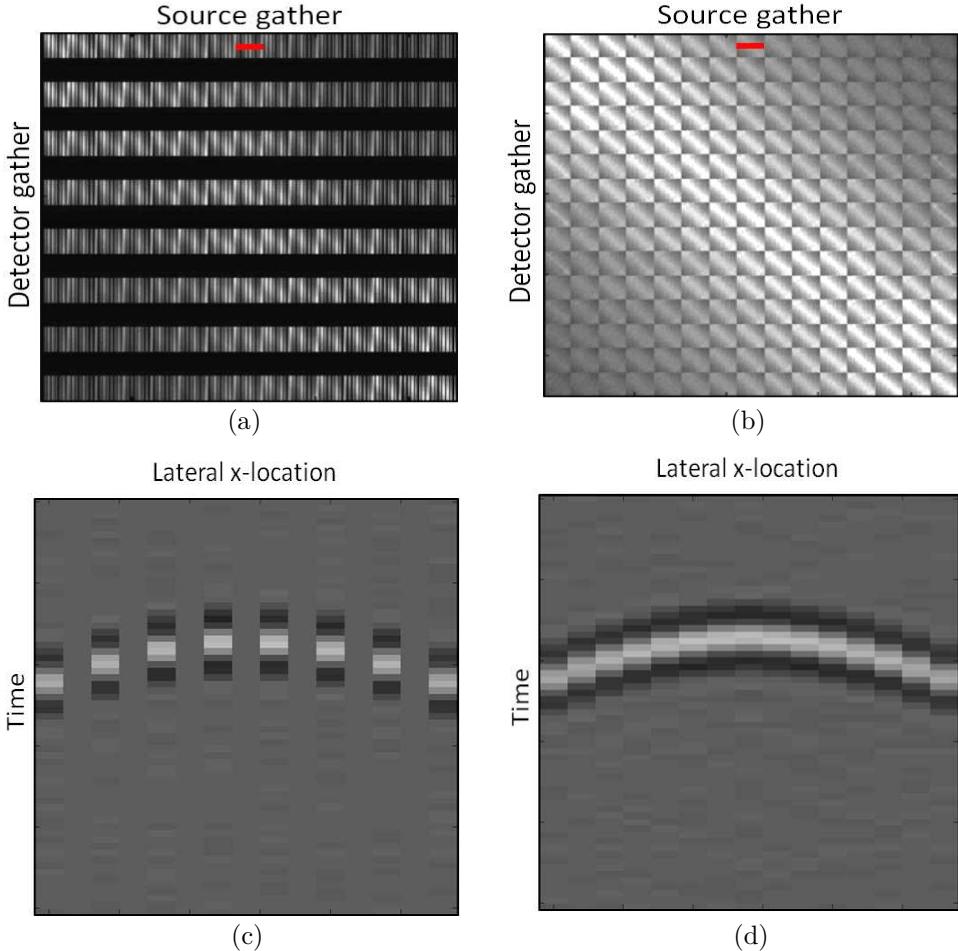


Figure 5.3: Data matrix (a) corresponding to the coarse input data as visualised in Figure 5.2b. In (c) a 2D coarse detector gather is displayed in the time domain. It corresponds to the red line in the monochromatic data matrix in (a). In (b) the reconstructed data matrix is visible and in (d) the reconstructed 2D detector gather corresponding to the red line in (b).

Again the l_1 norm of the the focal domain is minimised, while the data-misfit is kept below the noise-level σ . Like the data, the monochromatic 3D focal domain consists of L by L (15^2) cells each containing K by K (15^2) monochromatic traces. The time-domain equivalent is displayed in Figure 5.4. This is the output after

200 spgl1 iterations. It is important to realise that the whole 3D dataset of $K^2 * L^2$ traces is represented by a band-limited line in the focal domain. In the ideal case only $K * L$ traces in the focal domain (along the main diagonal) are non-zero.

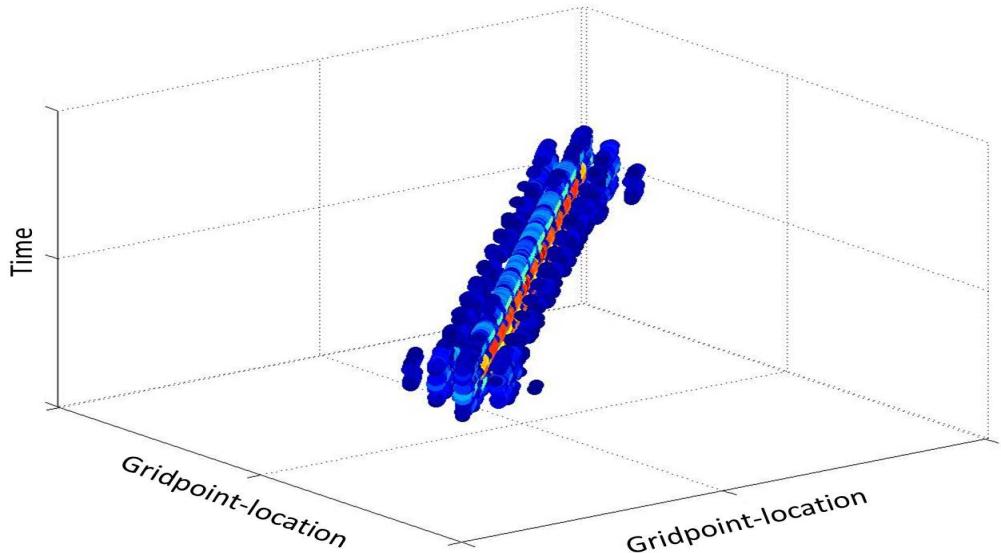


Figure 5.4: Scatter-plot of the 3D focal domain for the case of one reflector in the subsurface using the perfect corresponding focal operator. Blue dots represent weak values and red dots strong ones. Like in the monochromatic representation of the input data in Figure 5.3a also here both spatial axis are sub-divided in 15 cells in which the y -coordinate is constant.

Note that the data is much more sparsely represented in the 3D case than in the 2D case. In the 3D case K times L elements along the main diagonal of the 3D focal domain are used for the full 3D data representation for this simple example. Would the same data consist of L^2 2D experiments, where for L different source- y locations and L different receiver- y locations a 2D experiment is carried out, then L^2 focal domains with each K elements on their main diagonals would be required to create all the data. In conclusion, in 3D the same data can be represented by K times L traces for which with the 2D implementation K times L^2 traces were required. Please note that hereby it is already assumed that the data could be generated properly with a 2D implementation, where all receivers are at a y -location different from all the sources. This can be accomplished by adding a y -offset dependent term to the 2D \mathbf{W} operator.

Furthermore, the 3D implementation can make use of the redundancy of the data in both lateral directions, which increases the reconstruction capability.

The application of the inverse double focal transformation to the estimated focal domain generates the reconstructed data on the dense grid. The monochromatic reconstructed data matrix is displayed in Figure 5.3b. The red line corresponds to the reconstructed monochromatic detector gather corresponding to the under-sampled one in Figure 5.3a. The time domain equivalent is displayed in Figure 5.3d. The reconstruction is close to perfect and the data-misfit within the noise level.

Chapter 6

Utilising the linear Radon transformation

As one of the conclusions of previous chapters it can be stated that the sparser the input data representation is in a certain transform domain, the better aliasing noise can be suppressed. In this chapter the possibility of a sparser data representation, by the application an additional linear Radon transform in both spatial directions of the focal domain, is investigated. We will illustrate this process on a synthetic dataset, which consists, for this investigation, of a fixed spread acquisition corresponding to a model with one plane and one dipping reflector. In Figure 6.1a a typical shot record of the dense input data is displayed.

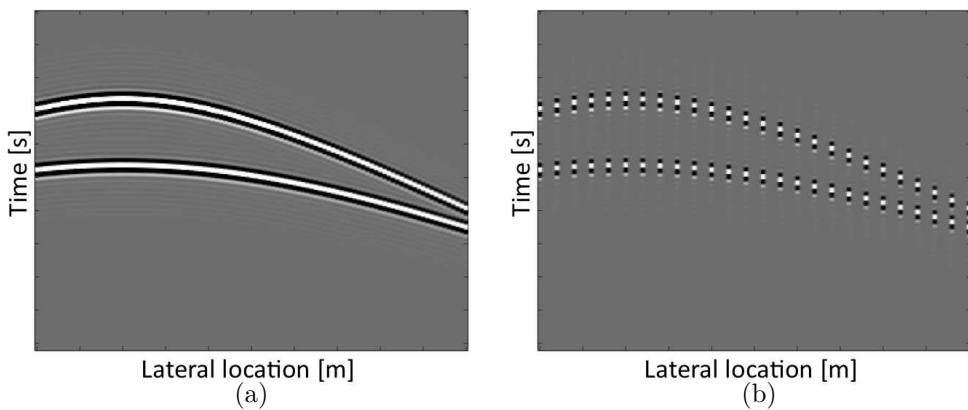


Figure 6.1: A typical shot record of the dense (a) and the coarse (b) input data.

In Figure 6.2a the corresponding focal domain is displayed for the case that the perfect focal operator for the first reflector is used. This display is done in a 3D fashion by making all data samples with values below a certain threshold invisible. As expected, the first reflector is focused to a diagonal event through the focal domain. The second reflector is not completely focused and has residual spatial phase.

6.1 The general work flow

In the case of a smooth subsurface, adjacent source gathers look very similar. As a result the subsequent slices in the focal domain will be very similar as well, as can be observed in Figure 6.2a. This correlation can be exploited for further data compression. With the following steps the data in the focal domain (Figure 6.2a) is transformed to the double-radon-focal domain (Figure 6.2b):

- For the sake of an easier physics-related understanding, the focal domain can be understood as an inversely propagated 2D fixed spread acquisition. Every (x_1, x_3) slice in Figure 6.2a represents an inversely propagated source gather with source and receivers at the depth of the first reflector z' . Accordingly, every (x_2, x_3) section can be understood as an inversely propagated receiver gather. Now the inversely propagated receiver gathers are resorted to offset gathers by applying the sorting operator $S(x_{off}, x_{rcv})$, which is the reverse operation to $S(x_{rcv}, x_{off})$ in equation 6.5. In that way the data-similarity is not along the diagonal in the focal domain any more, but parallel to the x_1 axis in Figure 6.2a.
- The application of a linear Radon transformation in the x_2 direction maps the linear event corresponding to the first reflector in the resorted focal domain, to a band-limited plane in (x_1, x_2) direction. The second event is transformed from a set of similar hyperbolas along the x_1 axis to a set of similar ellipses along the x_1 axis. The x_2 axis changes from lateral grid-point location to offset and from there to horizontal ray parameter p_{xoff} .
- The application of a linear Radon transformation in x_1 direction transforms the band-limited plane corresponding to the first reflector to a band-limited line in the middle of the transform domain ($p_{xsrc} = 0$) parallel to the x_2 axis. The set of similar ellipses along the x_1 axis is transformed to a single band-limited ellipse in the middle of the transform domain parallel to the x_2 axis. The x_1 axis changes from lateral grid-point location to horizontal ray parameter p_{xsrc} .
- The x_3 axis changes from time to intercept time (zero-offset time).

The result is the double-radon-focal domain $\theta(z')$ as displayed in Figure 6.2b and, in a zoomed fashion, in Figure 6.3a.

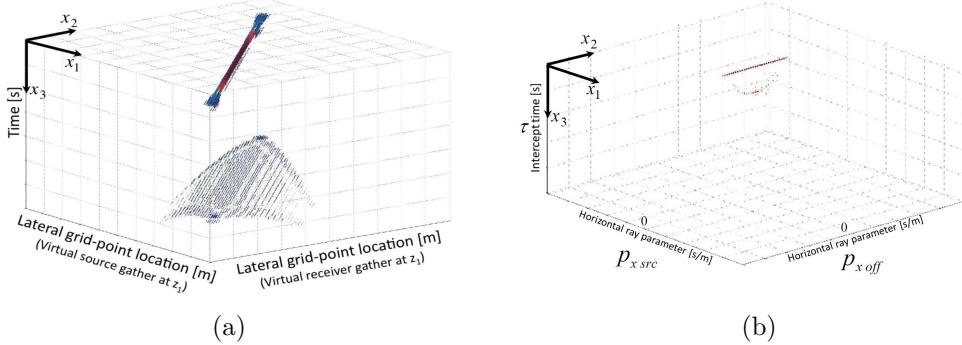


Figure 6.2: (a) The focal domain as a tensor for a 2D fixed spread acquisition and a two layer subsurface. In (b) the focal domain of (a) is transformed with the linear Radon transformation in both lateral directions to obtain $\theta(z')$. Note that a zoomed version of this data is displayed in Figure 6.3a.

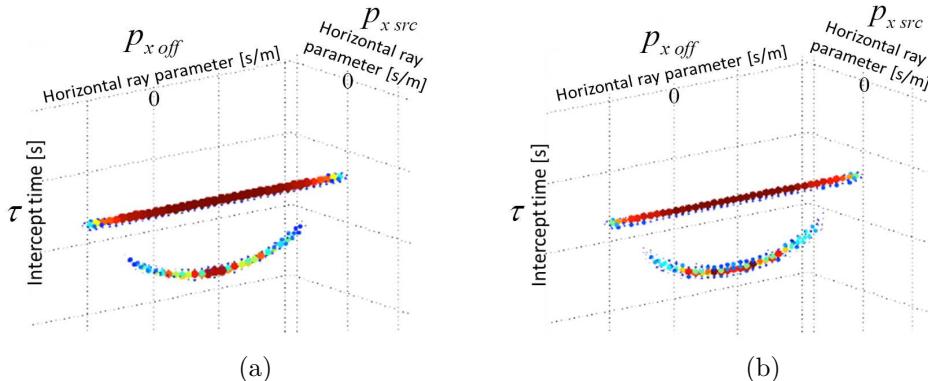


Figure 6.3: Zoomed double-radon-focal domain $\theta(z')$ after ten *spgl1* iterations for dense (a) and coarse (b) input data.

The sorting and the first radon transformation in the x_2 direction actually decreased the sparseness, but it generated a situation where the energy is distributed very homogeneously along the x_1 axis. Therefore, the second linear Radon transformation in the x_1 direction could focus all the aligned energy of the first flat reflector to the horizontal ray parameter $p_{xsrc} = 0$, and the aligned energy of the second dipping reflector to a single horizontal ray-parameter corresponding to this dip. No compression for the first event in the focal domain was accomplished, because it was transformed from a linear event to a linear event. But the energy corresponding to the second reflector, which consisted of a set of hyperbolic events

in the focal domain could be transformed to just a single band-limited ellipse. Note that utilising the linear Radon transformation also has the capability to compress the crosses in the focal domain (see section 4.1) that were generated by the AVO differences between the data and the focal operators.

6.2 Implementation

In the previous section the forward double-radon-double-focal transformation has been introduced schematically. However, in practice this transformation cannot be performed directly. Instead, the inverse transformation (equation 6.5) is defined and the desired model space is estimated iteratively by sparse inversion.

First, the discrete version of the linear Radon transformation in the time domain (see Schultz and Claerbout, 1978; Kostov, 1990) is formulated as:

$$m(p_x, \tau) = \sum_x d(x, t = \tau + p_x x), \quad (6.1)$$

where $d(x, t)$ is the original data in the offset-time domain. The mapping from the data space to the model space is performed by summations along straight lines $t = \tau + p_x x$ for different intercept (zero-offset) times τ and different horizontal ray-parameters p_x representing the slopes of these lines. Transforming equation 6.1 to the frequency domain leads to:

$$M(p, \omega) = \sum_x D(x, \omega) e^{i\omega p_x x}, \quad (6.2)$$

where capital characters represent data in the frequency domain. The summation in equation 6.2 can be denoted as well by matrix-vector multiplications:

$$\vec{M} = \widetilde{\mathbf{L}}_r \vec{D}, \quad (6.3)$$

where \vec{D} is the monochromatic vectorised version of the data space and \vec{M} the monochromatic vectorised model space. The transform operator is defined as:

$$\widetilde{\mathbf{L}}_r = \widetilde{\mathbf{L}}_r(\vec{p}_x, \vec{x}) = \begin{pmatrix} e^{i\omega p_{x1} x_1} & \dots & e^{i\omega p_{x1} x_n} \\ \vdots & & \vdots \\ e^{i\omega p_{xm} x_1} & \dots & e^{i\omega p_{xm} x_n} \end{pmatrix}, \quad (6.4)$$

where $\widetilde{\mathbf{L}}_r$ is the forward Radon transformation matrix. To distinguish it from the double focal transformation it is indicated with the subscript (r) . Also, to indicate that data and model space are in the frequency domain, a tilde is added. n and m represent the number of traces in the original data domain and in the corresponding Radon transform domain respectively. To increase the resolution in the Radon domain a larger number m can be selected. If the linear Radon transformation is not only carried out for a single gather but for multi-gather

data, then \vec{D} and \vec{M} change from vectors to matrices \mathbf{D} and \mathbf{M} containing each the multi-gather data set for one frequency. The model of the double-radon-double-focal transformation (represents the inverse transformation) reads:

$$\begin{aligned}\mathbf{P}(z_0) &= \mathbf{W}(z_0, z')\mathbf{X}(z_n)\mathbf{W}(z', z_0) \quad , \text{with} \\ \mathbf{X}(z_n) &= S(x_{rcv}, x_{off}) \left\{ \widetilde{\mathbf{L}_r}^H(\vec{p}_{x\ off}, \vec{x}_{off})\Theta(z')\widetilde{\mathbf{L}_r}^*(\vec{p}_{x\ src}, \vec{x}_{src}) \right\},\end{aligned}\quad (6.5)$$

where $\mathbf{P}(z_0)$ is the monochromatic 2D data at the surface as a matrix. $\mathbf{W}(z_0, z')$ and $\mathbf{W}(z_0, z')$ are the focal operators that propagate sources and receivers from depth level z' to the surface z_0 . The outer part of the equation is analogue to the original definition of the model of the double focal transformation. The inner part of equation 6.5 contains the sorting operator $S(x_{rcv}, x_{off})$, which sorts the x_2 axis from offset gathers to receiver gathers. $\widetilde{\mathbf{L}_r}$ is the forward Radon transformation as it is defined by equation 6.4. The $(\cdot)^H$ denotes the conjugate transpose (inverse Radon transformation) and the $(\cdot)^*$ denotes the conjugate. $\Theta(z')$ is the desired model space.

All properties in equation 6.5 are denoted in the frequency domain. However, like for the traditional double focal transformation, first, the input data is transformed from the time to the frequency domain. Finally, after estimating $\Theta(z')$, it is inversely Fourier transformed to the (intercept-) time domain in order to obtain $\theta(z')$ as it is displayed in Figure 6.2b. Like for the traditional focal reconstruction in Section 3.2.2, also here $\theta(z')$ is the solution to a Basis Pursuit Denoise problem and is estimated with the spgl1 solver from the input data in the time domain $\mathbf{p}(z_0)$. Again, a model space is looked for, which is as sparse as possible and results in a minimum data-misfit at the known traces.

6.3 Results

With the sparseness constrained optimisation using the double-radon-double-focal transformation, the transform domain for the dense input data (Figure 6.1a) can be obtained (Figure 6.3a). It is clearly visible that the transform domain (Figure 6.3b) corresponding to the coarse input data (Figure 6.1b) doesn't show additional aliasing noise compared to the dense case in Figure 6.3a, which proves a successful reconstruction.

In Figure 6.4 the reconstruction results are compared for the double focal transformation and the double-radon-double-focal transformation each for ten spgl1 iterations. It is clearly visible that with the double focal transformation the first reflector is properly reconstructed but the second one, for which no focal operator was available, is less well reconstructed. On the other hand, with the double-radon-double-focal transformation both reflectors are reconstructed equally well. Thus, the extra linear Radon transformation improves the reconstruction for reflection events which are not completely focused by the focal operator. This

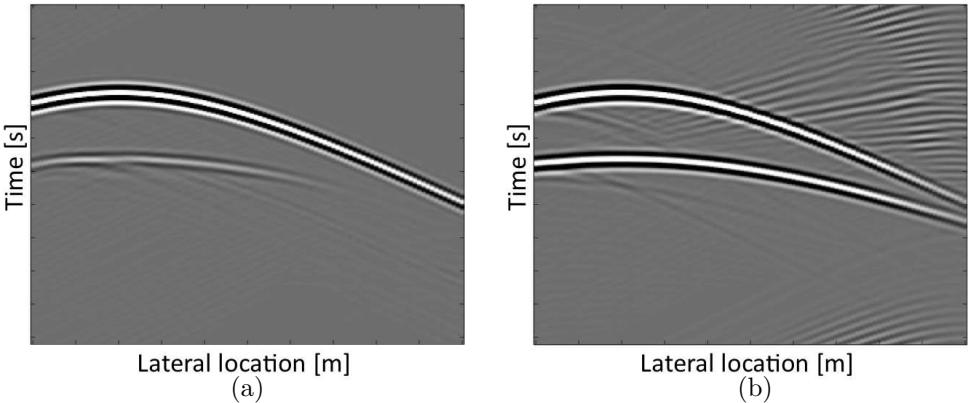


Figure 6.4: The reconstruction of a typical shot record after ten *spgl1* iterations using the double focal transformation (a) and the double-focal-double-radon transformation (b).

provides the opportunity to use less depth levels in the multi-level implementation of the double focal transformation (smaller M in equation 2.5 at page 24) in combination with a double-radon transformation.

However, some artefacts were introduced as well in Figure 6.4b. This is the result of an undersampling in the linear Radon domain, which means that the step-size Δp of the horizontal slowness, or ray-parameter, is not small enough to satisfy the sampling criterion:

$$\Delta p \leq 1/(\omega_{max} h_{max}), \quad (6.6)$$

where $\omega_{max} = 2\pi f_{max}$ gives the maximum angular frequency and h_{max} gives the maximum range of offsets covered by the input data. However, decreasing Δp increases the size of the inversion problem and the computational cost.

Another possibility to improve the contribution of the double Radon transformation to the double focal transformation is to use moving spread acquisition instead of fixed spread acquisition. The reason for this suggestion becomes clear by investigating the double-radon-double-focal transformation process in more detail. By sorting the inversely propagated receiver gathers (focal domain) to offset gathers, a strong similarity along the x_1 axis was generated. However, the similarity smoothly varies from the first virtual receiver gather, which has only negative offsets, to the last one, which has only positive ones. If, on the other hand, a moving spread configuration instead of a fixed spread configuration is used, adjacent offset gathers would actually be identical for the case of a plane reflector. In the case of a dipping reflector the offset gathers would be identical along a certain angle and, therefore, could still be represented by a single ray-parameter in the x_1 direction. It can, hence, be expected that we will have a sparser representation of the data from a combination of a moving-spread double focal transformation with a double linear Radon transformation.

Chapter 7

Conclusions, Discussion and Recommendations

In this chapter the conclusions of this thesis are summarised and recommendations for future work are discussed.

7.1 Conclusions

The double focal transformation is a powerful tool for data reconstruction, because it can make use of prior subsurface information. The focal domain can be understood as a more general form of the reflectivity operator that also has extent in the time domain (section 2.4). In that way the focusing effect of the propagation operator can be utilised for seismic data reconstruction even if the focusing is not perfect (section 3.2.2).

It has been shown that highly spatially aliased data can be properly reconstructed as well as large data gaps (chapter 4). Satisfactory reconstruction results were obtained for 2D and 3D synthetic data and a 2D subset from marine field data (Gulf of Mexico), especially in terms of proper reconstruction of the curved events, including diffractions.

In section 3.1.1 it has been demonstrated that focal reconstruction can be implemented in a rather direct way. The missing traces were considered as information and, therefore, aliasing noise was generated in the focal domain that had to be suppressed manually. However, in section 3.2.2 it has been shown that formulating the focal reconstruction as a sparseness-constrained inversion problem achieved superior results. The computational cost of such an iterative approach is substantially higher than with the direct approach. This disadvantage is compensated for by the better reconstruction result and the fact that in general and specifically in industry, user interaction becomes increasingly expensive compared to the costs for longer computations.

In section 3.3 it has been demonstrated that the multi-level implementation allows a stronger focussing of the seismic data in the focal domain and, hence, a better signal to noise separation.

Finally, in comparison to state-of-the-art sparse linear Radon reconstruction, it turned out that the double focal transformation was better able to properly reconstruct curved, aliased events and large data gaps. However, the linear Radon transformation showed slightly better reconstruction for the linear events at mid to far offsets. This opens the door for a hybrid focal-Radon approach.

7.2 Discussion and Recommendations

7.2.1 The effect of quality and quantity of the utilised focal operators

For the reconstruction examples in this research mainly focal operators are used that represent propagation between a flat surface and a flat reflector through a homogeneous layer. This is not a requirement. In general, any kind of propagation operator could be used. Adding more degrees of variation to the focal operator, e.g. propagating between complex reflectors, propagating through inhomogeneous layers or propagation with a more realistic AVO behaviour in general, helps to describe certain reflectors with less model space coefficients. This translates into a sparser representation of these reflectors and, therefore, into an improved reconstruction capability for the corresponding reflector. However, usually only a couple of focal operators are chosen, based on the most prominent reflectors in the data. In that way each focal operator is used to focus more than only one reflector. By attempting a perfect focusing for one of these reflectors the focusing of the others is not necessarily improved and even the opposite might be the case. Hence, to use more accurate focal operators in a case where there are more reflectors than focal operators does not generally improve the reconstruction result (see section 4.1).

To use a large amount of focal operators to describe more reflection events in the data is not very attractive either, because of two reasons. First, the computational expense would increase accordingly, assuming the current implementation of the double focal transformation, where every focal operator generates a focal domain with the same size as the dense data. However, different implementations are possible. For instance, considering that only the diagonals close to the main diagonal of the focal contain information. Hence, skipping the computation of the parts further away from the main diagonal (by loops instead matrix multiplications) has the potential to speed up the algorithm and lower the computational cost. This might be an interesting topic for further research. On the other hand, the second and more important reason, why using a large amount of focal operators does not necessarily guarantee a better result, is that data reconstruction is a preprocessing step. At this stage, normally, a high quality subsurface model is not available. To represent an inaccurate subsurface model more accurately by

utilising more focal operators is just not very efficient as the total focusing of the data is not improved. Hence, the gain of utilising more focal operators is related to the quality of the prior information they are based on.

However, in some cases a rather precise subsurface model is available, like for time-lapse seismic processing. Therefore, there is still scope in investigating the effect of a larger number of focal operators with small corresponding focal domains in future work. Also, a further comparison to parabolic and hyperbolic Radon implementations, particularly multi-domain implementations (see for instance Trad, 2003), in terms of efficiency, might be interesting.

7.2.2 Utilise the linear Radon transformation in combination with the double focal transformation

When a sparse data representation is used for seismic data reconstruction, it is intrinsically assumed that the data is well represented by the involved basis functions. If, for example, sparse Radon implementations are used, it is automatically assumed that the data can be explained by a couple of linear events, parabolas or hyperbolas. Correspondingly, sparse Fourier implementations assume that the data can be explained by a small number of plane waves. A sparse double focal transformation assumes that the data can be explained by a limited number of approximated reflection events, and so on. Real seismic data, however, don't care about these assumptions and are usually complex and have many different features. It is, therefore, important that a model is used that explains as many parts as possible of the seismic data, with a minimum amount of model coefficients. Hence, a trade-off has to be found that allows the representation of the most important events in the data as sparse as possible, in order to suppress aliasing noise and, at the same time, does neither suppress the details in the data nor events that don't fit the model well. This, of course, applies to the double focal transformation as well. One possibility to improve the reconstruction capability is to combine it with other transformations like the linear Radon transformation in order to allow the efficient reconstruction of linear events.

Hybrid implementation

The WRW model and the concept of double focal transformation both describe reflections events. However, linear events, like direct arrivals in the data, are not included in this model. As a result these events are not represented sparsely in the focal domain, because a combination of many focal operators is needed to represent a linear event in the data domain, which degrades the reconstruction capability of the double focal transformation for these events. The same is true for the far-offset range of reflection events, where the events are close to linear. An example for the degraded reconstruction capability of the double focal transformation compared to the linear Radon method for the far offset range of real seismic data was discussed in section 4.2.3 and was shown in Figure 4.10e. On

the other hand, it also became clear that the linear Radon reconstruction has its limitations for curved events and complex data.

Another criteria for comparison is the computational effectiveness of both methods. For the sake of convenience this analysis is demonstrated on a simple 2D fixed spread acquisition that is stored in a data cube as was shown in Figure 2.13 on page 26. Transforming this data cube with a 3D linear Radon transformation creates a 3D Radon domain of a certain size. Adding more sources or receivers to the dataset will increase the size of the corresponding Radon domain and the computational effort to perform the transformation linearly, especially when the data is subdivided into windows. This is not the case for the double focal transformation. To better visualise this please see Figure 2.19 at page 33 where the inverse double focal transformation is shown in the frequency domain. To transform the seismic energy from a certain depth level to the surface it is propagated to the sources and receivers with the corresponding focal operator. This is accomplished by multi-shift convolution that is equivalent to matrix multiplications in the frequency domain. Let's consider the case that the number of sources and receivers doubles in this example, hence the new dataset is four times larger. Also the focal domain and the focal operators will have four times their previous size. Hence, the memory requirement grows linearly. However, the computational effort does not. The data matrix in Figure 2.19 is computed by matrix multiplications of the monochromatic focal operators and the focal domain. If these matrices double in length and height (twice the amount of sources and receivers) then the number of row-vector multiplications would quadruple, but also the length of each row and vector would be doubled. Hence, the computational cost of the double focal transformation grows super-linearly with the data size.

The logical step is to utilise the advantages of both methods in one scheme. In order to be computationally and memory-wise efficient the two methods could be applied to different parts of the data in a separate way. For instance, the double focal transformation could be used for the near offset range and for complex areas (smaller part of the data), whereas the linear Radon transformation could be applied to the mid and far offset range (larger part of the data). However, there is usually not a clear distinction of areas where either linear or curved events are present. In that case the double focal transformation can be combined with the linear Radon transformation as a hybrid transformation.

The corresponding optimisation problem can be formulated as:

$$\begin{aligned} & \underset{\vec{x}, \vec{x}_r}{\text{minimise}} \quad ||\vec{x}||_1 + \alpha ||\vec{x}_r||_1 \\ & \text{subject to} \quad \|\vec{p} - \mathbf{L}\vec{x} - \mathbf{L}_r^H \vec{x}_r\|_2 \leq \sigma. \end{aligned} \tag{7.1}$$

Here \vec{p} is the vectorised input data, \vec{x} the vectorised focal domain and \mathbf{L} the inverse double focal transformation. The forward linear Radon transformation \mathbf{L}_r can be formulated as a function that consists of several steps analogous to the double focal transformation (see the list on page 36):

1. Resort the input data from a long vector \vec{p} to a 3D tensor with the dimensions (time, receiver-coordinate, source-coordinate).
2. Transform this tensor from the time to the frequency domain and multiply every frequency slice with the linear Radon matrix \mathbf{L}_r , which is defined by equation 6.4 at page 86.
3. Transform the output from the frequency domain to the time domain to obtain the 3D tensor. If the linear Radon transformation was applied in the receiver direction then the dimensions of the tensor are: intercept-time, horizontal slowness, source-coordinate.
4. Finally, the tensor is vectorised to obtain the model vector \vec{x}_r .

The inverse linear Radon transformation \mathbf{L}_r^H correspondingly transforms the model vector \vec{x}_r to the data vector \vec{p} by executing the steps in the above list reversely and utilising the inverse Radon matrix $\widetilde{\mathbf{L}}_r^H$. Please note that equation 7.1 actually represents two simultaneous transformations with two model spaces. The first part of this equation ensures that the total data is represented as sparsely as possible in these two model spaces. Naturally, linear events can be represented very sparsely in the linear Radon domain (\vec{x}_r) and reflection events at the smaller offsets can be represented very sparsely in the focal domain (\vec{x}). With the scalar α the balance between these two domains is regulated. For example, if it is desired that more events of the data are represented in the linear Radon domain the scalar α can be decreased. The second part in equation 7.1 represents the minimum data-misfit criterion.

In Figure 7.1 a synthetic shot record is displayed. The area with mostly curved events and the area with mostly linear events are each indicated by different colours. Note that it is, of course, possible to apply to the near- and mid-offset part (green area in Figure 7.1) of the data only the focal reconstruction and to the far-offset part and the deep reflections (red area in Figure 7.1) only the linear Radon reconstruction. This certainly is computationally less expensive. But then the user interaction is increased, which also increases the chance of misinterpretations. Furthermore, linear and curved events might be overlapping, which makes it difficult to separate them into windows. On the other hand, with a hybrid implementation the sparseness constraint will automatically choose the adequate transform domain to explain the data, such that an optimum distribution between both domains is obtained.

Consecutive implementation

The most important property that is utilised in this thesis is that seismic data can be represented sparsely in the focal domain. A sparser representation potentially leads to better reconstruction results. As demonstrated in section 3.3, the sparser data representation using the multi-level implementation, compared to

the single-level implementation, leads to a better reconstruction result. Besides the hybrid approach, as described in the above, it is also possible to carry out the double focal transformation and the linear Radon transformation after each other and, in that way, define a new transformation domain. In Chapter 6 it is demonstrated that especially for events that are not entirely focused in the focal domain, a better reconstruction result can be achieved. From the observations in Chapter 6 it sounds reasonable to assume that an implementation of the double focal transformation that relies, on a moving spread (non-stationary acquisition) geometry rather than on a fixed spread (stationary acquisition) geometry, will increase this improvement. Therefore, it is suggested to investigate this possibility in future research.

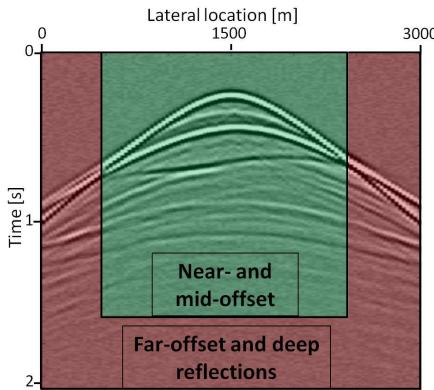


Figure 7.1: Synthetic shot record (see section 4.1). The near- and mid-offset range contains mostly curved events and is indicated by the green rectangle. The far-offset range and the deep data area contains mostly linear events and is indicated by the red shape.

7.2.3 Decreasing computational costs

The strength of the double focal transformation is that it can utilise the information of the full seismic dataset at once, in a physical manner, for data reconstruction. However, especially in 3D, the corresponding inversion problem can become very large and computationally expensive to solve. In order to cope with this problem, the focal reconstruction can be applied in little sub-windows of the data, e.g. only using shot records within a certain area. In that way a big inversion problem is split into many smaller sub-problems, which is computationally more effective and more suited for parallel implementations. However, in that way the redundancy of the data is not fully utilised.

7.2.4 3D versus 2D

A 3D implementation of the focal reconstruction makes in general more use of the redundancy of the data than a 2D implementation. It has also been discussed that a 3D implementation increases the sparseness of the data representation. Therefore, the comparison of a 2D and a 3D implementation for real data should be made in future research. However, a 3D implementation for a realistic dataset might require too much memory to be handled with the current implementation. As mentioned before, one solution is to apply the data reconstruction in windows. However, in that case the redundancy of the whole data is not fully utilised. Another solution is a 'pseudo-3D' implementation. Hereby, only four out of five data dimensions are varying. Hence, for the reconstruction algorithm only parts of the whole dataset, where one dimension is constant (e.g. y_{src}), are used at the same time. One example where this implementation could be applied, is a marine acquisition, where a boat pulls an acoustic source (or source array) and a number of streamers. The receivers are distributed over a certain range in both lateral (x_{rcv} and y_{rcv}) directions, but the source positions are approximately distributed along a line (along x_{src}).

7.2.5 Iterative implementation of the focal reconstruction

As discussed in Sections 4.1 and 7.2.1, it is not necessarily efficient to use a large number of focal operators or very complex focal operators, because the available prior information is usually not very accurate. However, an iterative approach should be investigated, where, first, the data is reconstructed based on only a couple of simple focal operators. After several pre-processing and processing steps based on these reconstructed data, much better prior knowledge might become available. It is reasonable to go back then and perform a new focal reconstruction based on more and more accurate focal operators. To repeat these steps several times should eventually lead to a very good reconstruction result. This result then aids the processing steps that require a dense data sampling. For instance artefacts can be avoided in the migration of seismic data. This issue is addressed, for instance, in Downton et al. (2010). Imagine that the prior knowledge is accurate and a very large number of focal operators is chosen. If then only the zero-time slice of the focal domains is considered, the double focal transformation resembles a sparse migration algorithm. Therefore, it might be interesting to iteratively combine these two processes (migration and reconstruction).

7.2.6 Utilising multiple reflections

Within this thesis multiple reflection events (i.e. surface-related multiples and internal multiples) are described in the same way as primaries. In the absence of strong multiple reflections the usage of one-way propagation operators based on a macro velocity model leads to a sufficient focusing of the data. However,

this is not the case in the presence of strong multiple reflections. With the current implementation of the double focal transformation, multiples are deliberately misinterpreted as primaries and corresponding one-way propagation operators are used that focus them. For instance, in the case of a surface-related multiple, the corresponding focal operator would represent a relative deep reflection event with relatively low velocity (e.g. water velocity). Furthermore, several focal operators for roughly the same time level might be used simultaneously, e.g. one corresponding to a primary event and another one corresponding to a multiple reflection. This approach might seem an oversimplification, but the main objective is to have an efficient tool for aliasing noise removal by inverse propagation of the data with focal operators. Even though these operators don't always have a physical meaning.

Surface-related multiples can be more naturally described by the double focal transformation. Equation 7.2 denotes the forward model including surface-related multiples:

$$\begin{aligned}\mathbf{P}(z_0) &= \hat{\mathbf{D}}(z_0)\mathbf{X}_0(z_0)\left(\hat{\mathbf{S}}(z_0) + \mathbf{R}^\cap(z_0)\mathbf{P}(z_0)\right), \text{ with} \\ \mathbf{X}_0(z_0) &= \sum_{n=1}^N \left[\hat{\mathbf{W}}^-(z_0, z_n)\mathbf{X}(z_n)\hat{\mathbf{W}}^+(z_n, z_0) \right].\end{aligned}\quad (7.2)$$

The upper line of equation 7.2 denotes the data model including surface-related multiples as it is described in Verschuur and Berkhouw (1997).

$\hat{\mathbf{W}}^-(z_0, z_n)\mathbf{X}(z_n)\hat{\mathbf{W}}^+(z_n, z_0)$ in the lower line of equation 7.2 describes the downward propagation of a wavefield from the surface to a number of depth levels, a reflection process and the upward propagation of the reflected wavefield to the surface, like in equation 2.6 at page 26. Since the downward and upward propagation operators $\hat{\mathbf{W}}^+(z_n, z_0)$ and $\hat{\mathbf{W}}^-(z_0, z_n)$ are imprecise, the focal domains $\mathbf{X}(z_n)$ are not only representing pure reflectivity but also include residual spatial phase terms that correct for the inaccuracy of the propagation operators. The focal domains should be chosen such that, together with the focal operators, they describe the impulse response of the subsurface without surface multiples (second part of equation 7.2). Furthermore, not only the source wavefield $\hat{\mathbf{S}}(z_0)$ is propagated, but also the up-going wavefield $\mathbf{P}(z_0)$ that was reflected at the free surface. The reflectivity matrix $\mathbf{R}^\cap(z_0)$ does not contain spatial phase shifts and is considered to be known as $-\mathbf{I}$ (negative identity matrix) for the marine case. In this form equation 7.2 can be used for multiple suppression. Since it has been demonstrated in Doulgeris et al. (2012) that multiples are beneficial for the process of separating blended sources, it is reasonable to assume that the process of multiple separation also contributes to seismic data reconstruction. In this way we can consider equation 7.2 as an alternative implementation of the EPSI algorithm (Estimation of primaries by sparse inversion, van Groenestijn and Verschuur, 2009a,b). In fact the combination of data reconstruction and primary estimation by sparse inversion is investigated by Lopez and Verschuur (2013a,b).

However, in a conventional processing flow the seismic data is reconstructed before multiple suppression or primary estimation. Also the separation of blended shot records relies on dense data and, therefore, on the quality of the reconstruction of coarse data. The simultaneous solution of all three problems is suggested: reconstruction of coarse data, estimation of primaries and separation of blended sources. In other words, is it more efficient to suppress multiples, blending noise and aliasing noise simultaneously or consecutively? Using the double focal transformation all three processes could be combined optimally.

Appendices

Appendix A

The derivation of the Rayleigh integral

In this chapter the theoretical foundation of acoustical wave propagation, the Rayleigh integral, is described. At first the Kirchhoff integral is derived, which allows wave field extrapolation to any point in an object from measurements taken at the surface all around this object. However, in most cases of field surveys only one-sided illumination and detection is possible. This situation can be described by the Rayleigh integral, which will be developed from the Kirchhoff integral. It allows the computation of a wavefield in a point that is separated by an infinitely extended flat surface from the wavefield generating sources. For more information on the derivation of the Rayleigh integral the reader is referred to Gisolf and Verschuur (2010), which we have used as basis for this appendix. Finally, the Rayleigh integral is reformulated as a convolution integral, which is the foundation of the WRW concept and the double focal transformation. As is demonstrated in Chapter 2 the WRW concept and the double focal transformation are fully based on these convolution integrals.

A.1 The acoustic wave equation

The Kirchhoff integral is based on the propagation of point sources, which are defined by the acoustic wave equation. In a homogeneous medium, in the presence of a point source at location \vec{r}_s the acoustic wave equation is formulated as:

$$\nabla^2 p(\vec{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\vec{r}, t)}{\partial^2 t} = -w(t) \delta(\vec{r} - \vec{r}_s), \quad (\text{A.1})$$

where t is time, \vec{r} is the position vector (x, y, z), $p(\vec{r}, t)$ is the pressure field, c is the propagation velocity, and $w(t)$ is the source wavelet. In the absence of sources

the right hand side of the equation would be set to zero. The solution to equation A.1 (with a point source) is given by:

$$p(\vec{r}, t) = \frac{w(t - |\vec{r} - \vec{r}_s|/c)}{4\pi|\vec{r} - \vec{r}_s|}. \quad (\text{A.2})$$

Transforming the wavefield of a point source (equation A.2) to the frequency domain leads to:

$$P(\vec{r}, \omega) = W(\omega) \frac{e^{-i\omega|\vec{r} - \vec{r}_s|/c}}{4\pi|\vec{r} - \vec{r}_s|}. \quad (\text{A.3})$$

If the point source in equation A.2 is assumed to produce a delta pulse wavelet in the time domain $w(t)$ the corresponding frequency equivalent $W(\omega)$ is equal to one for all frequencies. Therefore, $W(\omega)$ can be skipped in the formulation.

A.2 The Kirchhoff integral

The mathematical foundation of wavefield extrapolation is Gauss's Theorem:

$$\int_V \nabla \cdot \vec{a}(\vec{r}) dV = \oint_S \vec{a}(\vec{r}) \cdot \vec{n} dS, \quad (\text{A.4})$$

where \vec{r} is the position vector (x, y, z) and \vec{n} the normal to the closed surface S pointing outward. Please note that the lower case index $()_s$ denotes the source of a point source and that the capital S denotes a surface. Equation A.4 states that integrating the divergence of any vector field $\vec{a}(\vec{r})$ over the interior of the region (V), is equal to the integral of the vector-field over the region's boundary (S). Hence, if the sum of all sources and sinks within V (Figure A.1a) is not equal to zero then the divergence should be equal to the flow into or out of the volume V through the boundary S .

Next, \vec{a} is chosen as:

$$\vec{a}(\vec{r}) = f(\vec{r})\nabla g(\vec{r}) - g(\vec{r})\nabla f(\vec{r}), \quad (\text{A.5})$$

where f and g are bounded, twice differentiable functions.

Substituting equation A.5 in A.4 and applying the product rule:

$$\begin{aligned} & \nabla \cdot (f(\vec{r})\nabla g(\vec{r}) - g(\vec{r})\nabla f(\vec{r})) \\ &= [\nabla f(\vec{r}) \cdot \nabla g(\vec{r}) + f(\vec{r})\nabla^2 g(\vec{r})] - [\nabla g(\vec{r}) \cdot \nabla f(\vec{r}) + g(\vec{r})\nabla^2 f(\vec{r})] \\ &= f(\vec{r})\nabla^2 g(\vec{r}) - g(\vec{r})\nabla^2 f(\vec{r}), \end{aligned}$$

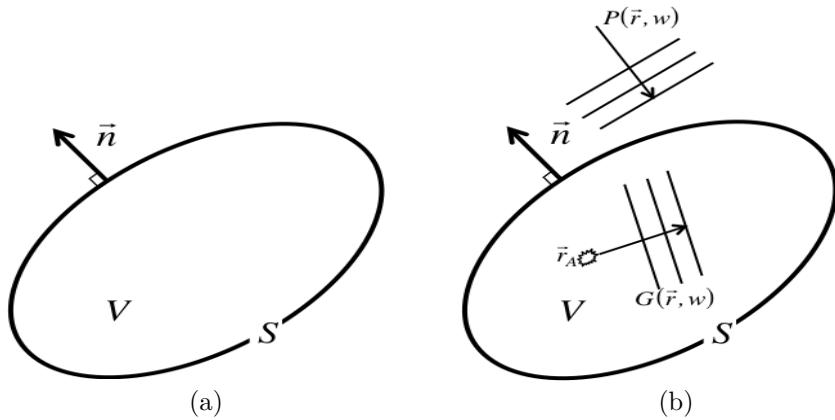


Figure A.1: In (a) a closed surface S containing volume V is displayed. The normal to S is pointing outwards. In (b) the wavefields $P(\vec{r}, \omega)$ and $G(\vec{r}, \omega)$ are added according to equations A.7 and A.8.

leads to the well known Greens Theorem:

$$\int_V [f(\vec{r})\nabla^2 g(\vec{r}) - g(\vec{r})\nabla^2 f(\vec{r})]dV = \oint_S [f(\vec{r})\nabla g(\vec{r}) - g(\vec{r})\nabla f(\vec{r})] \cdot \vec{n} dS. \quad (\text{A.6})$$

Please note that until now only mathematical theorems were developed. In the next step they will be used to describe the physical process of wavefield extrapolation. This is accomplished by choosing physical wavefields for $f(\vec{r})$ and $g(\vec{r})$ in equation A.6.

For $f(\vec{r})$ a wavefield $P(\vec{r}, \omega)$ is selected, that is generated by a source or by sources outside the surface S and is recorded everywhere on S (Figure A.1b). The aim is to extrapolate this field to an arbitrary point in V . Since $P(\vec{r}, \omega)$ is defined in the frequency domain it has to obey the Helmholtz equation inside V without sources:

$$\nabla^2 P(\vec{r}, \omega) + \frac{\omega^2}{c^2} P(\vec{r}, \omega) = 0, \quad (\text{A.7})$$

which is essentially the frequency equivalent of the acoustic wave equation in a homogeneous medium (equation A.1) without the source term on the right hand side of the equation. For $g(\vec{r})$ a wavefield $G(\vec{r}, \omega)$ is chosen that is generated by a point-source inside the volume V (Figure A.1b). Since it is presumed that the propagation velocity inside and outside V is the same, $G(\vec{r}, \omega)$ has to obey the Helmholtz equation with a source:

$$\nabla^2 G(\vec{r}, \omega) + \frac{\omega^2}{c^2} G(\vec{r}, \omega) = -\delta(\vec{r} - \vec{r}_A), \quad (\text{A.8})$$

where \vec{r}_A is the location of the source inside V .

Next, equation A.7 is multiplied by $-G(\vec{r}, \omega)$ and equation A.8 is multiplied by $P(\vec{r}, \omega)$ and then these two new equations are summed:

$$P(\vec{r}, \omega) \nabla^2 G(\vec{r}, \omega) - G(\vec{r}, \omega) \nabla^2 P(\vec{r}, \omega) = -P(\vec{r}, \omega) \delta(\vec{r} - \vec{r}_A). \quad (\text{A.9})$$

Integrating equation A.9 over volume V leads to the left hand side of equation A.6:

$$\begin{aligned} \int_V [P(\vec{r}, \omega) \nabla^2 G(\vec{r}, \omega) - G(\vec{r}, \omega) \nabla^2 P(\vec{r}, \omega)] dV = \\ - \int_V [P(\vec{r}, \omega) \delta(\vec{r} - \vec{r}_A)] dV = \\ -P(\vec{r}_A). \end{aligned} \quad (\text{A.10})$$

Substituting the result of equation A.10 in the left hand side of equation A.6 and multiplying by (-1) leads to the Kirchhoff integral.

$$P(\vec{r}_A, \omega) = - \oint_S [P(\vec{r}, \omega) \nabla G(\vec{r}, \omega) - G(\vec{r}, \omega) \nabla P(\vec{r}, \omega)] \cdot \vec{n} dS. \quad (\text{A.11})$$

It predicts the wavefield $P(\vec{r}_A)$ in an arbitrary point A inside the closed surface S from measurements of the wavefield on S . $G(\vec{r}, \omega)$ is the wavefield of a point source in the frequency domain (compare to equation A.3), which is defined as:

$$G(\vec{r}_A, \vec{r}, \omega) = \frac{e^{-i\omega|\vec{r} - \vec{r}_s|/c}}{4\pi|\vec{r} - \vec{r}_s|}. \quad (\text{A.12})$$

Thus, equation A.11 can be evaluated. However, extrapolating the wavefield to a point in a bounded volume V requires not only the measurement of $P(\vec{r}, \omega)$ on the surface S but also $\nabla P(\vec{r}, \omega) \cdot \vec{n}$. With only a single type measurement it will be impossible to establish the direction of propagation of the wavefield perpendicular to the surface. To resolve this ambiguity the pressure and the velocity have to be measured simultaneously.

A.3 Causality of the Kirchhoff integral

The frequency representation of the Kirchhoff integral provides no insight about causality. Therefore, equation A.11 is transformed to the time domain. In order to accomplish that, first, the gradient of wavefield G (equation A.12) needs to be derived.

Skipping $W(\omega)$ in equation A.3 (assuming an impulsive source-wavelet), setting the source location to point A and calling the wavefield G instead of P leads to:

$$G(\vec{r}, \omega) = \frac{e^{-i\omega|\vec{r} - \vec{r}_A|/c}}{4\pi|\vec{r} - \vec{r}_A|}. \quad (\text{A.13})$$

Using the chain rule of differentiation:

$$\nabla f(g(\vec{r})) = \frac{\partial f}{\partial g} \frac{\partial g(\vec{r})}{\partial \vec{r}} = \frac{\partial f}{\partial g} \nabla g(\vec{r}) \quad (\text{A.14})$$

and the product rule

$$\nabla(f(\vec{r})g(\vec{r})) = \nabla f(\vec{r})g(\vec{r}) + f(\vec{r})\nabla g(\vec{r}), \quad (\text{A.15})$$

where $f(\vec{r})$ and $g(\vec{r})$ are arbitrary functions, the gradient of G can be derived as:

$$\begin{aligned} \nabla G &= \nabla[(4\pi|\vec{r} - \vec{r}_A|)^{-1} e^{-i\omega|\vec{r} - \vec{r}_A|/c}] \\ &= (-1)(4\pi|\vec{r} - \vec{r}_A|)^{-2} \nabla(4\pi|\vec{r} - \vec{r}_A|) e^{-i\omega|\vec{r} - \vec{r}_A|/c} \\ &\quad + (4\pi|\vec{r} - \vec{r}_A|)^{-1} e^{-i\omega|\vec{r} - \vec{r}_A|/c} \nabla(-i\omega|\vec{r} - \vec{r}_A|/c) \\ &= -\frac{1}{4\pi} \left(1 + i\omega \frac{|\vec{r} - \vec{r}_A|}{c} \right) \frac{e^{-i\omega|\vec{r} - \vec{r}_A|/c}}{|\vec{r} - \vec{r}_A|^2} \nabla|\vec{r} - \vec{r}_A|. \end{aligned} \quad (\text{A.16})$$

Keeping in mind that the multiplication with $i\omega$ in the frequency domain translates to a differentiation in the time domain and that a phase factor $\exp(-i\omega|\vec{r} - \vec{r}_A|/c)$ in the frequency domain leads to a time shift $|\vec{r} - \vec{r}_A|/c$ in the time domain, the inverse Fourier transformation of the Kirchhoff integral (equation A.11) using equation A.16 reads:

$$\begin{aligned} p(\vec{r}_A, t) &= \oint_S \left\{ \frac{p(\vec{r}, t)}{4\pi} \left(1 + \frac{\partial}{\partial t} \frac{|\vec{r} - \vec{r}_A|}{c} \right) \frac{\nabla|\vec{r} - \vec{r}_A|}{|\vec{r} - \vec{r}_A|^2} - \frac{\nabla p(\vec{r}, t)}{4\pi|\vec{r} - \vec{r}_A|} \right\}_{t - |\vec{r} - \vec{r}_A|/c} \cdot \vec{n} \, dS \\ &= \frac{1}{4\pi} \oint_S \left\{ \left[\frac{1}{|\vec{r} - \vec{r}_A|} p(\vec{r}, t) + \frac{1}{c} \frac{\partial p(\vec{r}, t)}{\partial t} \right] \frac{\nabla|\vec{r} - \vec{r}_A|}{|\vec{r} - \vec{r}_A|} \frac{\nabla p(\vec{r}, t)}{|\vec{r} - \vec{r}_A|} \right\}_{t - |\vec{r} - \vec{r}_A|/c} \cdot \vec{n} \, dS. \end{aligned} \quad (\text{A.17})$$

Equation A.17 provides information about causality in wavefield extrapolation. The wavefield in A is predicted at a time t , from measurements of the wavefield on S at an earlier time $t - |\vec{r} - \vec{r}_A|/c$. This is the causal form of the Kirchhoff integral: The known values on the boundary S are the cause of the predicted field value in A , which is the effect. The time delay between cause and effect is due to wave propagation from the measurement location to the predicted location (Figure A.2a).

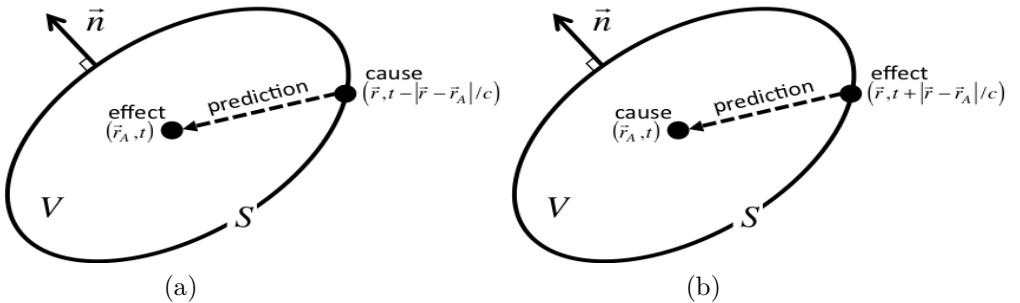


Figure A.2: Causal (a) and anti-causal (b) prediction of the wavefield in \vec{r}_A from measurements at S .

The Kirchhoff integral is the mathematical formulation of the Huygens' Principle, where the known wavefield at S is treated as secondary sources that together describe the wavefield at a later time.

If in equation A.12 the wavefield of a point sink:

$$G(\vec{r}_A, \vec{r}, \omega) = \frac{e^{i\omega|\vec{r} - \vec{r}_A|/c}}{4\pi|\vec{r} - \vec{r}_A|}. \quad (\text{A.18})$$

instead of a point source would have been chosen, the Kirchhoff integral wouldn't look any different in the frequency domain (equation A.11). The difference is only revealed in the time domain, where the anti-causal form of the Kirchhoff integral reads:

$$\begin{aligned} p(\vec{r}_A, t) &= \frac{1}{4\pi} \oint_S \left\{ \left[\frac{1}{|\vec{r} - \vec{r}_A|} p(\vec{r}, t) + \frac{1}{c} \frac{\partial p(\vec{r}, t)}{\partial t} \right] \frac{\nabla |\vec{r} - \vec{r}_A|}{|\vec{r} - \vec{r}_A|} \frac{\nabla p(\vec{r}, t)}{|\vec{r} - \vec{r}_A|} \right\}_{t+|\vec{r} - \vec{r}_A|/c} \cdot \vec{n} \, dS. \end{aligned} \quad (\text{A.19})$$

In the anti-causal form (Figure A.2b) the wavefield in A at time t is predicted from measurements at S at later times $t + |\vec{r} - \vec{r}_A|/c$. Hence, the known values of p on S are the effect, caused by the predicted value of p in A .

Since the Kirchhoff integral in the frequency domain (equation A.11) has two terms (P and ∇P) the direction of propagation of the wavefield P on S can be obtained. Depending on this direction for some measurements on S the field in A is the effect of those secondary sources, whereas for other measurements the field in A is the cause of those measurements. If the wavefield P is generated by a single source (Figure A.3a) it can be easily visualised which part of S acts as a cause (blue) of the field in A and which part of S acts as the effect (red) of the field in A . Hence, through the blue part of S energy only flows into the volume V and through the red part only out of V . This separation into two parts is

fully dependent on the position of the source and the shape of S , but independent of the position of A (within V). The presence of two terms in the Kirchhoff integral allows to differentiate between blue and red parts of the surface S . If equation A.17 is used, only the blue part contributes, whereas if equation A.19 is used, only the red part contributes. The total field can be a linear superposition of point-source fields, each of them having their specific blue and red parts of surface S .

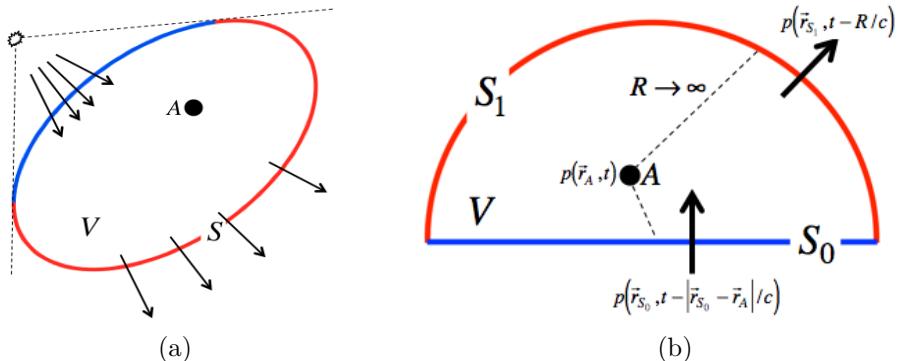


Figure A.3: Depending on the source-position in (a) through one part (blue) energy is flowing into volume V and through the other part (red) energy is flowing out of V . In (b) the wavefield in A is causal predicted from two points at the boundary $S = S_0 + S_1$, where it is assumed that all sources are below S_0 (Gisolf and Verschuur, 2010).

A.4 The Rayleigh integrals in a homogeneous medium

The topic of this thesis is seismic data reconstruction, where in most cases the sources and receivers are placed along a relatively flat plane rather than all around the area of interest. For the sake of the theoretical derivation it is assumed that all sources are placed below an infinitely extended flat plane S_0 (blue part in Figure A.3b), which is part of an interior-enclosing surface S . The surface S is closed in the direction away from the sources. Through the blue part S_0 energy can only flow into the interior (volume V) and through the closing part (red) S_1 energy can only flow out of it.

In the remainder of this section only forward extrapolation (i.e. extrapolation away from the sources), using the causal form of the Kirchhoff integral (equation A.17) will be considered.

Since the direction of propagation through S_0 is by definition inward and through S_1 by definition outward, energy along the red part of S can't contribute in a causal way to the wavefield in A . This suggests that the wavefield doesn't need to be measured all around the area of interest and, therefore, the closing part S_1

can be discarded and with it one term in the Kirchhoff integral. But since S_0 has an infinite extend, the wavefield has to be measured along an infinite flat plane. To derive an extrapolation integral that only needs P or $\nabla P \cdot n$ Rayleigh argued that there is a degree of freedom in the Kirchhoff integral. By adding any solution Γ of the source-free Helmholtz equation:

$$\nabla^2 \Gamma(\vec{r}, \omega) + \frac{\omega^2}{c^2} \Gamma(\vec{r}, \omega) = 0 \quad (\text{A.20})$$

to the Green's function G in equation A.11 the prediction result is not changed:

$$P(\vec{r}_A) = - \oint_S [P \nabla(G + \Gamma) - (G + \Gamma) \nabla P] \cdot \vec{n} \, dS. \quad (\text{A.21})$$

First, as was discussed before, S_1 doesn't contribute to the wavefield in A . Hence, the integration boundary in equation A.21 can be changed from S to S_0 . Furthermore, Rayleigh suggested to define Γ in a way that it is generated by a point-source in position A' , which is the mirror position to point A with respect to S_0 (see Figure A.4a).

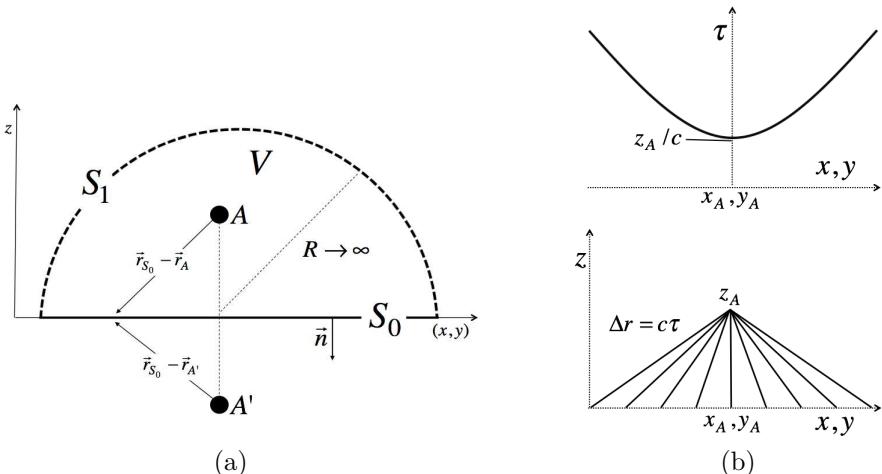


Figure A.4: (a) Configuration for the derivation of the Rayleigh integral. All sources generating the wavefield P are below S_0 . The wavefield prediction point (also source point for the green's function G) is in A inside V . The source point for Γ is in the mirror position A' . In (b) the delay time τ is plotted as a function of lateral position. The hyperbolic shape is typical for homogeneous media. For inhomogeneous media (section A.5) τ describes a more complex function.

Please note that this choice is allowed as far as the source-free Helmholtz equation (equation A.20) is concerned, because A' is outside the volume V . Hence, Γ is source-free within V . Secondly, it is decided that the point-source in A' will have a negative source strength compared to the one in A :

$$G(\vec{r}_A, \vec{r}_{S_0}) = -\Gamma(\vec{r}'_A, \vec{r}_{S_0}). \quad (\text{A.22})$$

This leads to a cancelling of the second term in equation A.21 everywhere on S_0 . With the wavefields cancelling each other at S_0 , the normal components of their gradients are equal on S_0 :

$$\nabla G(\vec{r}_{S_0}) = \nabla \Gamma(\vec{r}_{S_0}). \quad (\text{A.23})$$

Substituting equation A.22 and A.23 in A.21 leads to:

$$\begin{aligned} P(\vec{r}_A) &= - \oint_{S_0} [P \nabla(G + G) - (G - G) \nabla P] \cdot \vec{n} \, dS_0 \\ &= -2 \oint_{S_0} [P \nabla G] \cdot \vec{n} \, dS_0 \end{aligned} \quad (\text{A.24})$$

Please note that in the following it is assumed that surface S_0 is only extended in the x, y -direction and is not dipping and has no structure. Because of the outward pointing normal (negative z -direction) everywhere on S_0 in Figure A.4a the gradient of G becomes:

$$\nabla G \cdot \vec{n} = \begin{pmatrix} \partial G / \partial x \\ \partial G / \partial y \\ \partial G / \partial z \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} = -\partial G / \partial z, \quad (\text{A.25})$$

which changes equation A.24 to:

$$P(\vec{r}_A) = 2 \oint_{S_0} dS_0 P \frac{\partial G}{\partial z}, \quad (\text{A.26})$$

where G is still given by the point-source solution (equation A.12). With Δr defined as:

$$|\vec{r} - \vec{r}_A| = \sqrt{(x - x_A)^2 + (y - y_A)^2 + (z - z_A)^2} = \Delta r, \quad (\text{A.27})$$

its derivation with respect to z :

$$\begin{aligned} \left(\frac{\partial \Delta r}{\partial z} \right)_{z=0} &= \\ \left(\frac{1}{2} [(x - x_A)^2 + (y - y_A)^2 + (z - z_A)^2]^{-\frac{1}{2}} \, 2(z - z_A) \frac{\partial(z - z_A)}{\partial z} \right)_{z=0} \\ &= \frac{-z_A}{\Delta r} \end{aligned} \quad (\text{A.28})$$

and its derivation with respect to z_A :

$$\left(\frac{\partial \Delta r}{\partial z_A} \right)_{z=0} = \frac{z_A}{\Delta r}, \quad (\text{A.29})$$

the derivative of G with respect to z at z_0 becomes:

$$\begin{aligned} \left(\frac{\partial G}{\partial z} \right)_{z=0} &= \left(\frac{\partial}{\partial z} \left((4\pi\Delta r)^{-1} e^{-i\omega\Delta r/c} \right) \right)_{z=0} \\ &= -1(4\pi\Delta r)^{-2} \left(4\pi \frac{-z_A}{\Delta r} \right) e^{-i\omega\Delta r/c} \\ &\quad + (4\pi\Delta r)^{-1} e^{-i\omega\Delta r/c} (-i\omega/c) \frac{-z_A}{\Delta r} \\ &= \frac{z_A}{4\pi(\Delta r)^3} e^{-i\omega\Delta r/c} \left(1 + \frac{i\omega}{c} \Delta r \right) \end{aligned} \quad (\text{A.30})$$

and the following equality can be derived:

$$\begin{aligned} \left\{ \frac{\partial}{\partial z_A} \left[(\Delta r)^{-1} e^{-i\omega\Delta r/c} \right] \right\}_{z=0} &= \\ -(\Delta r)^{-2} \frac{z_A}{\Delta r} e^{-i\omega\Delta r/c} + (\Delta r)^{-1} e^{-i\omega\Delta r/c} \frac{-i\omega}{c} \frac{z_A}{\Delta r} &= \\ -\frac{e^{-i\omega\Delta r/c} z_A}{(\Delta r)^3} \left(1 + \frac{i\omega}{c} \Delta r \right). \end{aligned} \quad (\text{A.31})$$

With this little bit of math the derivation of the Rayleigh integrals becomes rather simple. Just by substituting equation A.30 in equation A.26 and at the same time writing out the integral over S_0 as an infinite integral over x and y in the $z = 0$ plane, the following version of the Rayleigh forward extrapolation integral can be derived:

$$P(\vec{r}_A, \omega) = \frac{z_A}{2\pi} \iint_{-\infty}^{\infty} dx dy P(x, y, 0, \omega) \left(1 + i\omega \frac{\Delta r}{c} \right) \frac{e^{-i\omega\Delta r/c}}{(\Delta r)^3}. \quad (\text{A.32})$$

With the help of equation A.31 this equation can also be formulated as:

$$P(\vec{r}_A, \omega) = \frac{-1}{2\pi} \frac{\partial}{\partial z_A} \iint_{-\infty}^{\infty} dx dy P(x, y, 0, \omega) \frac{e^{-i\omega\Delta r/c}}{\Delta r}. \quad (\text{A.33})$$

These two versions are called the Rayleigh II integral. If a positive source strength would have been chosen in point-source A' such that $G(\vec{r}_A, \vec{r}_{S_0}) = \Gamma(\vec{r}'_A, \vec{r}_{S_0})$, then

the first term in equation A.21 would have cancelled, which would have led to the derivation of the Rayleigh I integral.

Since within this thesis only the Rayleigh II integral is used, it will hereafter just be referred to as 'the Rayleigh integral'. It can be transformed to the time domain in the same way as was done for the Kichhoff integral in equation A.17:

$$p(\vec{r}_A, t) = \frac{z_A}{2\pi c^3} \iint_{-\infty}^{\infty} dx dy \frac{1}{(\Delta r)^3} [p(x, y, 0, t - \frac{\Delta r}{c}) + \frac{\Delta r}{c} \frac{\partial}{\partial t} p(x, y, 0, t - \frac{\Delta r}{c})]. \quad (\text{A.34})$$

Again the causality is easily visible. By introducing the delay time $\tau = \Delta r/c$ equation A.34 can be written as:

$$p(\vec{r}_A, t) = \frac{z_A}{2\pi c^3} \iint_{-\infty}^{\infty} dx dy \frac{1}{\tau^3} [p(x, y, 0, t - \tau) + \tau \frac{\partial}{\partial t} p(x, y, 0, t - \tau)]. \quad (\text{A.35})$$

The time representation of equation A.33 is formulated as:

$$\begin{aligned} p(\vec{r}_A, t) &= -\frac{1}{2\pi} \frac{\partial}{\partial z_A} \iint_{-\infty}^{\infty} dx dy \frac{1}{\Delta r} p(x, y, 0, t - \Delta r/c) \\ &= -\frac{1}{2\pi c} \frac{\partial}{\partial z_A} \iint_{-\infty}^{\infty} dx dy \frac{1}{\tau} p(x, y, 0, t - \tau). \end{aligned} \quad (\text{A.36})$$

In Figure A.4b the delay times τ are plotted as a function of (x, y) -position along the data acquisition plane. As illustrated, for homogeneous media this is a hyperbolic shape. The delay times are a direct visualisation of Huygens' principle and the causality of wave propagation. Hence, the wavefield in A is generated by contributions of all the secondary sources along the acquisition plane. These contributions are delayed corresponding to their positions relative to point A (Figure A.4b).

Most acoustical imaging applications work in the far-field, by which is meant that the distance Δr to the source is larger than a number of wave-lengths (typically more than ten). For example, for $\Delta r = 10\lambda$ the following is valid:

$$|1| \ll \left| i\omega \frac{\Delta r}{c} \right| = \left| i\omega \frac{10\lambda}{c} \right| = |i20\pi| \quad (\text{A.37})$$

and the first term in equation A.32 can be skipped, resulting in:

$$P(\vec{r}_A, \omega) = \frac{i\omega z_A}{2\pi c} \iint_{-\infty}^{\infty} dx dy P(x, y, 0, \omega) \frac{e^{-i\omega\Delta r/c}}{(\Delta r)^2}. \quad (\text{A.38})$$

A.4.1 The Rayleigh integral in 2D space

The real world has three spatial dimensions, which makes the 3D implementation of the double focal transformation (introduced in Chapter 5.2) the ultimate goal. However, the majority of the content and all field data examples are in 2D. Therefore, it is important to provide the theoretical foundation in two spatial dimensions as well. In the following the 2D Rayleigh integral is derived.

Following again Gisolf and Verschuur (2010) a 2D wavefield extrapolation algorithm can be derived from the 3D equations by stating that the recorded wavefield is constant along one lateral direction (e.g. y -coordinate) in the acquisition plane. In that sense the wavefield is still propagating in a 3D space, but since it has been generated by y -direction-line-sources below S_0 the wavefield is indeed independent of y . Equations A.32 and A.33 are valid independent of the source-configuration that generates the wavefield (e.g. line-sources) below S_0 . The integral A.33 can be re-written:

$$P(x_A, z_A, \omega) = \frac{-1}{2\pi} \frac{\partial}{\partial z_A} \int_{-\infty}^{\infty} dx P(x, z=0, \omega) \int_{-\infty}^{\infty} dy \frac{e^{-i\omega\Delta r/c}}{\Delta r}, \quad (\text{A.39})$$

with $\Delta r = \sqrt{\rho^2 + y^2}$ and $\rho = \sqrt{(x - x_A)^2 + (0 - z_A)^2}$.

Note, that the recorded wavefield at $z = 0$ is independent of y and, therefore, can be excluded from the integral over y in equation A.39. Also, since line-sources are used, that are infinitely extended in y , the delay with which the secondary sources for different y -coordinates contribute is independent of the y -locations of A . Hence, shifting A along the y -axis wouldn't change the wavefield in A . Therefore, it is allowed to simply set $y_A = 0$ in Δr .

Using the fact that the integrand in equation A.39 is symmetrical in y , it can be re-formulated as:

$$P(x_A, z_A, \omega) = \frac{-1}{\pi} \int_{-\infty}^{\infty} dx P(x, z=0, \omega) \frac{\partial}{\partial z_A} \int_0^{\infty} dy \frac{e^{-i\omega\Delta r/c}}{\Delta r}. \quad (\text{A.40})$$

The integral over y represents a zero-order Hankel function of the second kind (Abramowitz and Stegun, 1970):

$$H_0^{(2)}(\omega\rho/c) = \frac{2i}{\pi} \int_0^{\infty} dy \frac{e^{-i\omega\sqrt{\rho^2+y^2}/c}}{\sqrt{\rho^2+y^2}}. \quad (\text{A.41})$$

The differentiation with respect to z_A in equation A.40 turns the zero order Hankel function into a first order Hankel function:

$$\frac{\partial}{\partial z_A} H_0^{(2)}(\omega\rho/c) = -\frac{\omega z_A}{c\rho} H_1^{(2)}(\omega\rho/c). \quad (\text{A.42})$$

Using equations A.41 and A.42 in equation A.40 leads to the following version of the 2D Rayleigh integral:

$$P(x_A, z_A, \omega) = -\frac{i\omega z_A}{2c} \int_{-\infty}^{\infty} dx P(x, z=0, \omega) \frac{1}{\rho} H_1^{(2)}(\omega\rho/c). \quad (\text{A.43})$$

Please note, even though the attempt is to generate a 2D extrapolation integral, the space over which the integration takes place is still 3D. Through an asymptotic expansion of the Hankel function,

$$\lim_{|z| \rightarrow \infty} H_1^{(2)}(z) = i\sqrt{\frac{2i}{\pi z}} e^{-iz}, \quad (\text{A.44})$$

the far-field 2D Rayleigh extrapolation integral is denoted as:

$$P(x_A, z_A, \omega) = z_A \sqrt{\frac{i\omega}{2\pi c}} \int_{-\infty}^{\infty} dx P(x, z=0, \omega) \frac{e^{-\omega\rho/c}}{\rho^{3/2}}. \quad (\text{A.45})$$

A time domain version of the 2D Rayleigh integral can be easily obtained from the 3D version in equation A.36. Since the recorded data at S_0 is independent of y the integrand is only dependent on y through Δr , which is symmetrical in y . Applying the y -symmetry to equation A.36 and changing the order of integration (first over x) simply leads to:

$$p(\vec{r}_A, t) = -\frac{1}{\pi} \frac{\partial}{\partial z_A} \int_0^\infty dy \int_{-\infty}^\infty dx \frac{1}{\Delta r} p(x, y, 0, t - \Delta r/c). \quad (\text{A.46})$$

In the integration over x , $\Delta r = c\sqrt{\tau^2 + (x - x_A)^2/c^2}$, where: $\tau = \sqrt{y^2 - z_A^2}/c^2$. The integration variable shall be changed from y to τ . Hence:

$$\frac{d\tau}{dy} = 0 + c^{-1} \frac{1}{2} (y^2 + z_A^2)^{-\frac{1}{2}} 2y = \frac{y}{c\sqrt{y^2 + z_A^2}} \quad (\text{A.47})$$

and, therefore:

$$dy = \frac{c\sqrt{y^2 + z_A^2}}{y} d\tau = \frac{c\tau}{\sqrt{\tau z_A^2/c^2}} d\tau. \quad (\text{A.48})$$

With this new integration variable equation A.46 becomes:

$$p(\vec{r}_A, t) = -\frac{c}{\pi} \frac{\partial}{\partial z_A} \int_{\frac{z_A}{c}}^{\infty} d\tau \frac{\tau}{\sqrt{\tau z_A^2/c^2}} \int_{-\infty}^{\infty} dx \frac{1}{\Delta r} p(x, y, 0, t - \Delta r/c). \quad (\text{A.49})$$

Note that even though equation A.49 has a pole at the integration boundary $\tau(y=0) = \frac{z_A}{c}$ this equation, or rather its time-reverse version, was used in many industrial imaging algorithms.

It is important to notice, that in actual 2D seismic field data applications the recorded data is not generated by line sources in the y -direction, but by point sources along a 2D profile. Thus, before the 2D Rayleigh integral can be applied, the point-source data have to be transformed to line-source data by the following equation:

$$P_L(x_S, x_R, \omega) = \int_{-\infty}^{\infty} dy_S P(x_S, y_S, x_R, 0, \omega), \quad (\text{A.50})$$

where $P_L(x_S, x_R, \omega)$ is the simulated line-source data with a line source in x_S and a point receiver in x_R . $P(x_S, y_S, x_R, y_R = 0, \omega)$ is the point-source data that would have been acquired from a point-source at (x_S, y_S) and a point-receiver at $(x_R, y_R = 0)$. Hence, to achieve line-source data, an integration over point-source data for different source- y locations is performed.

The actual measured 2D point-source data (e.g. 2D profile measurement) has of course a constant y -coordinate (e.g. 0) for all sources and receivers: $P(x_S, 0, x_R, 0)$. To evaluate the integral in equation A.50 a model for wavefields generated by a point source is needed. Here a simple convolution model is chosen. Please note that the task is to generate two-way data. However, assuming flat reflectors this can be simulated with the model of 'mirror receiver' (see Figure A.5). The two-way reflection data from a reflector at z_j is equivalent to a one-way data with the receiver being at $2z_j$. Hence, the wavefield, generated by a point source in (x_S, y_S) and received in $(x_R, 0)$ is given by:

$$P(x_S, y_S, x_R, 0, \omega) = W(\omega) \sum_j R_j \frac{e^{-i\omega t_j}}{r_j}, \quad (\text{A.51})$$

where $W(\omega)$ is the source pulse, R_j the reflectivity at depth z_j ,

$r_j = \sqrt{2z_j^2 + (x_R - x_S)^2 + y_S^2}$ and $t_j = r_j/c$. The resulting integral is again a Hankel function (see equation A.41). As done previously it can be shown that for the far-field approximation the integration over y_S results, because of the symmetry in y -direction, in a 'half integration'. Hence, in the frequency domain a division by $\sqrt{i\omega}$ (like for equation A.45) and a multiplication in the time domain by \sqrt{t} has to be applied. While the $\sqrt{i\omega}$ scaling can be performed at any stage in the processing or even omitted all together, the \sqrt{t} scaling is crucial for a proper functioning of the 2D extrapolation integral. The cause of this scaling factors is the absence of infinite receivers in y -direction.

The derived equations assume a constant velocity of the medium, in which the field propagates. For the case of velocities that vary in the z -direction $c(z)$ see Wapenaar et al. (1992).

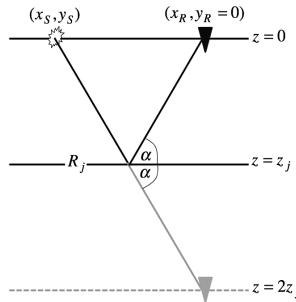


Figure A.5: Schematic illustration of a mirror receiver. The travel-path from the source to the actual receiver (black) has the same length as the travel-path to the virtual mirror receiver (grey).

A.5 The Rayleigh integral in inhomogeneous low contrast media

The wavefield of a point source (Green's function) in homogeneous media is defined by equation A.12 in the frequency domain. For the sake of convenience it is repeated here:

$$G(\vec{r}_A, \vec{r}, \omega) = \frac{e^{-i\omega|\vec{r} - \vec{r}_A|/c}}{4\pi|\vec{r} - \vec{r}_A|}. \quad (\text{A.52})$$

It basically consists of two parts: Phase shift (enumerator in equation A.52) and an amplitude term (denominator in equation A.52). Since c is a constant, the greens function is only a function of the distance to point A .

For the case of an inhomogeneous medium the phase shift term and the amplitude term $a(\vec{r}_A, \vec{r})$ are both dependent on the actual location of the point A . In a non-reflecting non-dispersive background, the causal Green's function $G(\vec{r}_A, \vec{r}, \omega)$ with a point source in A , generally should consist of a single arrival of an undistorted delta-pulse with the arrival time $t_G(\vec{r}_A, \vec{r})$:

$$G(\vec{r}_A, \vec{r}, \omega) = a(\vec{r}_A, \vec{r})e^{-i\omega t_G(\vec{r}_A, \vec{r})}. \quad (\text{A.53})$$

The time equivalent to equation A.53 reads:

$$g(\vec{r}_A, \vec{r}, t) = a(\vec{r}_A, \vec{r})\delta[t - t_G(\vec{r}_A, \vec{r})]. \quad (\text{A.54})$$

The amplitude factor $a(\vec{r}_A, \vec{r})$ represents geometrical spreading and can be determined from the travel times $t_G(\vec{r}_A, \vec{r})$ and the density background values in \vec{r} and \vec{r}_A .

Using the Green's function for inhomogeneous media, the Rayleigh integral in low contrast inhomogeneous media in the frequency domain (far field approximation) is defined as:

$$P(\vec{r}_A, \omega) = 2i\omega \iint_{-\infty}^{\infty} dx dy P(x, y, 0, \omega) \frac{a}{\rho_0} \left(\frac{\partial t_G}{\partial z} \right)_{z=0} e^{-i\omega t_G}. \quad (\text{A.55})$$

The corresponding time-domain-equivalent is denoted as:

$$p(\vec{r}_A, t) = 2 \iint_{-\infty}^{\infty} dx dy \frac{a}{\rho_0} \left(\frac{\partial t_G}{\partial z} \right)_{z=0} \frac{\partial}{\partial t} p(x, y, 0, t - t_G), \quad (\text{A.56})$$

For a detailed derivation the reader is referred to Gisolf and Verschuur (2010). The procedure is very similar to the homogeneous case, just starting from the inhomogeneous wave equation and, hence, utilising the inhomogeneous Helmholtz equation instead of equation A.8 on page 103.

A.6 Rayleigh integrals in terms of convolution

In this section the Rayleigh integral is formulated as a convolution-integral in order to describe wavefield propagation in terms of convolution with a wavefield operator. For more detailed information the reader is referred to Berkhoult (1987), Wapenaar and Berkhoult (1989), Kinneging et al. (1989), Berkhoult (1993) and Berkhoult et al. (1999).

First, the general 3D case is derived. The notation is slightly adjusted to cover the case where acoustic data are downward propagated from one depth level to another (see Figure A.6):

- Sources are above z_0 and are downward extrapolated to z' .
- Depth difference $|z_0 - z_A| = |z_0 - z'|$ is simply denoted with Δz .
- Angle ϕ is introduced such that $\cos(\phi) = \Delta z / \Delta r$.
- Wavenumber $k = \omega/c$ is used.

With these changes in notation of the 3D Rayleigh integral in equation A.32 the new formulation reads:

$$P(x_A, y_A, z', \omega) = \iint_{-\infty}^{\infty} dx dy \left[\frac{1}{2\pi} \frac{1 + ik\Delta r}{(\Delta r)^2} \cos(\phi) e^{-ik\Delta r} \right] P(x, y, z_0, \omega). \quad (\text{A.57})$$

By defining the term in the square brackets in equation as the propagation operator:

$$W^+(x, x_A, y, y_A, \Delta z, \omega) = \frac{1 + ik\Delta r}{(2\pi\Delta r)^2} \cos(\phi) e^{-ik\Delta r} \quad (\text{A.58})$$

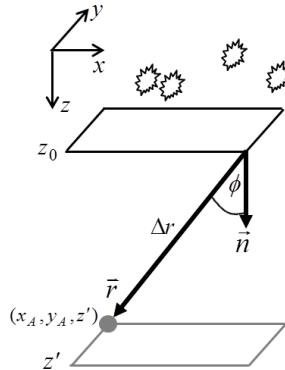


Figure A.6: Schematic illustration of the downward propagation of the wavefield at a certain point at the surface to a certain point in the subsurface.

with its corresponding far-field approximation:

$$W^+(x, x_A, y, y_A, \Delta z, \omega) \approx \frac{ik}{2\pi} \cos(\phi) \frac{e^{-ik\Delta r}}{\Delta r} \quad (\text{A.59})$$

that extrapolates the wavefield from a plane in x, y direction at depth level z_0 to point A at depth z' , equation A.57 changes to a convolution integral along the x and y axes:

$$P(x_A, y_A, z', \omega) = \iint_{-\infty}^{\infty} dx dy W^+(x, x_A, y, y_A, \Delta z, \omega) P(x, y, z_0, \omega). \quad (\text{A.60})$$

Please note that the ' W^+ ' denotes downwards propagation.

For the case of inhomogeneous velocity distributions, as discussed in section A.5, the Rayleigh integral can still be formulated as a convolution of a propagation operator with the data. In that case in equation A.60 only another propagation operator is used. Hence instead of using a W^+ that describes propagation in homogeneous media the propagation operator (far-field approximation) for inhomogeneous media is used, which is defined as:

$$W^+(x, x_A, y, y_A, \Delta z, \omega) = 2i\omega \frac{a}{\rho_0} \left(\frac{\partial t_G}{\partial z} \right)_{z=0} e^{i\omega t_G}. \quad (\text{A.61})$$

In principal any suited propagation operator could be chosen. W^+ doesn't even need to be based on a velocity model and could be derived in a data-driven manner like it is done in applications of the Common Focus Point (CFP) technology (see Thorbecke and Berkhouwt, 1994; Berkhouwt, 1997; Thorbecke, 1997). However, for

the sake of convenience in the remainder of this chapter the case of homogeneous velocity distribution will be considered.

Similarly the 2D Rayleigh integral in equation A.43 can be formulated as:

$$P(x_A, z', \omega) = \int_{-\infty}^{\infty} dx \left[-\frac{ik}{2} \cos(\phi) H_1^{(2)}(k\rho) \right] P(x, z_0, \omega), \quad (\text{A.62})$$

using $\cos(\phi) = \Delta z / \Delta \rho$ and $H_1^{(2)}$ being the first order Hankel function of the second kind (see Abramowitz and Stegun, 1970).

The propagation operator in 2D (within the square brackets in equation A.62) for homogeneous media is denoted as:

$$W^+(x, x_A, \Delta z, \omega) = -\frac{ik}{2} \cos(\phi) H_1^{(2)}(k\rho) \quad (\text{A.63})$$

with its corresponding far-field approximation:

$$W^+(x, x_A, \Delta z, \omega) \approx \sqrt{\frac{ik}{2\pi}} \cos(\phi) \frac{e^{-ik\rho}}{\sqrt{\rho}}. \quad (\text{A.64})$$

It downward extrapolates the wavefield from a line in x direction at depth level z_0 to point A at depth z' . Equation A.62 changes to a convolution integral along the x axis:

$$P(x_A, z', \omega) = \int_{-\infty}^{\infty} dx W^+(x, x_A, \Delta z, \omega) P(x, z_0, \omega). \quad (\text{A.65})$$

The possibility to express wave-propagation by convolution integrals (equations A.60 and A.65) in the frequency domain is the theoretical foundation of the detail-hiding operator notation.

Appendix B

The wave character of the WRW model

In section 2.1 the data generation with the WRW model was derived. However, sources and receivers are distributed along the surface in a discrete fashion. The grid-points also have discrete locations. The question is, how can reflections be explained that physically would reflect in between the grid points, as is visualised by the black ray-path in Figure B.1?

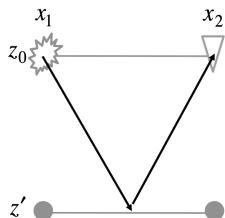


Figure B.1: Simple situation where the incident ray-path would be reflected in between (laterally) two grid-points .

Using just ray-paths between source, receiver and grid-points it would be impossible to generate the true reflection path. The answer to this question is provided by the wave character of the WRW model. This means that no ray-paths are propagated, but wave-fronts, which have a spatial extent. This allows a certain communication between the grid-points. In the following, data generation with the WRW model is defined utilising the formulation displayed in Figure 2.7 on page 18.

Using only one source experiment with one source at location (x_1, z_0) changes the source matrix to a column vector:

$$\begin{pmatrix} S_1(x_1) \\ 0 \end{pmatrix}. \quad (\text{B.1})$$

Propagating the wavefield from all possible surface locations to depth level z' with the downward propagation operator $\mathbf{W}^+(z', z_0)$ is then formulated as:

$$\begin{pmatrix} W^+(x_{1,1}) & S_1(x_1) + W^+(x_{1,2}) & 0 \\ W^+(x_{2,1}) & S_1(x_1) + W^+(x_{2,2}) & 0 \end{pmatrix}. \quad (\text{B.2})$$

Assuming for the sake of convenience a point source with a delta-pulse wavelet, $S_1(x_1)$ can be set to 1.

Equation B.2 then simply reads:

$$\begin{pmatrix} W^+(x_{1,1}) \\ W^+(x_{2,1}) \end{pmatrix}. \quad (\text{B.3})$$

Since only a single source at x_1 was fired, the vector in equation B.3 represents the downward propagation of the one source to both grid-points. The reflector is reacting locally. This means in the current example, that the reflection at grid-point (x_1, z') is determined only by the wavefield propagating from (x_1, z_0) to (x_1, z') and not by the wavefield propagating from (x_1, z_0) to (x_2, z') . Of course from a ray-theoretical point of view that wouldn't have made sense anyway, but with a wave-theoretical approach this is possible, because of the spatial extent of the wave-fronts. Similarly, the reflection at (x_2, z') is only determined by the wavefield propagation from (x_1, z_0) to (x_2, z') and not by the propagation from (x_1, z_0) to (x_1, z') . Utilising the locally reacting reflectivity matrix to turn the down-going wavefield into an up-going wavefield reads:

$$\begin{pmatrix} R(x_{1,1}) & 0 \\ 0 & R(x_{2,2}) \end{pmatrix} \begin{pmatrix} W^+(x_{1,1}) \\ W^+(x_{2,1}) \end{pmatrix} = \begin{pmatrix} R(x_{1,1})W^+(x_{1,1}) \\ R(x_{2,2})W^+(x_{2,1}) \end{pmatrix}. \quad (\text{B.4})$$

Next, the reflected wavefield is upward propagated to both surface locations with the upward propagation operator $\mathbf{W}^-(z_0, z')$:

$$\begin{aligned} & \begin{pmatrix} W^-(x_{1,1}) & W^-(x_{1,2}) \\ W^-(x_{2,1}) & W^-(x_{2,2}) \end{pmatrix} \begin{pmatrix} R(x_{1,1})W^+(x_{1,1}) \\ R(x_{2,2})W^+(x_{2,1}) \end{pmatrix} = \\ & \begin{pmatrix} W^-(x_{1,1})R(x_{1,1})W^+(x_{1,1}) + W^-(x_{1,2})R(x_{2,2})W^+(x_{2,1}) \\ W^-(x_{2,1})R(x_{1,1})W^+(x_{1,1}) + W^-(x_{2,2})R(x_{2,2})W^+(x_{2,1}) \end{pmatrix}. \end{aligned} \quad (\text{B.5})$$

Because of having only one receiver record with a single detector at (x_2, z_0) , the receiver matrix becomes a row $(0 \quad D_1(x_2))$. Furthermore, the receiver response is assumed to have a flat spectrum such that for the sake of convenience $D_1(x_2)$ can be set to 1.

Multiplying $(0 \quad 1)$ at the left hand side to equation B.5 reads:

$$\begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} W^-(x_{1,1})R(x_{1,1})W^+(x_{1,1}) + W^-(x_{1,2})R(x_{2,2})W^+(x_{2,1}) \\ W^-(x_{2,1})R(x_{1,1})W^+(x_{1,1}) + W^-(x_{2,2})R(x_{2,2})W^+(x_{2,1}) \end{pmatrix} = W^-(x_{2,1})R(x_{1,1})W^+(x_{1,1}) + W^-(x_{2,2})R(x_{2,2})W^+(x_{2,1}). \quad (\text{B.6})$$

The output of the WRW model for the wavefield that is generated by a point source at (x_1, z_0) and received at (x_2, z_0) is displayed in equation B.6, which is clearly a mixture of two wavefields. However, in combination they represent the wavefield that was (in ray-theoretical terms) reflected at the midpoint between x_1 and x_2 . Next investigate Figure B.1 again and comparing it with equation B.6. The first term (read from right to left) is the wavefield that was propagated from the source at (x_1, z_0) to the left grid-point at (x_1, z') , reflected and from there propagated to the receiver at (x_2, z_0) . The second term is the wavefield that was generated by the source at (x_1, z_0) , propagated to the right grid-point at (x_2, z') , reflected and from there propagated to the receiver at (x_2, z_0) . By mixing the travel-times and the reflectivities of both of these wavefields, the actual reflection is generated. It is left to the reader to investigate what wavefield will be received, if the the reflector is not locally reacting, hence, the off-diagonal elements of $R(z')$ are non-zero, as it is usually the case with angle-dependent reflection.

Please note that for a proper description of wavefield propagation with the WRW model the surface points and the grid-points must be distributed densely enough in the lateral direction, such that the Nyquist criterion is obeyed. Otherwise aliasing artefacts occur.

Appendix C

Inversion and Optimisation

C.1 The pseudoinverse

In section 3.2.1 the pseudoinverse of a matrix is used to derive the unknown, densely sampled vectorised focal domain \vec{x} from the coarsely sampled vectorised input data \vec{p} . Please note that within this thesis the data and model space consists of real values (both in the space-time domain). However, it is also possible to define the double focal transformation in the frequency domain. Hereby, the sparse inversion is performed for every frequency separately. In that case data and model space would be complex. Therefore, the general case for complex values is derived here. Furthermore, it is demonstrated how the stabilisation term λ influences the optimisation problem. Please note that $()^*$ denotes the conjugate, $()^T$ the transpose and $()^H$ the adjoint (conjugate-transpose). Without loosing generality \vec{p} and \vec{x} are complex vectors of different length. The mapping from \vec{x} to \vec{p} is defined by the non-square complex matrix \mathbf{L} :

$$\vec{p} = \mathbf{L}\vec{x}. \quad (\text{C.1})$$

C.1.1 Basic definitions

First the derivation with respect to complex numbers is defined. For that purpose any complex number z can be split in real part a and an imaginary part ib . Thus: $z = a + ib$. The complex derivation of the complex number z with respect to z is defined as:

$$\frac{\partial}{\partial z} z = \frac{\partial}{\partial(a + ib)}(a + ib) \equiv \frac{1}{2} \left(\frac{\partial}{\partial a}(a + ib) - i \frac{\partial}{\partial b}(a + ib) \right) = \frac{1}{2}(1 + 1) = 1. \quad (\text{C.2})$$

The complex derivation of the conjugate of z (using $-ib$ for the imaginary part) with respect to z results in zero:

$$\frac{\partial}{\partial z} z^* = 0. \quad (\text{C.3})$$

Consider a scalar t and a complex column vector $\vec{x} = (x_1, x_2)^T$. The derivative of t with respect to \vec{x} gives a column vector:

$$\frac{\partial t}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial t}{\partial x_1} \\ \frac{\partial t}{\partial x_2} \end{pmatrix}. \quad (\text{C.4})$$

C.1.2 Utilised equations

Utilising previous definitions the following result can be obtained:

$$\frac{\partial \vec{x}^H \vec{x}}{\partial \vec{x}} = \vec{x}^*. \quad (\text{C.5})$$

To see this, we can expand the inner product as a function:

$$\vec{x}^H \vec{x} = x_1^* x_1 + x_2^* x_2. \quad (\text{C.6})$$

Taking the derivative with respect to the column vector \vec{x} yields:

$$\frac{\partial \vec{x}^H \vec{x}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial x_1^* x_1 + x_2^* x_2}{\partial x_1} \\ \frac{\partial x_1^* x_1 + x_2^* x_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix}, \quad (\text{C.7})$$

which proofs equation C.5. Note that $\frac{\partial x_1^* x_1 + x_2^* x_2}{\partial x_1}$ is the derivative of a scalar with respect to a scalar. Therefore, the product rule and equations C.2 and C.3 can be used.

Let $\vec{a} = (a_1, a_2)^T$ be a complex column vector. Similarly, we have:

$$\frac{\partial \vec{x}^H \vec{a}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial x_1^* a_1 + x_2^* a_2}{\partial x_1} \\ \frac{\partial x_1^* a_1 + x_2^* a_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (\text{C.8})$$

and:

$$\frac{\partial \vec{a}^H \vec{x}}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial a_1^* x_1 + a_2^* x_2}{\partial x_1} \\ \frac{\partial a_1^* x_1 + a_2^* x_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \vec{a}^*. \quad (\text{C.9})$$

Furthermore, we can derive for the complex matrix \mathbf{B} that:

$$\frac{\partial \vec{x}^H \mathbf{B} \vec{x}}{\partial \vec{x}} = \mathbf{B}^T \vec{x}^*. \quad (\text{C.10})$$

This can be simply verified by expanding the function $\vec{x}^H \mathbf{B} \vec{x}$ and then using previous definitions. For further information on matrix calculus and complex-valued derivation see Petersen and Pedersen (2012).

C.1.3 Derivation of the pseudoinverse

Basing on previous definitions and equations the objective is to proof that:

$$\vec{x} = (\mathbf{L}^H \mathbf{L} + \lambda \mathbf{I})^{-1} \mathbf{L}^H \vec{p} \quad (\text{C.11})$$

is the solution to:

$$\begin{aligned} J &= \|\vec{p} - \mathbf{L} \vec{x}\|_2^2 + \lambda \|\vec{x}\|_2^2 \\ &= (\mathbf{L} \vec{x} - \vec{p})^H (\mathbf{L} \vec{x} - \vec{p}) + \lambda \vec{x}^H \vec{x}. \end{aligned} \quad (\text{C.12})$$

Note that the cost function J is a scalar. Hence, its derivative with respect to column vector \vec{x} will have the dimensions of a column vector (see definition C.4). The objective is to find the minimum of J . Therefore, its derivation with respect to \vec{x} is set to zero:

$$\begin{aligned} \vec{0} &= \frac{\partial J}{\partial \vec{x}} = \frac{\partial (\vec{x}^H \mathbf{L}^H \mathbf{L} \vec{x})}{\partial \vec{x}} - \frac{\partial (\vec{x}^H \mathbf{L}^H \vec{p})}{\partial \vec{x}} - \frac{\partial (\vec{p}^H \mathbf{L} \vec{x})}{\partial \vec{x}} \\ &\quad + \frac{\partial (\vec{p}^H \vec{p})}{\partial \vec{x}} + \lambda \frac{\partial \vec{x}^H \vec{x}}{\partial \vec{x}}. \end{aligned} \quad (\text{C.13})$$

Now, substitute:

$\mathbf{L}^H \mathbf{L} = \mathbf{B}$; $\mathbf{L}^H \vec{p} = \vec{a}$; $\vec{p}^H \mathbf{L} = \vec{a}^H$, which changes equation (C.13) to:

$$\vec{0} = \frac{\partial J}{\partial \vec{x}} = \frac{\partial (\vec{x}^H \mathbf{B} \vec{x})}{\partial \vec{x}} - \frac{\partial (\vec{x}^H \vec{a})}{\partial \vec{x}} - \frac{\partial (\vec{a}^H \vec{x})}{\partial \vec{x}} + \frac{\partial (\vec{p}^H \vec{p})}{\partial \vec{x}} + \lambda \frac{\partial \vec{x}^H \vec{x}}{\partial \vec{x}}. \quad (\text{C.14})$$

For the first, second and third term we use equations C.10, C.8 and C.9. Term four will obviously be a zero vector and term five is defined in equation C.7. The resulting equation reads:

$$\vec{0} = \frac{\partial J}{\partial \vec{x}} = \mathbf{B}^T \vec{x}^* - \vec{0} - \vec{a}^* + \vec{0} + \lambda \vec{x}^*. \quad (\text{C.15})$$

Taking the conjugate of the previous equation leads to:

$$\vec{0} = \mathbf{B}^H \vec{x} - \vec{a} + \lambda \vec{x}. \quad (\text{C.16})$$

After re-substituting \mathbf{B}^H and \vec{a} this reads:

$$(\mathbf{L}^H \mathbf{L} + \lambda \mathbf{I}) \vec{x} = \mathbf{L}^H \vec{p}. \quad (\text{C.17})$$

Solving previous equation for \vec{x} leads to:

$$\vec{x} = (\mathbf{L}^H \mathbf{L} + \lambda \mathbf{I})^{-1} \mathbf{L}^H \vec{p}, \quad q.e.d. \quad (\text{C.18})$$

Please note, setting $\lambda = 0$ in equation C.13 will change equation C.18 to the unconstrained pseudoinverse (setting also $\lambda = 0$ in equation C.18).

C.2 The spgl1 solver

The spgl1 solver was developed for large-scale l_1 regularised least-squares problems by van den Berg and Friedlander (2007). The theoretical foundation of this solver can be found in van den Berg and Friedlander (2008b). In this thesis the spgl1 solver was used to solve the Basis Pursuit Denoise problem for seismic data reconstruction. The method utilises a steepest decent type iterative process, called Spectral Projected Gradient, hereafter referred to as SPG, see Birgin et al. (2003).

As stated on page 49 spgl1 solves the Basis Pursuit Denoise problem:

$$\underset{\vec{x}}{\text{minimise}} \quad ||\vec{x}||_1 \text{ subject to } ||\vec{p} - \mathbf{L}\vec{x}||_2 \leq \sigma. \quad (\text{C.19})$$

However, this problem is closely related to the Lasso problem:

$$\underset{\vec{x}}{\text{minimise}} \quad ||\vec{p} - \mathbf{L}\vec{x}||_2 \text{ subject to } ||\vec{x}||_1 \leq \tau. \quad (\text{C.20})$$

In equation C.19 the l_1 norm of the model space is minimised, while the l_2 norm of the data residual is kept lower than σ . In equation C.20 the l_2 norm of the data residual is minimised while the upper limit of the l_1 of the model space is set to τ . For certain values of σ and τ the optimal points of these two problems coincide. To establish the relationship between the two problems, a function $\Phi(\tau)$ is defined as:

$$\Phi(\tau) = ||\vec{p} - \mathbf{L}\vec{x}_\tau||_2. \quad (\text{C.21})$$

Hereby \vec{x}_τ denotes the optimal solution to equation C.20. Although $\Phi(\tau)$ is a scalar function, in order to evaluate it for certain τ_0 equation C.20 needs to be solved for $\tau = \tau_0$.

From equation C.21 it becomes clear that finding the root of:

$$\Phi(\tau) = \sigma, \quad (\text{C.22})$$

means finding the parameters σ and τ for which the two problems share the same optimal points. Hence, instead of solving equation C.19, equation C.20 can be solved for certain parameters. Efficient tools for solving the Lasso problem exist and one of them is the SPG method.

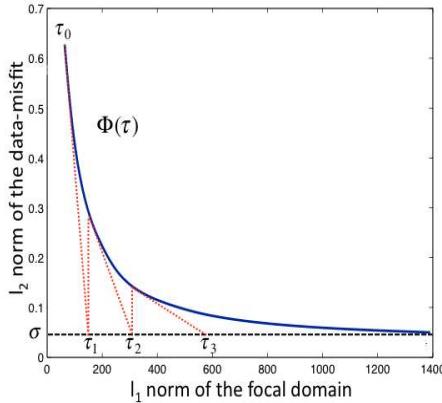


Figure C.1: The blue line denotes the Pareto Curve $\Phi(\tau)$. The red line represents three iterations of Newton's algorithm while finding the root of equation C.22.

Figure C.1 visualises function $\Phi(\tau)$. This curve is often referred to as the Pareto Curve. It describes the trade-off between the sparsity of the solution and the data-misfit. The root of equation C.22 lies on the intersection of the Pareto Curve with the red dotted line (σ). Please note that the spgl1 solver closely tracks the Pareto Curve. It consists of three nested processes:

- At the highest level the method tries to find the root of the non-linear function $\Phi(\tau)$ by utilising the Newton algorithm.
- At a deeper level, every evaluation of $\Phi(\tau)$ for the Newton method requires to solve a Lasso problem (equation C.20), which is done with the SPG method.
- At the deepest level a projection step is performed on each iteration of the SPG method.

C.2.1 The Newton iteration

The root-finding algorithm is an iterative method that can handle non-linear equations like equation C.22. It is described by:

$$\tau_{n+1} = \tau_n - \frac{\Phi(\tau_n) - \sigma}{\Phi'(\tau_n)}, \quad (\text{C.23})$$

where n is the iteration number. In Figure C.1 three iterations towards the root of equation C.22 are displayed. This algorithm generally relies heavily on the given initial condition. However, it can be proven that $\Phi(\tau)$ is convex and non-increasing. Therefore, Newton's method will always converge to the true root. An analytical form of the derivative can be found in van den Berg and Friedlander (2008b). The exact evaluation of $\Phi(\tau)$ and its derivative $\Phi'(\tau)$ is too expensive, since it requires solving exactly a l_1 norm regularised least-squares problem. In the case that only the derivative of $\Phi(\tau)$ is not known exactly, this method is often referred to as inexact Newton's Method and has a super-linear convergence rate. However, in this implementation, both $\Phi(\tau)$ and $\Phi'(\tau)$ are approximate, leading to sub-linear convergence. Please note, however, that the convergence rate of the Newton iteration is controlled by the accuracy of the evaluation of $\Phi(\tau)$.

C.2.2 The spectral projected gradient method

The SPG method is used to solve the least-squares problem required to evaluate $\Phi(\tau)$. The method originates from the steepest descent method, using the Barzilai - Borwein step length, see Barzilai and Borwein (1988). It involves a projection of a candidate solution onto the feasible set, that, in this case, is the l_1 ball with the radius τ . In form of an equation, this becomes:

$$\vec{x}_{n+1} = P_\tau[\vec{x}_n - \alpha \vec{g}_n], \quad (\text{C.24})$$

where \vec{g}_n is the gradient, α the Barzilai - Borwein step length, and P_τ is the projection operator.

In Figure C.2 the first iteration of the SPG method is illustrated for the simple example of two variables. Starting from an initial guess \vec{x}_0 the solver moves along the opposite direction of the gradient \vec{g}_0 for the distance α (step length). This candidate point, depicted as an empty bullet in the drawing, is then projected onto the feasible set, illustrated as the space confined by the dotted lines.

C.2.3 The projection

The projection process resides in the core of the spgl1 solver. Its mathematical definition is given by:

$$P_\tau[\vec{c}] := \{\arg \min_{\vec{x}} \|\vec{c} - \vec{x}\|_2 \text{ subject to } \|\vec{x}\|_1 \leq \tau\}. \quad (\text{C.25})$$

It is implemented as a soft thresholding process that picks the closest to the candidate point that resides in the feasible set, the point \vec{x}_1 in the drawing. This serves as the starting point for the next iteration of the SPG method.

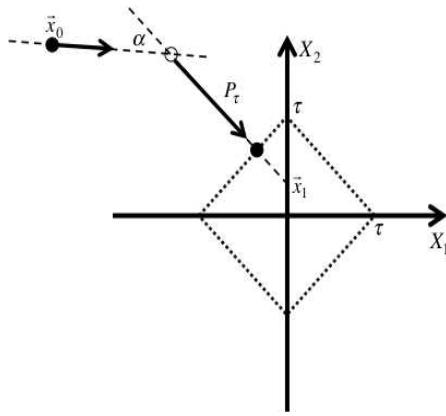


Figure C.2: Schematic representation of an iteration of the SPG method. The dashed line indicates the gradient direction.

C.3 The matching pursuit solver

There are many different implementations of the matching pursuit algorithm, however for data reconstruction it basically consists of these general steps:

1. Set the data residual \vec{r}_i to the coarse input data \vec{p}_0 and initiate an empty model space \vec{x}_i .
2. Transform the residual to the model space.
3. Pick the strongest values and add them to \vec{x}_i .
4. Transform \vec{x}_i to the data space and scale it to fit the input data. The result is the reconstructed data \vec{p}_i .
5. Subtract \vec{p}_i from \vec{p}_0 at the known traces to derive new residual \vec{r}_{i+1} .
6. Go back to step 2 or stop.

Picking the strongest events in step 3 ensures, that no aliasing noise is picked up. When the strong events are subtracted in step 5 only weaker events are left in the residual. These weak events can potentially only generate aliasing noise that is even weaker. If then in the next iteration again the strongest values are picked in the model space it is certain that again only signal content will be selected in step 3. For more information see Wang et al. (2010). Also note the close relation of this method to POCS (Projections Onto Convex Sets). For further information see Abma and Kabir (2006).

Appendix D

Notation

Mathematical notation

- A superscript '+' denotes down-going and a superscript '-' denotes up-going wavefields or propagation operators.
- A hat-symbol denotes an approximated property. For instance the approximated propagation matrix from the surface z_0 to depth level z' is denoted by $\widehat{\mathbf{W}}^+(z', z_0)$.
- Superscript 'T' denotes the transpose, superscript '*' denotes the conjugate and superscript 'H' denotes the adjoint (i.e. the conjugate transpose) of a vector or a matrix.
- Multidimensional data tensors in the time domain are denoted by lower case bold-font letters and have a depth indication, like the data tensor at the surface $\mathbf{p}(z_0)$. In the 2D case this tensor has three dimensions: time, lateral receiver location and lateral source location (both in the x-direction).
- Monochromatic data matrices are denoted by capital bold-font letters and have a depth indication, like the data matrix $\mathbf{P}(z_0)$. In the 2D case the matrix has the dimensions: lateral receiver location and lateral source location (both in the x-direction).
- The vectorised data tensors in the time domain are denoted by lower case letters with a vector-arrow, like the data vector \vec{p} .
- Transformation matrices (explicit) and their implicit counterparts are denoted by a capital bold font letter. For instance the inverse focal transformation is denoted by \mathbf{L} .

- Vector-matrix multiplications are denoted by the absence of a symbol in between them. For example the multiplication of the inverse focal matrix \mathbf{L} with the vectorised focal domain in the time domain \vec{x} is denoted by $\mathbf{L}\vec{x}$. Also for matrix-matrix multiplications, no symbol is used. An example is the downward propagation of the source matrix: $\mathbf{P}(z')^+ = \mathbf{W}^+(z', z_0)S(z_0)$.
- \mathbf{L} denotes the inverse double focal transformation matrix, which describes the propagation (defocusing) of a multi-gather measurement from a deeper level to a higher level (see the enumeration at page 36). \mathbf{L}_r denotes the forward Radon transformation matrix, which is also applied to a multi-gather dataset (see the enumeration at page 92). Hereby every gather is transformed from the time-space domain to the intercept time-slowness domain. If this transformation takes place in the frequency domain a tilde is added. $\widetilde{\mathbf{L}}_r$ then describes the transformation from the space to the slowness domain for one frequency for every gather from a multi-gather measurement. This matrix is defined by equation 6.4 at page 86.

Abbreviations

- AVO: Amplitude Versus Offset
- CFP: Common Focus Point
- CMP: Common Midpoint
- COVT: Common Offset Vector Tile
- EPSI: Estimation of Primaries by Sparse Inversion
- NMO: Normal Move Out
- DMO: Dip Move Out
- AMO: Azimuth Move Out
- POCS: Projections Onto Convex Sets
- spgl1: Spectral Projected Gradient l_1

Bibliography

- Abma, R., and Kabir, N., 2006, 3D interpolation of irregular data with a POCS algorithm: *Geophysics*, **71**, no. 6, E91–E97.
- Abramowitz, M., and Stegun, I. A., 1970, Handbook of mathematical function; with formulas, graphs, and mathematical tables: Dover Publications, Inc.
- Bagaini, C., and Spagnolini, U., 1996, 2-D continuation operators and their applications: *Geophysics*, **61**, no. 6, 1846–1858.
- Barzilai, J., and Borwein, J. M., 1988, Two-point step size gradient methods: *IMA Journal of Numerical Analysis*, **8**, no. 1, 141–148.
- Berkhout, A. J., and Verschuur, D. J., 2005, Removal of internal multiples with the common-focus-point (CFP) approach: Part 1 - Explanation of the theory: *Geophysics*, **70**, no. 3, V45–V60.
- Berkhout, A. J., and Verschuur, D. J., 2006, Focal transformation, an imaging concept for signal restoration and noise removal: *Geophysics*, **71**, no. 6, A55–A59.
- Berkhout, A. J., de Vries, D., Baan, J., and van den Oetelaar, B. W., 1999, A wave field extrapolation approach to acoustical modeling in enclosed spaces: *The Journal of the Acoustical Society of America*, **105**, no. 3, 1725–1733.
- Berkhout, A. J., Blacquière, G., and Verschuur, D. J., 2009, The concept of double blending: Combining incoherent shooting with incoherent sensing: *Geophysics*, **74**, A59–A62.
- Berkhout, A. J., 1982, Seismic migration, imaging of acoustic energy by wave field extrapolation, A: theoretical aspects: Elsevier (second edition).
- Berkhout, A. J., 1987, Applied seismic wave theory: Elsevier.
- Berkhout, A. J., 1993, A unified approach to acoustical reflection imaging, part 1: the forward model: *Journal Acoustical Society of America*, **93**, no. 4, 2005–2017.

- Berkhout, A. J., 1997, Pushing the limits of seismic imaging, part II: integration of prestack migration, velocity estimation and AVO analysis: *Geophysics*, **62**, no. 3, 954–969.
- Berkhout, A. J., 2008, Changing the mindset in seismic acquisition: The Leading Edge, **27**, no. 7, 924–938.
- Birgin, E. G., Mario, J., and Raydan, M. M., 2003, Inexact spectral projected gradient methods on convex sets: *IMA Journal on Numerical Analysis*, **23**, 539–559.
- Cary, P., 1999, Common-offset-vector gathers: An alternative to cross-spreads for wide-azimuth 3-D surveys: *Common-offset-vector gathers: An alternative to cross-spreads for wide-azimuth 3-D surveys*, SEG Technical Program Expanded Abstracts 1999, 1496–1499.
- Chopra, S., Negut, D., and Cilensek, S., 2009, Azimuth moveout transformation - some promising applications: *The Leading Edge*, **28**, no. 5, 510–521.
- Claerbout, J., 1990, Interpolation error filtering:, *in* SEG Technical Program Expanded Abstracts 1990, 1602–1603.
- de Bruin, C. G. M., Wapenaar, C. P. A., and Berkhout, A. J., 1990, Angle-dependent reflectivity by means of prestack migration: *Geophysics*, **55**, no. 9, 1223–1234.
- Deregowski, S. M., 1986, What is DMO?: *in* First Break 1986, 4.
- Donoho, D. L., 2004, For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution: *Comm. Pure Appl. Math.*, **59**, 797–829.
- Doulgeris, P., Mahdad, A., and Blacquiere, G., 2009, An R-based overview of the WRW concept: 3rd International Conference and Exhibition, Nafplion, Greece, 21–26 June, 2009.
- Doulgeris, P., Mahdad, A., and Blacquiere, G., 2010, Separation of blended data by iterative estimation and subtraction of interference noise: Separation of blended data by iterative estimation and subtraction of interference noise:, SEG, 3514–3518.
- Doulgeris, P., Mahdad, A., and Blacquiere, G., 2011, Iterative separation of blended marine data: Discussion on the coherence-pass filter: Iterative separation of blended marine data: Discussion on the coherence-pass filter:, SEG, 26–31.
- Doulgeris, P., Verschuur, D. J., and Blacquiere, G., 2012, Separation of blended data by sparse inversion utilizing surface-related multiples: 74rd Ann. Internat. Mtg., Eur. Ass. of Expl. Geophys., Expanded abstracts.

- Downton, J., Holy, D., Trad, D., Hunt, L., Reynolds, S., and Hadley, S., 2010, The effect of interpolation on imaging and azimuthal AVO: A nordegg case study: SEG Technical Program Expanded Abstracts, **29**, no. 1, 383–387.
- Duijndam, A., Schonewille, M., and Hindriks, C., 1999, Reconstruction of band-limited signals, irregularly sampled along one spatial direction: Geophysics, **64**, no. 2, 524–538.
- Fokkema, J. T. ., and van den Berg, P. M., 1993, Seismic applications of acoustic reciprocity: Elsevier Science Publishers B.V.
- Fomel, S., and Claerbout, J. F., 2001, Three-dimensional seismic data regularization: Ph.D. thesis, Stanford University.
- Gill, P., Wang, A., and Molnar, A., 2011, The in-crowd algorithm for fast basis pursuit denoising: Signal Processing, IEEE Transactions on, **59**, no. 10, 4595–4605.
- Gisolf, A., and Verschuur, D. J., 2010, The principles of quantitative acoustical imaging: EAGE publications bv.
- Hennenfent, G., and Herrmann, F. J., 2008, Simply denoise: Wavefield reconstruction via jittered undersampling: Geophysics, **73**, no. 3, V19–V28.
- Hennenfent, G., Fenelon, L., and Herrmann, F. J., 2010, Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach: Geophysics, **75**, no. 6, WB203–WB210.
- Herrmann, P., Mojesky, T., Magasan, M., and Hugonnet, P., 2000, De-aliased, high-resolution Radon transforms: 70th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded abstracts, 1953–1956.
- Kabir, M. M. N., and Verschuur, D. J., 1995, Restoration of missing offsets by parabolic radon transform: Geophys. Prosp., **43**, no. 3, 347–368.
- Kinneging, N. K., Budejicky, V., Wapenaar, C. P. A., and Berkhouwt, A. J., 1989, Efficient 2D and 3D shot record redatuming: Geophys. Prosp., **37**, no. 5, 493–530.
- Knutsson, H., and Westin, C. F., June 16-19 1993, Normalized and differential convolution: Methods for interpolation and filtering of incomplete and uncertain data:, in Proceedings of Computer Vision and Pattern Recognition ('93), 515–523.
- Kostov, C., 1990, Toeplitz structure in slant-stack inversion: SEG Technical Program Expanded Abstracts, **9**, no. 1, 1618–1621.
- Kuehl, H., 2002, Least-squares wave-equation migration/inversion: Ph.D. thesis, University of Alberta (Canada).

- Kutsch, H., and Verschuur, D. J., 2012a, The double focal transformation and its application to seismic data reconstruction: 74th Ann. Internat. Mtg., Eur. Ass. of Geosc. and Eng., Expanded abstracts.
- 2012b, Data reconstruction via sparse double focal transformation: An overview.: IEEE Signal Process. Mag., **29**, no. 4, 53–60.
- Kutsch, H., Verschuur, D., and Berkhouit, A., 2010, High resolution double focal transformation and its application to data reconstruction:, *in* SEG Technical Program Expanded Abstracts 2010.
- Lopez, G. A., and Verschuur, D. J., 2013a, 3D primary estimation by sparse inversion using the focal domain parameterization: 83th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded abstracts.
- 2013b, Estimation of primaries by sparse inversion - parameterization via the focal domain: 75rd Ann. Internat. Mtg., Eur. Ass. of Expl. Geophys., Expanded abstracts.
- Mahdad, A., Doulgeris, P., and Blacquiere, G., 2011, Separation of blended data by iterative estimation and subtraction of blending interference noise: Geophysics, **76**, no. 3, Q9–Q17.
- Naghizadeh, M., and Sacchi, M. D., 2010, Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data: Geophysics, **75**, no. 6, WB189–WB202.
- Naghizadeh, M., 2012, Seismic data interpolation and denoising in the frequency-wavenumber domain: Geophysics, **77**, no. 2, V71–V80.
- Nemeth, T., Wu, C., and Schuster, G., 1999, Least squares migration of incomplete reflection data: Geophysics, **64**, no. 1, 208–221.
- Paige, C. C., and Saunders, M. A., March 1982, Lsqqr: An algorithm for sparse linear equations and sparse least squares: ACM Trans. Math. Softw., **8**, no. 1, 43–71.
- Petersen, K. B., and Pedersen, M. S. The matrix cookbook:, nov 2012. Version 20121115.
- Schonewille, M., Klaedtke, A., and Vigner, A., 2009, Anti alias anti leakage Fourier transform:, *in* SEG Technical Program Expanded Abstracts 2009, 3249–3253.
- Schultz, P. S., and Claerbout, J. F., 1978, Velocity estimation and downward continuation by wavefront synthesis: Geophysics, **43**, no. 4, 691–714.
- Spagnolini, U., and Opreni, S., 1996, 3D shot continuation operator:, *in* SEG Technical Program Expanded Abstracts 1996, 439–442.

- Spitz, S., 1991, Seismic trace interpolation in the f-x domain: *Geophysics*, **56**, no. 6, 785–794.
- Thorbecke, J. W., and Berkhout, A. J., 1994, 3-D recursive extrapolation operators: An overview: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded abstracts, 1262–1265.
- Thorbecke, J. W., 1997, Common focus point technology: Ph.D. thesis, Delft University of Technology.
- Trad, D. O., Ulrych, T., and Sacchi, M., 2003, Latest views of the sparse radon transform: *Geophysics*, **68**, 386–399.
- Trad, D. O., 2002, Accurate interpolation with high-resolution time-variant radon transforms: *Geophysics*, **67**, no. 2, 644–656.
- Trad, D. O., 2003, Interpolation and multiple attenuation with migration operators: *Geophysics*, **68**, 2043–2054.
- Trad, D. O., 2009, Five-dimensional interpolation: Recovering from acquisition constraints: *Geophysics*, **74**, no. 6, V123–V132.
- van den Berg, E., and Friedlander, M. P. SPGL1: A solver for large-scale sparse reconstruction:, June 2007. <http://www.cs.ubc.ca/labs/scl/spgl1>.
- van den Berg, E., and Friedlander, M. P., 2008a, Probing the pareto frontier for basis pursuit solutions: *SIAM J. Sci. Comput.*, **31**, no. 2, 890–912.
- 2008b, Probing the pareto frontier for basis pursuit solutions: *SIAM Journal on Scientific Computing*, **31**, no. 2, 890–912.
- van Groenestijn, G. J. A., and Verschuur, D. J., 2009a, Estimating primaries by sparse inversion and application to near-offset data reconstruction: *Geophysics*, **74**, A23–A28.
- 2009b, Estimation of primaries and near offsets by sparse inversion: Marine data applications: *Geophysics*, **74**, R119–R128.
- Vermeer, G. J. O., 2002, 3-d seismic survey design: Society of Exploration Geophysicists.
- Verschuur, D. J., and Berkhout, A. J., 1997, Estimation of multiple scattering by iterative inversion, part II: practical aspects and examples: *Geophysics*, **62**, no. 5, 1596–1611.
- Verschuur, D. J., and Kutscha, H., 2009, High resolution reconstruction of irregularly sampled, aliased measurements: International Conference on Acoustics, Proceedings of NAG/DAGA 2009.

- Verschuur, D. J., and Marhfoul, B. E., 2005, Estimation of complex near surface focusing operators by global optimization: 67th Ann. Internat. Mtg., Eur. Ass. of Geosc. and Eng., Expanded abstracts, A035.
- Verschuur, E., Vrolijk, J. W., and Tsingas, C., 2012, 4d reconstruction of wide azimuth (waz) data using sparse inversion of hybrid radon transforms: *in* SEG Technical Program Expanded Abstracts 2012, 1–5.
- Verschuur, D. J., 2006, Seismic multiple removal techniques - past, present and future: EAGE Publications BV.
- Wang, J., Ng, M., and Perz, M., 2010, Seismic data interpolation by greedy local radon transform: *Geophysics*, **75**, no. 6, WB225–WB234.
- Wapenaar, C. P. A., and Berkhouit, A. J., 1989, Elastic wave field extrapolation: redatuming of single- and multi-component seismic data: Elsevier Science Publ. Co., Inc.
- Wapenaar, C. P. A., Verschuur, D. J., and Herrmann, P., 1992, Amplitude pre-processing of single- and multicomponent seismic data: *Geophysics*, **57**, no. 9, 1178–1188.
- Xu, S., Zhang, Y., Pham, D., and Lambare, G., 2005, Antileakage fourier transform for seismic data regularization: *Geophysics*, **70**, no. 4, V87–V95.
- Zwartjes, P. M., and Gisolf, A., 2006, Fourier reconstruction of marine-streamer data in four spatial coordinates: *Geophysics*, **71**, no. 6, V171–V186.
- Zwartjes, P. M., and Gisolf, A., 2007, Fourier reconstruction with sparse inversion: *Geophys. Prosp.*, **55**, no. 2, 199–221.
- Zwartjes, P., and Sacchi, M., 2007, Fourier reconstruction of nonuniformly sampled, aliased seismic data: *Geophysics*, **72**, no. 1, V21–V32.
- Zwartjes, P. M., 2005, Fourier reconstruction with sparse inversion: Ph.D. thesis, Delft University of Technology.

Summary

Many seismic data processing and imaging processes require densely and regularly sampled data, whereas the actual measurements are mostly irregularly and sparsely sampled. Therefore, seismic data reconstruction methods are utilised as a pre-processing step.

Within the class of transformation-based reconstruction techniques, observed seismic data is decomposed into certain basis functions, such as plane waves, parabolae or curvelets. In the corresponding model space the aliasing noise is assumed to have different properties than the seismic signal and can be suppressed. However, in many cases subsurface information is available that cannot be used in these traditional reconstruction methods. Therefore, the double focal transformation was derived as a way to incorporate knowledge about the subsurface in the reconstruction algorithm.

The basic principle of the double focal transformation is to focus seismic energy by a back-propagation of the seismic data at the source and receiver side to certain depth levels. As a result, the seismic data are represented by a limited number of samples in the focal domain in a localized area, whereas aliasing noise spreads out. By imposing a sparse solution in the focal domain, aliasing noise is suppressed and data reconstruction beyond aliasing is achieved.

To facilitate the process, only a few effective depth levels need to be included, preferably along the major boundaries in the subsurface. Propagation operators from these boundaries to the surface (focal operators) serve as the basis functions of this data decomposition method. Including more depth levels allows a sparser data representation, and hence, increases the reconstruction capability. The more precise the subsurface information is known, the more accurate these propagators can be computed. However, very precise operators are not necessary for a good reconstruction result, because in the reconstruction step (the inverse focal transformation) the effect of these operators is again removed.

The calculation of the double focal transformation requires a non-linear inversion process, where the samples in the focal domain are estimated such that they - after inverse transformation - match the input data at the measurement locations. Because the inversion process is under-determined, an extra constraint on the focal domain is applied, for which the minimum L1 norm is chosen. This forces the

distribution in the focal domain to be sparse and - thereby - suppresses the aliasing noise. For the inversion a so-called spgl1 solver has been used that is guaranteed to converge to the desired minimum of the defined objective function. It utilises a steepest decent type iterative process, called Spectral Projected Gradient. Seismic data reconstruction via the double focal transform method appears to be robust against inaccuracies in the focal operators up to roughly ten percent velocity error. Furthermore, the method was extended to the full 3D case, where each focal transform sub-domain in principle contains a 5D data space. In addition to the basic focal transformation, the method can be combined with other transforms in order to increase data compression. As an example, the double focal transformation can be combined with the linear Radon transformation, such that the seismic data can represented sparser and fewer focal operators are necessary. Satisfactory results of focal domain data reconstruction beyond aliasing on 2D and 3D synthetic and 2D field data illustrate the method's virtues.

Samenvatting

Veel seismische data verwerkings- en beeldvormingsprocessen vereisen regelmatig bemonsterde data met een hoge dichtheid, terwijl de metingen in de praktijk onregelmatig en dun bemonsterd zijn. Daarom worden reconstructiemethoden gebruikt als een voorbewerkingstap voor deze metingen.

Binnen de klasse van transformatie gebaseerde reconstructietechnieken, wordt de waargenomen seismische data ontbonden in bepaalde basisfuncties, zoals vlakke golven, parabolen of curvelets. Vervolgens wordt aangenomen dat in de desbetreffende modelruimte de bemonsteringsruis andere eigenschappen heeft dan seismische signalen en dus kan worden onderdrukt. Echter, in veel gevallen is informatie over de ondergrond beschikbaar welke niet kan worden gebruikt in deze traditionele reconstructiemethoden. Daarom is de dubbele focale transformatie voorgesteld als een manier om kennis over de ondergrond in het reconstructiealgoritme mee te nemen.

Het basisprincipe van de dubbele focale transformatie is het focussen van seismische energie door een inverse propagatie van de seismische metingen vanaf de bron en ontvanger naar bepaalde diepte niveaus. Hierdoor wordt de seismische data weergegeven in een beperkt aantal punten in het focale domein in een gelokaliseerd gebied, terwijl bemonsteringsruis wordt uitgespreid over het hele domein. Door het opleggen van een schaarse oplossing in het focale domein wordt bemonsteringsruis onderdrukt en een optimale datareconstructie bewerkstelligd.

Om het proces mogelijk te maken, zijn alleen maar een paar effectieve diepte niveaus nodig, bij voorkeur langs de belangrijke grenslagen in de ondergrond. Propagatie operatoren van deze grenslagen naar het oppervlak (focale operatoren) dienen als de basis functies van deze decompositiemethode. Het inbrengen van meer diepte niveaus laat een schaarsere data representatie toe, en dus verhoogt het reconstructievermogen. Des te nauwkeuriger informatie over de ondergrond bekend is, des te nauwkeuriger deze propagatie operatoren kunnen worden berekend. Echter, zeer nauwkeurige operatoren zijn niet nodig voor een goed reconstructieresultaat, omdat bij de reconstructie stap (de inverse focale transformatie) het effect van deze operatoren weer wordt verwijderd.

De berekening van de dubbele focale transformatie vereist een niet - lineair inversie proces, waarbij de waarden van de punten in het focale domain worden

geschat zodat zij - na inverse transformatie - overeenkomen met de input data op de meetlocaties. Omdat het inversie proces ondergedetermineerd is, wordt een extra beperking op het focale domein toegepast, waarvoor de L1 norm wordt gekozen. Dat dwingt de distributie in het focale domain schaars te zijn en onderdrukt daardoor de bemonsteringsruis. Voor de inversie is een zogenaamde spgl1 oplosser gebruikt, die gegarandeerd convergeert naar het gewenste minimum van de gedefinieerde minimalisatiefunctie. Het maakt gebruik van een ‘steepest decent’ type iteratief proces, genaamd ‘Spectral Projected Gradient’.

Seismische data reconstructie via de dubbele focale transformatie methode blijkt robuust te zijn tegen fouten in de focale operatoren tot ongeveer 10% snelheidfout. Bovendien werd de methode uitgebreid naar de 3D situatie, waarbij ieder focale sub-domein in principe een 5D dataruimte bevat. In aanvulling op de basis focale transformatie, kan de methode ook worden gecombineerd met andere transformaties om de datacompressie te verhogen. Als voorbeeld, een dubbele focale transformatie kan gecombineerd worden met de lineaire Radon transformatie. Daardoor kunnen de seismische metingen nog compacter worden weergegeven, hetgeen de reconstructiemogelijkheden verbetert.

Bevredigende resultaten van de dubbele focale reconstructie met onder-bemonsterde metingen voor 2D en 3D synthetische metingen en 2D veldmetingen illustreren de goede potentie van deze methode.

CV note

Hannes Kutschä was born in Sondershausen, Germany in 1982. He received his diploma in Geophysics in 2008 from the Westfälische Wilhelmsuniversität Münster (University of Münster) in Germany. His specialisation was the investigation of a geologically interesting area with reflection seismic and refraction seismic measurements.

He joined the Imaging Science and Technology department of Delft University of Technology in April 2008. During his time as PhD candidate he performed research for the DELPHI Consortium under the supervision of Dr. ir. D. J. (Eric) Verschuur, who's main interests are seismic modelling, data processing and imaging techniques.

Acknowledgements

First, I would like to use this opportunity to express my deep gratitude for the possibility of doing a PhD thesis in one of the best research groups in geophysics around the world.

I am grateful to my supervisor and co-promotor Eric Verschuur for the many inspiring discussions we had and for the time and effort he has put into my thesis. His friendly and informal way to deal with people made working with him very enjoyable. My grateful thanks are extended to my promoter Dries Gisolf for his support of my work. His honest and critical looks into presentations, Delphi chapters and publications were always very valuable to me. I also gratefully recognise that Dries provided his office for the annual "Christmas drinks" and financially supported the annual "Christmas lunch". Professor Berkhou accepted me in the Delphi family and provided with his WRW model the theoretical foundation of my work. Describing rather complex processes was greatly simplified by the detail hiding operator notation of Professor Berkhou. Also, his inspirational detail hiding talks and presentations taught me how to communicate efficiently with an inhomogeneous audience.

One of the great advantages to be a PhD of the Delphi Consortium is the possibility to directly communicate with representatives of the Oil and Gas industry during the Delphi meetings. In that way it is possible to keep the focus of the research to realistic and relevant problems. In this regard I would like to express my gratitude to the sponsors of the Delphi consortium for their financial support and constructive feedbacks and discussions. Furthermore, I had the opportunity to do two internships during my PhD. The first I did with FairfieldNodal, where I implemented the double focal reconstruction in the frequency domain. In this regard Josef Paffenholz and Paul Docherty are greatly acknowledged for inviting me and letting me have a fantastic, unforgettable time there. My second internship I did at the EXPEC Advanced Research Center in Dhahran, Saudi Arabia, where I extended the concept of double focal transformation to three dimensions. Thanks to many encouraging helpful discussions with Constantinos Tsingas, thanks to the great hospitality of Panos Kelamis and Saudi Aramco and thanks to the wonderful working atmosphere at EXPEC ARC this was an outstanding experience.

At the beginning of my PhD Prof. Felix J. Herrmann visited our research group

at the TU Delft for a sabbatical and brought his former student Tim Lin along. They shared their insights about convex optimisation and brought the "spgl1" solver to my attention. In this regard I also want to acknowledge Ewout van den Berg and Michael P. Friedlander from the Department of Computer Science, University of British Columbia, Canada for their continuing support.

The PhD defense committee is highly appreciated for their comments and suggestions, which further improved my thesis.

WesternGeco is acknowledged for providing the Gulf of Mexico field data. In any working environment not only the conditions but also the colleagues are very important. I consider myself very lucky to have had such amazing companions. With all the below mentioned people I share many unforgettable memories. Also their doors were open at any time for questions and ideas. I particularly recognise my great colleagues, good friends and even former house mates Araz, Peter and Panos who made my PhD very enjoyable. I am, furthermore, a proud member of the "WRW-rescue team", which next to me consists of Gert-Jan, Ahmad, Mohammad and Khalid. Also my special friend and colleague Alexey should be mentioned here, who often supported me with mathematical help. Xander had great patience in discussing scientific issues with me and gave me the possibility to practice my Dutch with him. Furthermore, Alok, Mikhayl, Gabriel, Gabrio, Apostolos, Abdulrahman, Justyna, Dana, Rolf and Jan Thorbecke are recognised for all the interesting and helpful discussions and the sharing of knowledge and ideas. Sincere thanks are given to Jan Willem, who kindly helped me with some parameter tweaking for the linear Radon reconstruction results and the translation of my summary into Dutch. Also many thanks to Yue Ma, a Geophysicist at Saudi Aramco for some help with mathematical derivations.

I express my great appreciation to my former colleagues from the medical acoustic group and the room acoustic group. Foremost, I have to mention Anton and Lars, who helped me to settle in at the beginning of my PhD. Also, many thanks to Erwin, Suzanne, Neslihan, Silvian and Libertario for having a good time with me.

In an environment that deals with a variety of software and hardware the role of a system administrator cannot be overestimated. Edo not only knows solutions to programming issues or Seismic Unix problems but also is a source for useful tips and tricks regarding things like booking the fastest flight, find the best cab or avoiding long cues at conferences. Henry visited me at least a million times in my office to help with software installations. He also shares my affection for outdoor activities, travelling and beautiful places.

Of course, like in any group the true power is held by the secretary. Thank you Margaret for your great administrative support and for organising annual events like "Christmas drinks", "Christmas lunch", "international lunch" and the "day out". I appreciated your special sense of humour and your stamina answering my endless questions.

Diemer de Vries is recognised for culturally enriching some of my lunch breaks by practising double bass with me.

What would be life without friends? I had a great time with many people during my PhD. Especially, Alex, Anke, Florian and Sonia. But also many others contributed positively to my distraction after work.

If somebody expected to find his or her name here but doesn't please forgive me as it happened unintentionally.

I would like to thank my family for their encouragement, trust and love. Having three siblings is a blessing and was very valuable for my personal development. A special thanks to my mother for her outstanding support before and during my professional career. Many thanks to my father, who studied geophysics him self and, thus, fascinated me already as a child with explanations and stories about earth and the universe.

Last but certainly not least I would like to thank my beloved partner in life Dimka (a.k.a. Dimi) for just being who she is. Her love and trust in me were always a great motivation and an inexhaustible source of energy. Thank you Dimi for having patience, stamina and confidence during our weekend-relationship and for recently leaving your precious Germany and your great friends to join me in a common future in the Netherlands.

*Delft,
January 2014*

Hannes Kutschä