Carlos Alberto da Costa Filho

Applications of independent component analysis to the attenuation of multiple reflections in seismic data

Aplicações da análise de componentes independentes à atenuação de reflexões múltiplas em dados sísmicos

Campinas
2013

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica

Carlos Alberto da Costa Filho

Applications of independent component analysis to the attenuation of multiple reflections in seismic data

Orientador: Prof. Dr. Martin Tygel

Aplicações da análise de componentes independentes à atenuação de reflexões múltiplas em dados sísmicos

Master's thesis submitted to the Institute of Mathematics, Statistics and Scientific Computing at Unicamp for the obtainment of the title of Master in Applied Mathematics.

Dissertação de Mestrado apresentada ao Instituto de Matemática, Estatística e Computação Científica da Unicamp para a obtenção do título de Mestre em Matemática Aplicada.

Este exemplar corresponde à versão final da dissertação defendida pelo aluno Carlos Alberto da Costa Filho, e orientada pelo Prof. Dr. Martin Tygel.

Assinatura do Orientador

Campinas
2013

FICHA CATALOGRÁFICA ELABORADA POR
MARIA FABIANA BEZERRA MULLER - CRB8/6162
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA - UNICAMP

Costa Filho, Carlos Alberto da, 1988-
C823a Aplicações da análise de componentes independentes à atenuação de reflexões múltiplas em dados sísmicos / Carlos Alberto da Costa Filho. – Campinas, SP : [s.n.], 2013.

Orientador: Martin Tygel.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Método sísmico de reflexão. 2. Ondas sísmicas. 3. Processamento de sinais. 4. Análise de componentes independentes. I. Tygel, Martin,1946-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em inglês: Applications of independent component analysis to the attenuation of multiple reflections in seismic data
Palavras-chave em inglês:
Seismic reflection method
Seismic waves
Signal processing
Independent component analysis
Área de concentração: Matemática Aplicada
Titulação: Mestre em Matemática Aplicada
Banca examinadora:
Martin Tygel [Orientador]
Leonardo Tomazeli Duarte
Wander Nogueira de Amorim
Data de defesa: 15-03-2013
Programa de Pós-Graduação: Matemática Aplicada

Dissertação de Mestrado defendida em 15 de março de 2013 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Prof.(a). Dr(a). MARTIN TYGEL

Prof.(a). Dr(a). LEONARDO TOMAZELI DUARTE

Prof.(a). Dr(a). WANDER NOGUEIRA DE AMORIM

Epigraph

16.
Aufwärts.

„Wie komm ich am besten den Berg hinan?“
Steig nur hinauf und denk nicht dran!

Friedrich Wilhelm Nietzsche
Die fröhliche Wissenschaft.

16.
Upward

“How do I best get to the top of this hill?”
Climb it, don’t think it! and maybe you will.

Friedrich Wilhelm Nietzsche
The Gay Science

28.
Trost für Anfänger.

Seht das Kind umgrunzt von Schweinen,
Hülflos, mit verkrümmten Zeh’n!
Weinen kann es, Nichts als weinen —
Lernt es jemals stehn und gehn?
Unverzagt! Bald, sollt’ ich meinen,
Könnt das Kind ihr tanzen sehn!
Steht es erst auf beiden Beinen,
Wird’s auch auf dem Kopfe stehn.

Friedrich Wilhelm Nietzsche
Die fröhliche Wissenschaft.

28.
Consolation for Beginners

See the child, with pigs she’s lying,
helpless, face as white as chalk!
Crying only, only crying —
will she ever learn to walk?
Don’t give up! Stop your sighing,
soon she’s dancing ‘round the clock!
Once her own two legs are trying,
she’ll stand on her head and mock.

Friedrich Wilhelm Nietzsche
The Gay Science

A vida é a arte do encontro,
embora haja tanto desencontro pela vida.

Vinicius de Moraes
Samba da Benção

Life is the art of encounters,
though there are so many misconnections in life.

Vinicius de Moraes
Samba da Benção

Acknowledgments

Agradeço em primeiro lugar à minha mãe Rosane, que através dos seus sacrifícios tornou possível esta dissertação. Agradeço também à minha irmã Tais e meu pai Carlos pelo apoio e carinho.

Pelas oportunidades oferecidas, pela liberdade dada, e pelo encorajamento durante estes dois anos, devo agradecer ao meu orientador Martin Tygel. Devo um agradecimento especial ao Leonardo Duarte, que me guiou pelos mares do ICA, além de sempre incentivar o trabalho. Também agradeço ao Wander Amorim que emprestou seus olhos de geofísico ao trabalho. Ao Paulo de Carvalho, agradeço a ajuda na parte do SRME.

Agradeço aos amigos que fiz na Unicamp, que tanto me ajudaram e por que sem eles esta dissertação não teria sido concretizada. Também agradeço aos amigos que deixei no Rio, por haver estado tão presentes nesta jornada, mesmo distantes.

Gostaria de agradecer aos professores que tive o prazer de conhecer e que contribuíram para o meu amadurecimento acadêmico.

Agradeço finalmente às seguintes instituições: Unicamp, IMECC, CEPETRO e LGC, pela infraestrutura; CNPq pela bolsa concedida; Petrobras pelos dados; consórcio DELPHI pelo *software*.

Resumo

As reflexões de ondas sísmicas na subsuperfície terrestre podem ser colocadas em duas categorias disjuntas: reflexões primárias e múltiplas. Reflexões primárias carregam informações pontuais sobre um refletor específico, enquanto reflexões múltiplas carregam informações sobre interfaces e pontos de reflexão variados. Consequentemente, é usual tentar atenuar reflexões múltiplas e trabalhar somente com reflexões primárias. Neste trabalho, a teoria de ondas acústicas é desenvolvida somente a partir da equação da onda. Um resultado que demonstra como a propagação de ondas acústicas pode ser descrita somente com uma única multiplicação por matriz é exposta. Este resultado permite que um algoritmo seja desenvolvido que, em teoria, pode ser usado para remover todas as reflexões múltiplas que refletiram na superfície pelo menos uma vez. Uma implementação prática deste algoritmo é mostrada. Por conseguinte, a teoria de análise de componentes independentes é apresentada. Suas considerações teóricas e práticas são abordadas. Finalmente, ela é usada em conjunção com o método de eliminação de múltiplas de superfície para atenuar múltiplas de quatro dados diferentes. Estes resultados são então analisados e a eficácia do método é avaliada.

Palavras-chave: Método sísmico de reflexão, Ondas sísmicas, Processamento de sinais, Análise de componentes independentes

Abstract

The reflections of seismic waves in the subsurface of the Earth can be placed under two disjoint categories: primary and multiple reflections. Primary reflections carry point-wise information about a specific reflector while multiple reflections carry informations about various interfaces and reflection points. Consequently, it is customary to attempt to attenuate multiple reflections and work solely with primary reflections. In this work, the theory of acoustic waves is developed solely from the wave equation. A result that shows how acoustic wave propagation can be described as a single matrix multiplication is exposed. This result enables one to develop an algorithm that, in theory, can be used to remove all multiple reflections that haved reflected on the surface at least once. The practical implementation of this algorithm is shown. Thereafter, the theory of independent component analysis is presented. Its theoretical and practical considerations are addressed. Finally, it is used in conjuction with the surface-related multiple elimination method to attenuate multiples in four different datasets. These results are then analyzed and the efficacy of the method is evaluated.

Keywords: Seismic reflection method, Seismic waves, Signal processing, Independent component analysis

List of Figures

List of Tables

Nomenclature

BSS Blind Source Separation

CLT Central Limit Theorem

FSME Free-Surface Multiple Elimination

IC Independent Component

ICA Independent Component Analysis

KKT Karush-Kuhn-Tucker

KL Kullback-Leibler

PCA Principal Component Analysis

pdf Probability Density Function

SRME Surface-Related Multiple Elimination

Contents

Chapter 1

Introduction

Все счастливые семьи похожи друг на друга, каждая несчастливая семья несчастлива по-своему.

Лев Николаевич Толстой
Анна Каренина

Every direct problem resembles on another; each inverse problem is inverse in it's own way.

1.1 Description of the problem

A typical problem in exploration geophysics is the imaging of the distribution of materials in the interior of the Earth. These materials are commonly arranged in a stratified fashion, that is, one layer sits atop another. These layers possess their own set of physical properties, such as density, porosity, permeability, etc., that may or may not vary within the layer. One particular property that these layers have is the velocity with which seismic waves propagate within them. This is interesting to the problem because when seismic waves cross from one medium to another medium with differing velocity, part of the wave will be reflected, while another will be refracted. Furthermore, the reflected wave will carry information of the media it has traveled through, and of the medium it reflected from. The reflected waves can be recorded and their analysis can give rise to information

of the disposition of material in the subsurface. Roughly, this is the method of reflection seismology.

Amongst the many methods in exploration geophysics, reflection seismology is by far the most common one used when attempting to build a detailed image of the Earth's interior. It uses the reflection of seismic waves to estimate the positions of layers inside the earth and some of their physical properties. A typical experiment consists in creating seismic waves and measuring their reflections in what is called a trace. These seismic waves can reflect and refract multiples times. The events that come from a wave that reflected once is called a primary,[*] while an event that derives from a wave that was reflected more than once is called a multiple.

1.2 Multiples and their removal

The primary reflections have the interesting property that they carry information of the medium it traveled in and of the reflector it rebounded on. On the other hand, multiple reflections carry information of many reflectors. Such a distinction is important since it means that by studying primary reflections, one may gain information of its associated reflector, something that cannot be done with multiple reflections. Consequently, primary reflections are considered more valuable in the seismic reflection method.

Furthermore, the most common imaging algorithms make the assumption that the received waves only reflected once, that is, it assumes all events are primary. However, for all but the most simple geologies, this is untrue, and multiples arrive intermixed with primaries. Therefore, in order to use the gathered data in these algorithms, one must first remove the multiples.

Verschuur (2006) and Yilmaz (2001) separate multiple removal methods in two categories: methods that rely on the different spacial behavior of multiples, and methods that use their predictability. Methods of the first type are based on the assumption that multiples have differing (moveout) velocities, and exhibit different local dips when compared to primaries. These methods include: f-k, slant-stack and Radon filtering, and high-resolution Radon filtering. Methods of the second type are based on the fact that, if the velocity distribution and the source wavelet are known, then it is possible to predict the multiples. These methods are usually divided in two parts: prediction and subtraction. The prediction step concerns itself with estimating the source wavelet and

[*] A primary event is not to be confused with a P-wave. The adjective *primary* in the former refers to the type of reflection event, while in the latter, refers to the type of wave (compressional or longitudinal)

effectively predicting the multiples. Estimation of the wavelet can be done with methods like predictive deconvolution and blind deconvolution.

For the prediction of the multiples, there exists methods such as surface-related multiple elimination (SRME), free-surface multiple elimination (FSME) and internal multiple removal by inverse scattering.

In the present work, the methods of SRME will be studied, in connection to independent component analysis (ICA).

1.3 Independent component analysis

ICA is a technique which, among other applications, can solve the problem of blind deconvolution. ICA can also solve the more general case of blind source separation (BSS). The problem of blind source separation is the following: given a combination of a number of source signals, separate these "mixtures" into their individual, original signals. The fact that how these combinations are performed are unknown gives meaning to the term *blind*. It is quite clear that in such formulation, the problem is tremendously ill-posed. Depending on the mixtures, it can impossible solve uniquely, or even impossible to solve at all. Therefore, in order to recover the sources, additional hypotheses are necessary. Different hypotheses lead to different solutions. The method of ICA supposes that the sources are statistically mutually independent. Furthermore, in linear ICA, which will be used exclusively, the mixtures are linear combinations of the sources. With these *a priori* informations, the problem can be solved uniquely.[†]

The purpose of this work is to exploit the use of ICA in the extraction of both the primary and the multiple reflections. While the use of ICA for blind deconvolution has been used since the 1990s, its application to seismic signal processing is more recent, dating from the beginning of the 2000s. Newer still is the application of ICA in multiple removal, having being recently pioneered in Lu (2006), Kaplan and Innanen (2008) and Lu and Liu (2009).

1.4 SRME and ICA

The main focus of this work will be concerned with the SRME algorithm and how the technique of ICA can be used to enhance it. In order to develop the theory of SRME, one

[†]Under a few more assumptions, and except for scaling constants and permutations.

must first study the mathematical description of acoustic wave phenomena. Chapter 2 treats this subject in a way that will make the theoretical development of SRME simpler. The following chapter will do exactly that, describe how the algorithm of SRME works, in both one and two dimensions. In Chapter 4 the problem of ICA is more precisely formulated. A technique for performing ICA is also developed, based on negentropy.

With both SRME and ICA algorithms appropriately derived and described, Chapter 5 exposes how both can be used in conjunction to attenuate multiples. Synthetic and field data are presented showing positive results, which are then analyzed in the conclusion. The conclusion also includes observations on the present work and prospects for future work.

Chapter 2

Acoustic wave theory

The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.

Paul Dirac

Proceedings of the Royal Society A, Vol. 123, No. 792

2.1 Introduction

In order to derive the SRME process, a certain amount of theory must be developed. This is the purpose of this chapter, to establish the necessary results needed to derive SRME in a two- and three-dimensional setting, though it can be applied in any n-dimensional space.

Acoustic waves are generated by sources that excite the medium causing a variation in pressure and particle velocity in its vicinities. The well known acoustic wave equation is a partial differential equation that describes how the pressure evolves in time. Letting this acoustic pressure be denoted $p(\mathbf{x}, t)$ for each spacial coordinate $\mathbf{x} \in \mathbb{R}^3$ and each time $t \in \mathbb{R}$, the equation is given by

$$\Delta p(\mathbf{x}, t) - \frac{1}{c(\mathbf{x})^2} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) = f(\mathbf{x}, t) \tag{2.1}$$

The term f is the "source term" because it is a source of acoustic variation independent of the natural propagation of the waves. Inside of a "source-free region", the term is set

to zero. In the next section, we will establish how a point in such a region is affected by a previously existent wave field. It is paramount to the SRME algorithm as will be seen subsequently.

2.2 The second Rayleigh integral

Let $\mathbf{x}_A$ be a point inside a source-free region $V \subseteq \mathbb{R}^3$. Following Berkhout (1985) we will establish certain results on how $p(\mathbf{x}_A, t)$ can be described if $p(\mathbf{x}, t)$ is known elsewhere.

Inside V, p satisfies the wave equation 2.1. Therefore, the Fourier transformed* $P(\mathbf{x}, \omega) = \mathcal{F}[p(\mathbf{x}, t)](\omega)$ satisfies the Helmholtz equation, which is obtained Fourier transforming both sides of the wave equation:

$$\Delta P(\mathbf{x}, \omega) + k(\mathbf{x})^2 P(\mathbf{x}, \omega) = 0, \tag{2.2}$$

where and $k(\mathbf{x}) = \omega / c(\mathbf{x})$.

The first objective of this section will be to describe the pressure p inside V using only information of p on the boundary of S of V. To do that, we will make use of the Green's function of the wave equation. The Green's function g is the solution to the equation

$$\Delta g - \frac{1}{c(\mathbf{x})^2} \frac{\partial^2 g}{\partial t^2} = \delta(\mathbf{x}_A - \mathbf{x})\delta(t). \tag{2.3}$$

It is well known (Bleistein, 1984) that if $c(\mathbf{x})$ is a constant c, in three dimensions, g is given by

$$g(\mathbf{x}, t) = \frac{\delta\left(t - \frac{\|\mathbf{x}_A - \mathbf{x}\|}{c}\right)}{4\pi\|\mathbf{x}_A - \mathbf{x}\|}, \tag{2.4}$$

or, in the frequency domain,

$$G(\mathbf{x}, \omega) = \frac{e^{-i\omega\frac{\|\mathbf{x}_A - \mathbf{x}\|}{c}}}{4\pi\|\mathbf{x}_A - \mathbf{x}\|}. \tag{2.5}$$

Now, Green's second identity states that if u and v are twice continuously differentiable inside a compact region $E \subset \mathbb{R}^3$ which has a piecewise smooth boundary ∂E, then

*The Fourier transform used here will be $\mathcal{F}[f(t)](\omega) = \int_{\mathbb{R}} f(t) \exp(-it\omega)\, \mathrm{d}t$.

the following holds.

$$\int_E [u\Delta v - v\Delta u]\, dV = \oint_{\partial E} [u\nabla v - v\nabla u] \cdot \mathbf{n}\, dS, \tag{2.6}$$

where the unit normal vector $\mathbf{n}$ points outward. Let us suppose that V is compact, its boundary S is piecewise smooth and that P is twice continuously differentiable. Taking $u = P$ and $v = G$, Equation 2.6 becomes

$$\int_V [P\Delta G - G\Delta P]\, dV = \oint_S [P\nabla G - G\nabla P] \cdot \mathbf{n}\, dS. \tag{2.7}$$

By construction $\Delta G = \delta(\mathbf{x}_A - \mathbf{x}) - k^2 G$ and $\Delta P = -k^2 P$. Replacing these values in Equation 2.7, we obtain

$$\begin{aligned}\oint_S [P\nabla G - G\nabla P] \cdot \mathbf{n}\, dS &= \int_V [P(\delta(\mathbf{x}_A - \mathbf{x}) - k^2 G) + k^2 GP)]\, dV \\ &= \int_V P\delta(\mathbf{x}_A - \mathbf{x})\, dV.\end{aligned}$$

Finally, using the sift property of the delta function,

$$P(\mathbf{x}_A, \omega) = \oint_S [P\nabla G - G\nabla P] \cdot \mathbf{n}\, dS. \tag{2.8}$$

So far we can describe the wave field at a point $\mathbf{x}_A$ from the information of that wave field over some boundary S. We would like for that boundary to be a simple one, namely the plane $z = 0$. However, there is no compact region with a boundary such as that. A way around this issue is to take a succession of compact regions whose limit is $z = 0$.

Therefore, let us suppose that

$$V = \{\mathbf{x} \in \mathbb{R}^3 \,|\, z > 0, \|\mathbf{x} - \mathbf{x}_A\| < R\}. \tag{2.9}$$

and that V is source free for all $R \in \mathbb{R}$. Without loss of generality, $\mathbf{x}_A \in V$. It is clear that the boundary S of V is

$$S = \underbrace{\{\mathbf{x} \in \mathbb{R}^3 \,|\, z = 0, \|\mathbf{x} - \mathbf{x}_A\| < R\}}_{S_1} \cup \underbrace{\{\mathbf{x} \in \mathbb{R}^3 \,|\, z > 0, \|\mathbf{x} - \mathbf{x}_A\| = R\}}_{S_2}. \tag{2.10}$$

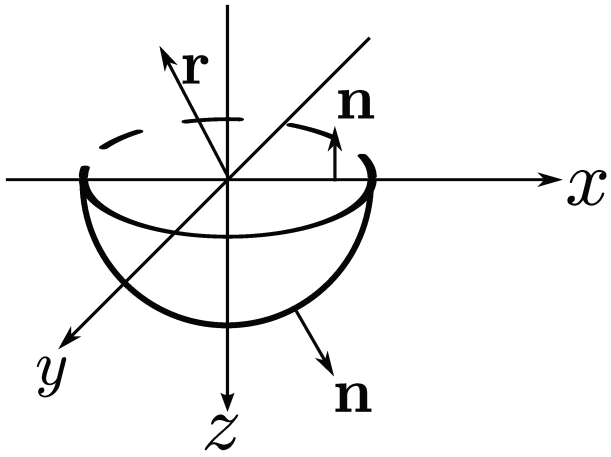


Figure 2.1: Volume V.

S_1 is the disk of radius R centered at $\mathbf{x}_A$ on the plane $z = 0$ and S_2 is the lower hemisphere of the sphere centered at $\mathbf{x}_A$ of radius R. Thusly, Equation 2.8 becomes

$$P(\mathbf{x}_A, \omega) = \int_{S_1} [P\nabla G - G\nabla P] \cdot \mathbf{n}\, dS_1 + \int_{S_2} [P\nabla G - G\nabla P] \cdot \mathbf{n}\, dS_2. \tag{2.11}$$

Ideally, we'd like the value of the second integral to tend to 0 as R increases unboundedly. This is not possible without additional hypotheses. Let $\mathbf{r}$ be the unit vector in the direction of $\mathbf{x}_A - \mathbf{x}$. A function $U(\mathbf{x}, \omega)$ satisfies the (displaced) Sommerfeld radiation conditions if

$$\|\mathbf{r}\| U(\mathbf{x}, \omega) \text{ is bounded for all } \mathbf{x}, \tag{2.12}$$

and

$$\lim_{\|\mathbf{r}\| \to \infty} \|\mathbf{r}\| \left[\frac{\partial U(\mathbf{x}, \omega)}{\partial \mathbf{r}} - ikU(\mathbf{x}, \omega) \right] = 0. \tag{2.13}$$

Physically, these conditions have the following interpretation:

> the sources must be *sources*, not *sinks*, of energy. The energy which is radiated from the sources must scatter to infinity; *no energy may be radiated from infinity into the prescribed singularities of the field* ... (Sommerfeld, 1949)

It is verifiable that G satisfies such conditions and given their physical interpretation, it is reasonable to suppose that P also does. We then establish that,

$$\int_{S_2} [P\nabla G - G\nabla P] \cdot \mathbf{n}\, dS_2 = \int_0^{2\pi} \int_0^{\pi/2} [P\nabla G - G\nabla P] \cdot (\mathbf{x}_A - \mathbf{x}) \|\mathbf{x}_A - \mathbf{x}\| \sin\theta \, \mathrm{d}\varphi \, \mathrm{d}\theta \tag{2.14}$$

Noting that on S_2, $\|\mathbf{x}_A - \mathbf{x}\| = R$,

$$\int_{S_2} [P\nabla G - G\nabla P] \cdot \mathbf{n}\, dS_2 = \int_0^{2\pi}\int_0^{\pi/2} [P\nabla G - G\nabla P] \cdot \mathbf{r}\, R^2 \sin\theta \,\mathrm{d}\varphi \,\mathrm{d}\theta$$
$$= \int_0^{2\pi}\int_0^{\pi/2} R^2 [P\partial_{\mathbf{r}} G - G\partial_{\mathbf{r}} P] \sin\theta \,\mathrm{d}\varphi \,\mathrm{d}\theta.$$

By simultaneously introducing and removing $R^2 ikGP \sin\theta$ to the integral one obtains

$$\int_{S_2} [P\nabla G - G\nabla P] \cdot \mathbf{n}\, dS_2 = \int_0^{2\pi}\int_0^{\pi/2} R^2 [P\partial_{\mathbf{r}} G - G\partial_{\mathbf{r}} P - ikGP + ikGP] \sin\theta \,\mathrm{d}\varphi \,\mathrm{d}\theta$$
$$= \int_0^{2\pi}\int_0^{\pi/2} \left[\underbrace{RP}_{\text{bounded}} \underbrace{R(\partial_{\mathbf{r}} G - ikG)}_{\to 0} - \underbrace{RG}_{\text{bounded}} \underbrace{R(\partial_{rB} P - ikP)}_{\to 0} \right] \sin\theta \,\mathrm{d}\varphi \,\mathrm{d}\theta.$$

Applying the Sommerfeld conditions, the above equation establishes that as $R \to \infty$, the integral over S_2 vanishes. As such, Equation 2.11 becomes

$$P(\mathbf{x}_A, \omega) = \int_{z=0} [P\nabla G - G\nabla P] \cdot \mathbf{n}\, dS. \tag{2.15}$$

Note that if G in the above equation was replaced with $G + H$, it remains unchanged as long as $\Delta H + k^2 H = 0$ and $G + H$ satisfies Sommerfeld conditions. Let us choose an H such that $G + H = 0$ on $z = 0$. It is not hard to verify that

$$H(\mathbf{x}', \omega) = -\frac{e^{-ik\|\mathbf{x}_A - \mathbf{x}'\|}}{4\pi\|\mathbf{x}_A - \mathbf{x}'\|}, \tag{2.16}$$

where $\mathbf{x}' = (x, y, -z)$ is such a function. Therefore, the second part of the integrand in Equation 2.15, now $(G + H)\nabla P \cdot \mathbf{n}$, is identically zero on $z = 0$ and the equation reduces to

$$P(\mathbf{x}_A, \omega) = \frac{1}{4\pi} \int_{z=0} P(\mathbf{x}, \omega) \nabla \left(\frac{e^{-ik\rho}}{\rho} - \frac{e^{-ik\rho'}}{\rho'} \right) \cdot \mathbf{n}\, \mathrm{d}S, \tag{2.17}$$

where $\rho = \|\mathbf{x}_A - \mathbf{x}\|$ and $\rho' = \|\mathbf{x}_A - \mathbf{x}'\|$.

The Rayleigh integral is obtained evaluating the above expression. A simple calcula-

tion yields

$$\nabla G \cdot \mathbf{n} = \frac{1 + ik\rho}{\rho} G \cos\phi, \tag{2.18}$$

where $\cos\phi = \mathbf{r} \cdot \mathbf{n}$ is the angle between $\mathbf{x}_A - \mathbf{x}$ and $\mathbf{n}$. Similarly for H,

$$\nabla H \cdot \mathbf{n} = \frac{1 + ik\rho'}{\rho'} H \cos\phi', \tag{2.19}$$

where $\cos\phi' = \mathbf{r}' \cdot \mathbf{n}'$ is the angle between $\mathbf{x}_A - \mathbf{x}'$ and $\mathbf{n}'$.

On $z = 0$, certain simplifications can be made. First, $\mathbf{x} = \mathbf{x}'$, so $\mathbf{r} = \mathbf{r}'$, $\rho = \rho'$ and $G = -H$. Also, $\mathbf{n} = -\mathbf{n}'$, and thus $\cos\phi = -\cos\phi'$. As consequence, $\partial_{\mathbf{n}} G = \partial_{\mathbf{n}} H$ and Equation 2.17 becomes

$$P(\mathbf{x}_A, \omega) = \frac{1}{2\pi} \int_{z=0} P(\mathbf{x}, \omega) \left(\frac{1 + ik\rho}{\rho^2} e^{-ik\rho} \cos\phi \right) \mathrm{d}S. \tag{2.20}$$

Equation 2.20 is known and the Rayleigh integral of the second kind.

2.3 Matrix formulation

By letting

$$W(\mathbf{x}_A - \mathbf{x}, \omega) = \frac{1 + ik\rho}{2\pi\rho^2} e^{-ik\rho} \cos\phi, \tag{2.21}$$

Equation 2.20 can be written as

$$P(\mathbf{x}_A, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(x_A - x, y_A - y, z_A, \omega) P(x, y, 0, \omega)\, \mathrm{d}x\, \mathrm{d}y. \tag{2.22}$$

This equation is a convolution in x and y between $W(\mathbf{x}, \omega)$ and $P(\mathbf{x}, \omega)$:

$$P(x, y, z, \omega) = W(x, y, z, \omega) \underset{x,y}{*} P(x, y, 0, \omega). \tag{2.23}$$

If $P(\mathbf{x}, \omega)$ does not vary laterally, that is, it does not depend on y, Equation 2.20 becomes

$$P(\mathbf{x}_A, \omega) = -\frac{ik}{2} \int_{z=0} P(x, z, \omega) H_1^{(2)}(k\rho)\, \mathrm{d}x, \tag{2.24}$$

where $H_1^{(2)}$ is the first order Hankel function of the second kind (Berkhout, 1985). In

this case, which is the one to be considered henceforth, we will take W to be given by $W(x_A - x, z_A - z, \omega) = -ikH_1^{(2)}(k\rho)/2$.

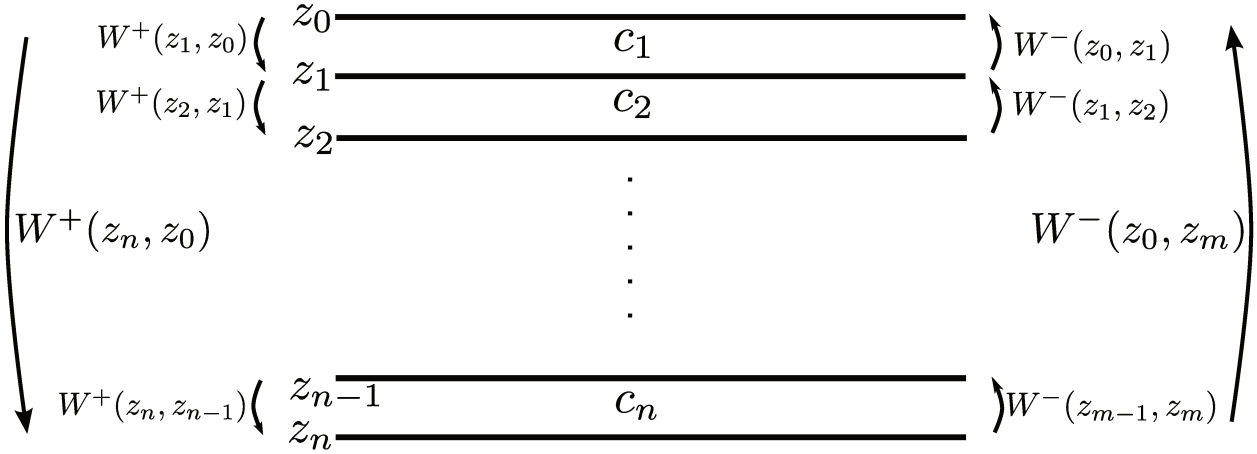


Figure 2.2: Horizontally sliced subsurface.

Now, we will change the notation slightly, in order for us to transition into a discrete setting. Let us calculate the field at a generic point at the depth z_{i-1}, supposing the wave field is known at depth level z_i. Therefore,

$$P(x, z_i, \omega) = (W * P)(x, z_{i-1}, \omega) \tag{2.25}$$

Consider the situation depicted in Figure 2.2. The idea behind such a scheme is to thinly slice the earth in such a way that each slice has a velocity that is independent of z. It is important to note that the slices should not be thought of as coinciding with potential reflectors. The notation which will be used in this work is based off of that in Berkhout (1985); Verschuur (1991). The notation is essentially discrete. Therefore, suppose first that $P(x, z_{i-1}, \omega)$ has compact support[†] and let us partition that support into N intervals of equal Δx spacing. Under those assumptions, the discrete form of Equation 2.24 is given by

$$P(x_i, z_j, \omega) = \sum_{\ell=0}^{N} W(x_i - x_\ell, z_j - z_{j-1}, \omega) P(x_\ell, z_{j-1}, \omega) \Delta x, \tag{2.26}$$

where now the wave field that affects the point $P(x_i, z_j)$ is measured on the plane $z = z_{j-1}$. The previous equation is susceptible to being written in matrix form. For that, let $(\mathbf{W}^+)_{i\ell}(z_j, z_{j-1}) = W(x_\ell - x_j, z_j - z_{j-1}, \omega)\Delta x$ and $(P^+)_\ell(z_{j-1}) = P(x_\ell, z_{j-1}, \omega)$ be a

[†]This is not unreasonable, since in practice, the pressure field is band limited.

matrix and a column vector, respectively. Equation 2.26 can be written as

$$P^+(z_j) = \mathbf{W}^+(z_j, z_{j-1})P^+(z_{j-1}). \tag{2.27}$$

There are two observations to be made about the previous equation. Firstly, $\mathbf{W}^+(z_j, z_{j-1})$ can both be seen as the *propagator matrix* from level z_{j-1} to level z_j. It can also be seen as the response to the spacial-temporal impulse. To see that, set $P(z_{j-1}) = [0, \dots, 0, 1, 0, \dots, 0]^T$. It is clear that $P(z_j) = (\mathbf{W}^+)_i(z_j, z_{j-1})$. Therefore, the i-th column of $\mathbf{W}^+(z_j, z_{j-1})$ is the response of an impulsive source at (x_i, z_{j-1}) measured along the level z_i.

Our model of wave propagation is almost concluded. To finalize it, we must add reflection and transmission through interfaces, since the matrices $\mathbf{W}^+$ only describe propagation through homogenous media. It can be shown (Berkhout, 1985; Verschuur, 1991) that they can both be described by matrix operations, just as propagation was. Transmission across the interface z_j can be written as

$$P^+(z_j^+) = \mathbf{T}^+(z_j)P^+(z_j^-), \tag{2.28}$$

where the superscripts $-$ and $+$ on z_j refer to the quantities before and after transmission, respectively. Now that reflection will be included in the model, we refer to P^+ as the downgoing wave field (which is the one we have been discussing so far) and P^- as the upgoing wave field, which will be presented shortly. Again in Berkhout (1985); Verschuur (1991) it is shown that

$$P^-(z_j) = \mathbf{R}(z_j)P^+(z_j), \tag{2.29}$$

where $\mathbf{R}(z_j)$ describes how the wave field is reflected on the level z_j. Each column of the matrix describes the upward-traveling impulse response of an impulsive source at (x_n, z_j) as described in de Bruin et al. (1990). To end the chapter, let us introduce a few more definitions: $\mathbf{W}^-(z_{j-1}, z_j) = \left[\mathbf{W}^+(z_j, z_{j-1})\right]^T$,

$$\mathbf{W}^+(z_n, z_0) = \mathbf{W}^+(z_n, z_{n-1})\mathbf{T}(z_{n-1})\mathbf{W}^+(z_{n-1}, z_{n-2})\mathbf{T}(z_{n-2}) \cdots \mathbf{T}(z_1)\mathbf{W}^+(z_1, z_0), \tag{2.30}$$

and

$$\mathbf{W}^-(z_0, z_n) = \mathbf{W}^+(z_n, z_0). \tag{2.31}$$

Finally, note that $P^+(z_j)$ and $P^-(z_j)$ are vectors and correspond to one shot. Let $P_n^+(z_0)$ correspond to a shot with source signature $S_n^+(z_0)$ that was placed at position x_n on the z_0 plane. The matrix $\mathbf{P}^+(z_0)$ is defined by having its n-th column be $P_n^+(z_0)$. The matrices $\mathbf{S}^+(z_0)$ and $\mathbf{P}^-(z_0)$ are defined similarly.

Chapter 3

Surface-Related Multiple Elimination

It was the future reflected
It felt familiar but new

MGMT
Future Reflections

3.1 Introduction to multiples

Multiple is the name given to an event in a seismic section resulting from waves that reflected multiple times in the subsurface (Sheriff and Geldart, 1995). In this sense, multiples oppose *primaries*, which incur only one reflection in the subsurface. It is wise to note that a seismic trace does not record only multiples and primaries, but all other sort of events, such as direct waves, refracted waves and so on; the seismic section contains all wave phenomena along with other sorts of noise. Multiples are nonetheless quite important in the analysis of the data, since they tend to have comparable amplitude to those of some primaries, and they are not always temporally separated from the data as much as direct waves.

It is common to depict wave paths as rays. Rays arise from applying the method of characteristics to the wave equation, and are perpendicular to the wavefront (see Bleistein, 1984; Červený, 2001). They not only provide a solid mathematical treatment of wave propagation, but they are also a powerful visualization tool.

Figure 3.1 below shows the ray paths of a primary reflection and a multiple reflection (solid blue and dashed red, respectively). They both result from a source placed at $\mathcal{S}$ and

are recorded at a receiver in $\mathcal{R}$. Since the path the multiple travels is longer, it will arrive at a later time in the receiver.

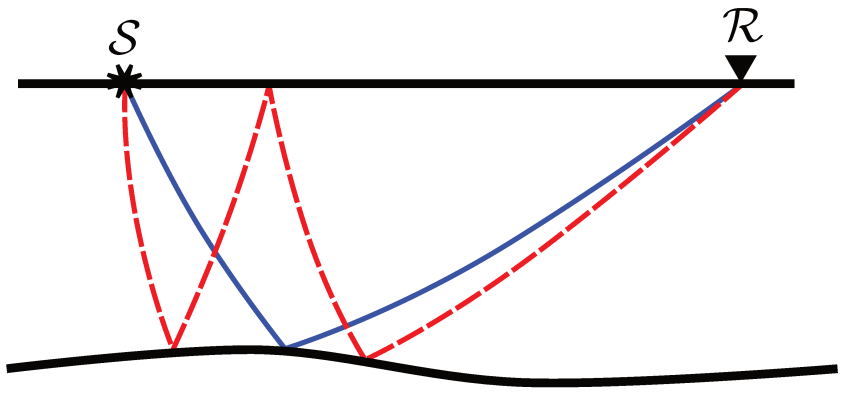


Figure 3.1: Ray paths of a primary reflection (blue) and a multiple reflection (red).

To illustrate even better the behavior of multiples, Figure 3.2 shows a common shot panel. The source is placed at 5 km and the receivers are spaced 10 m from each other, from 3 km to 7 km. The model is of a single horizontal reflector in a constant velocity medium.

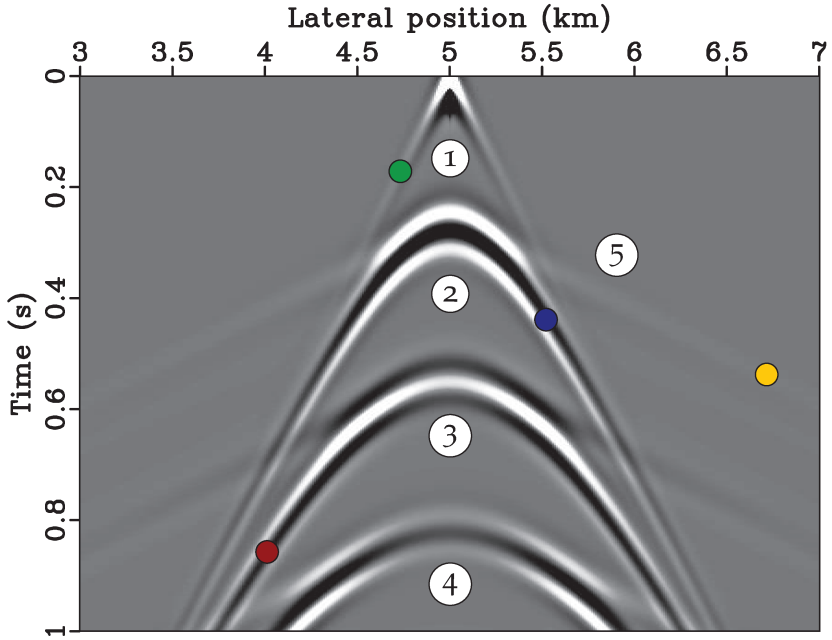


Figure 3.2: Common shot panel showing the direct wave (1), the primary wave (2), the first and second order multiples, (3) and (4) respectively, and the primary refraction (5).

Each colored dot in Figure 3.2 is an arrival of a certain feature of the wave. The green one for example, is an arrival of the direct wave. These different components of the wave field can be depicted, as we have before, as rays. The corresponding rays to these arrivals are shown in Figure 3.3. Thus, the green dot in Figure 3.2 at 4.75 km and 0.2 s means that, in Figure 3.3, the green ray took 0.2 s to travel from the source position (5 km) to the receiver at 4.75 km.

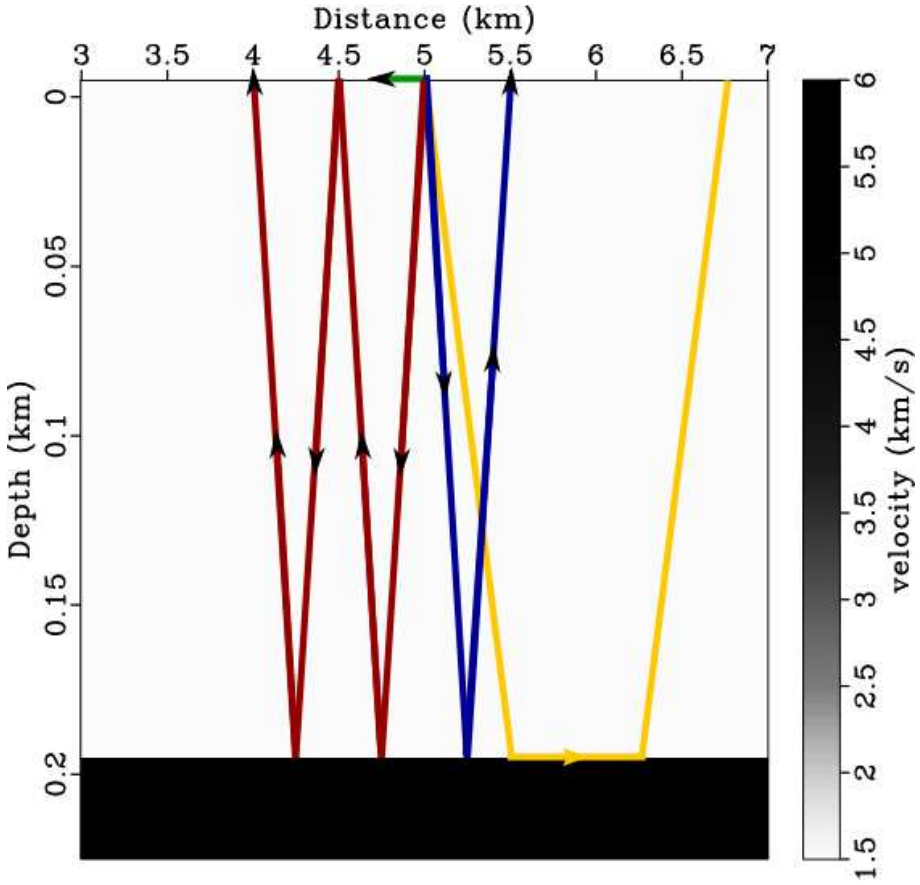


Figure 3.3: Ray paths of the different events of Figure 3.2.

It is important to observe that the multiples, though similar to the primaries in shape, are different. They have a different moveout behavior and thus cannot simply be treated as the primary shifted in time. However, it is important to note that for some types of multiples this treatment can be effective (see Robinson and Treitel, 2009, chap. 11). In this work, we shall exploit another method for predicting multiples, SRME.

SRME is a tool for predicting and eliminating surface multiples. Surface multiples are multiples that have at least one reflection in the acquisition surface, in contrast with internal multiples, which reverberate only in the internal layers of the earth. In the marine

case, surface multiples are very important. Since the acquisition surface is a free surface with normal reflection coefficient close to -1,[*] its multiples tend to appear prominently in the seismic gathers.

3.2 1D SRME

Before the 2D case is tackled, it is interesting to establish exactly how SRME works in one dimension. A detailed overview of the technique is given by Dragoset et al. (2010), and it includes the one dimensional case. A similar treatment is given by Verschuur (2006). As was stated in the introduction, SRME fits in the predict-subtract schemes of multiple removal techniques. The prediction part will be addressed first.

The convolutional model of the earth states that if a receiver is placed at the same position as the source, the source has a signature of $s(t)$ and the impulse response of the earth is $r(t)$, then the trace recorded will be given by $p(t) = (s * r)(t)$. The trace $p(t)$ contains the primaries and the multiples of all orders. First the case with a Dirac delta source will be considered.

3.2.1 Ideally impulsive source

Let $x(t)$ denote the trace measured with a Dirac delta source, that is, $x(t) = (\delta * r)(t)$. For simplicity the convolution will be denoted pointwise, that is, $x(t) = \delta(t) * r(t)$. Also, denote the part of $x(t)$ which only contains the primaries as $x_0(t)$. In this scenario, it is possible to write the first order multiples, that is, the multiples that reflected only once in the surface as a function of $x_0(t)$.

The multiple that reflected only once in the surface can be seen as two separate events. The first event is the wave path up until it reflects off the surface. The second event is the wave path after it reflects off the surface. The first event is clearly the primary. The second event can be seen as the response of the earth to reflected part of the first event. Since we are trying to model only the multiple, we suppose we only measure the second event. Therefore, the first order multiple can be written as

$$m_1(t) = \underbrace{r_0 x_0(t)}_{\text{Reflected primary}} * \underbrace{x_0(t)}_{\text{Reponse to primary}}. \tag{3.1}$$

[*]Considering the $v_a = 343\,\mathrm{m\,s^{-1}}$, $v_w = 1560\,\mathrm{m\,s^{-1}}$, $\rho_a = 1.2\,\mathrm{kg\,m^{-3}}$ and $\rho_w = 1020\,\mathrm{kg\,m^{-3}}$, a simple calculation yields $R_\perp = (v_a\rho_a - v_w\rho_w)/(v_a\rho_a + v_w\rho_w) \approx -0.99948$.

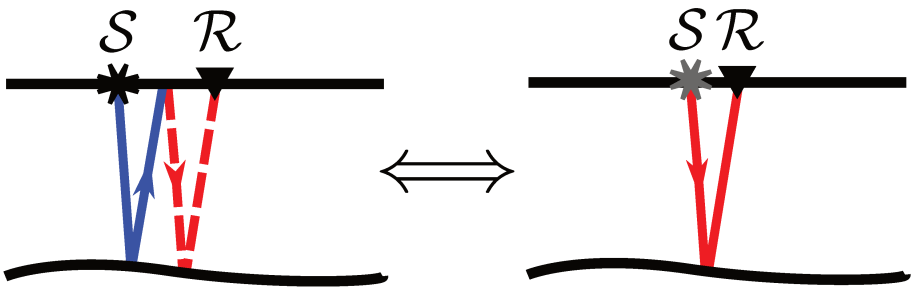


Figure 3.4: Display of the ray paths of a first order multiple. For visualization purposes, the source and receiver were separated.

In the convolutional model, the reflected primary acts as the source and the response to the primary acts as the reflectivity. This is possible because we are assuming a $\delta(t)$ source for the events.

Figure 3.4 is enlightening. It displays the ray path of a multiple. On the left side, the first event is represented by the solid blue ray while the second is represented by the dashed red ray. On the right side, the multiple, now a solid red line, can be seen as the earth's response to a primary whose source, represented in gray, is $r_0 x_0(t)$. That is, it can be seen as a primary resulting from a virtual source placed where the ray bounced off the surface. In order to better visualize the scenario, source and receiver were separated, but they are hypothetically in the same place. Therefore one should bear in mind that the blue and red paths are the same.

For a second order multiple, the rationale is similar. However, the source is now $r_0 m_1(t)$, therefore $m_2(t) = r_0 m_1(t) * x_0(t)$, or

$$m_2(t) = r_0^2 x_0(t) * x_0(t) * x_0(t). \tag{3.2}$$

In a general form, the nth order multiple can be written as

$$m_n(t) = r_0^n \underbrace{x_0(t) * \cdots * x_0(t)}_{n+1 \text{ times}}. \tag{3.3}$$

The full trace, $x(t)$ is the sum of the primaries and the multiples of all order, or

$$x(t) = \sum_{n \geq 0} r_0^n \underbrace{x_0(t) * \cdots * x_0(t)}_{n+1 \text{ times}}. \tag{3.4}$$

As it stands, the equation is not immediately useful; having the primaries in function of the trace is more appealing than the converse. A quick detour in the Fourier domain will

produce exactly that.

$$\begin{aligned}\mathcal{F}[x(t)](\omega) &= \mathcal{F}\left[\sum_{n\geq 0} r_0^n \underbrace{x_0(t) * \cdots * x_0(t)}_{n+1 \text{ times}}\right](\omega) \\ &= \sum_{n\geq 0} r_0^n \mathcal{F}[x_0(t) * \cdots * x_0(t)](\omega) \\ &= \sum_{n\geq 0} r_0^n \mathcal{F}[x_0(t)](\omega)^{n+1} \\ &= \mathcal{F}[x_0(t)](\omega) \sum_{n\geq 0} r_0^n \mathcal{F}[x_0(t)](\omega)^n \end{aligned} \tag{3.5}$$

From now on, $\mathcal{F}[f(t)](\omega)$ will be denoted simply as $F(\omega)$. In this notation, one can write Equation 3.5 as

$$X(\omega) = X_0(\omega) \sum_{n\geq 0} r_0^n X_0(\omega)^n. \tag{3.6}$$

Note that the sum above is the Laurent series of the function $1/(1 - r_0 X_0(\omega))$, which converges for $|X_0(\omega)| < 1$. In practice, one does not have $X_0(\omega)$. However, one does have $X(\omega)$, which can be normalized to have unit L_2 norm.[†] This will guarantee that $|X(\omega)| < 1$, which in consequence will guarantee that $|X_0(\omega)| < 1$. Hence,

$$X(\omega) = \frac{X_0(\omega)}{1 - r_0 X_0(\omega)}. \tag{3.7}$$

This equation can be manipulated to yield

$$X_0(\omega) = \frac{X(\omega)}{1 + r_0 X(\omega)} \tag{3.8}$$

which is remarkably similar to Equation 3.7. Again, with $|X(\omega)| < 1$, one can write the previous equation as

$$X_0(\omega) = X(\omega) \sum_{n\geq 0} (-r_0)^n X(\omega)^n. \tag{3.9}$$

This is the first main result of SRME. It provides the primaries in terms of full data record, containing all order multiples. However, it fails to be useful in practice for two reasons. First, r_0 needs to be known. This is not a major drawback in the marine case,

[†]By Parseval's Theorem, $\|x(t)\|_2 = \|X(\omega)\|_2$, therefore, normalizing to unit L_2 norm will force the signal to have $|X(\omega)| < 1$ for all ω.

since r_0 can be approximated to -1. In the land case, this is dealt with in a different manner. The second shortcoming is that the equation assumes a Dirac delta source. This shall be addressed henceforth.

3.2.2 Arbitrary source

In order to include the source signature $s(t)$, first note that the trace $p(t)$ resulting from such source can be written as

$$\begin{aligned} p(t) &= s(t) * r(t) \\ &= s(t) * \delta(t) * r(t) \\ &= s(t) * x(t). \end{aligned} \tag{3.10}$$

Equivalently, the multiple-free $p_0(t)$ response with $s(t)$ source is given by

$$p_0(t) = s(t) * x_0(t). \tag{3.11}$$

Recalling Equation 3.4, one may write it as

$$\begin{aligned} x(t) &= x_0(t) + r_0 x_0(t) * x_0(t) + r_0^2 x_0(t) * x_0(t) * x_0(t) + \dots \\ &= x_0(t) * [\delta(t) + r_0 x_0(t) + r_0^2 x_0(t) * x_0(t) + \dots] \\ &= x_0(t) * [\delta(t) + r_0 x(t)]. \end{aligned} \tag{3.12}$$

The equation above can be convolved with $s(t)$ to yield an expression for $p(t)$.

$$\begin{aligned} x(t) * s(t) &= x_0(t) * s(t) + r_0 x_0(t) * x(t) * s(t) \\ p(t) &= p_0(t) + r_0 x_0(t) * s(t) \end{aligned} \tag{3.13}$$

Let $a(t)$ be a function such that $a(t) * s(t) = r_0 \delta(t)$, that is, it acts as a deconvolution for the source $s(t)$. In this case, Equation 3.11 can be written in function of $a(t)$:

$$p_0(t) * a(t) = r_0 x_0(t) * a(t),$$

which can be incorporated into Equation 3.13 producing the following result

$$p(t) = p_0(t) + p_0(t) * a(t) * p(t). \tag{3.14}$$

In the Fourier domain, the equation becomes

$$P(\omega) = P_0(\omega) + P_0(\omega)A(\omega)P(\omega).$$

Solving for $P_0(\omega)$ one obtains

$$P_0(\omega) = \frac{P(\omega)}{1 + A(\omega)P(\omega)}, \tag{3.15}$$

which, expanded as a series in the time domain, becomes

$$p_0(t) = p(t) - a(t) * p(t) * p(t) + a(t) * a(t) * p(t) * p(t) * - \dots . \tag{3.16}$$

This result is the essence of 1D SRME. It establishes that the primary can be obtained by adding successive autoconvolutions of the trace to itself, as long as these autoconvolutions are properly deconvolved with respect to the source. The process can be proposed in an iterative manner.

Algorithm 3.1 Iterative 1D SRME

Choose $p_0^{(0)}(t) = p(t)$.
for $0 \leq i < \infty$ **do**
Calculate $a(t)$ such that it minimizes $\|p_0^{(i)} - a(t) * p_0^{(i)}(t) * p(t)\|$

$$p_0^{(i+1)} \leftarrow p_0^{(i)} - a(t) * p_0^{(i)}(t) * p(t)$$

end for

In practice, calculating $a(t)$ is complicated, since it involves knowledge of the source wavelet and the reflectivity coefficient. The usual way to surpass this obstacle, is by choosing $a(t)$ to be of a finite length N_f and finding the coefficients that lead to the smallest L_2 norm of $p_0^{(i)} - a(t) * p_0^{(i)}(t) * p(t)$. The reasoning behind this, is that when the correct $a(t)$ is chosen, Algorithm 3.1 will generate the series present in Equation 3.16. Therefore the multiples will be correctly removed, causing the L_2 norm to decrease. As such, the criteria for finding the correct $a(t)$ is chosen to be exactly that one which decreases the L_2 norm. When this process is done at each iteration, it is called adaptive SRME. It is a process with its own pitfalls, and the purpose of this work is to overcome some of them.

3.3 2D SRME

In the 2D case the power of the matrix notation and overall theory developed in Chapter 2 will become evident. Let $\mathbf{P}^-_{\text{tot}}(z_0)$ denote the total upgoing wave field measured (in the frequency domain) at the level z_0. It is composed of the response of the propagated source signature $\mathbf{S}^+(z_0)$, but also, in a recursive manner, the reflected upgoing wave field. This can be seen with the aid of the following diagram.

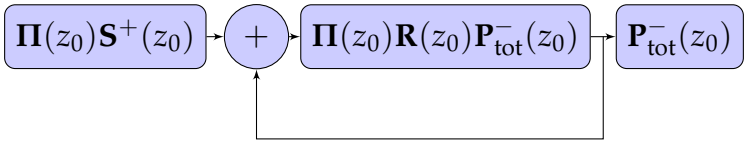


Figure 3.5: Block diagram representing Equation 3.18.

Following the introduction of the source energy $\mathbf{S}^+(z_0)$, it is propagated through the interfaces, reflecting off of each them as it travels through the subsurface. The operator $\mathbf{\Pi}(z_0)$, given by the following equation, describes this propagation.

$$\mathbf{\Pi}(z_0) = \sum_{m=1}^{n} \mathbf{W}^-(z_0, z_m)\mathbf{R}(z_m)\mathbf{W}^+(z_m, z_0) \tag{3.17}$$

After the source has been fully propagated downwards and upwards, the total wave field (at that point) reaches the surface. It then reflects off of it; the reflection is obtained multiplying it by $\mathbf{R}(z_0)$. This wave field is then propagated with $\mathbf{\Pi}(z_0)$. Once again it reaches the surface and is reflected off of it. This recursion, described by Berkhout (1985), can be expressed in the following manner.

$$\mathbf{P}^-_{\text{tot}}(z_0) = \mathbf{\Pi}(z_0)\left[\mathbf{S}^+(z_0) + \mathbf{R}(z_0)\mathbf{P}^-_{\text{tot}}(z_0)\right] \tag{3.18}$$

Remember that this is a multi-record notation, and each column of the matrices $\mathbf{P}^+_{\text{tot}}$ and $\mathbf{S}^+$ correspond to a different shot. Assuming a constant reflection coefficient of r_0 as was done in the previous section, the reflection matrix becomes $\mathbf{R}(z_0) = r_0\mathbf{I}$ and Equation 3.18 becomes

$$\mathbf{P}^-_{\text{tot}}(z_0) = \mathbf{\Pi}(z_0)\left[\mathbf{S}^+(z_0) + r_0\mathbf{P}^-_{\text{tot}}(z_0)\right]. \tag{3.19}$$

The primary reflections are obtained by only propagating the source signatures once, that is $\mathbf{P}^-(z_0) = \mathbf{\Pi}(z_0)\mathbf{S}^+(z_0)$. This term includes internal multiples, as SRME fails to

produce them in the recursive manner given above. Therefore we have

$$\mathbf{P}^{-}_{\text{tot}}(z_0) = \mathbf{P}^{-}(z_0) + r_0\mathbf{\Pi}(z_0)\mathbf{P}^{-}_{\text{tot}}(z_0). \tag{3.20}$$

Now, let us define a deconvolution matrix $\mathbf{A}$ such that $\mathbf{S}^{+}(z_0)\mathbf{A} = r_0\mathbf{I}$. If such matrix exists, $\mathbf{P}^{-}(z_0)\mathbf{A} = r_0\mathbf{\Pi}(z_0)$ and Equation 3.20 can be written as

$$\mathbf{P}^{-}_{\text{tot}}(z_0) = \mathbf{P}^{-}(z_0) + \mathbf{P}^{-}(z_0)\mathbf{A}\mathbf{P}^{-}_{\text{tot}}(z_0), \tag{3.21}$$

which can be solved for $\mathbf{P}^{-}(z_0)$ yielding

$$\mathbf{P}^{-}(z_0) = \mathbf{P}^{-}_{\text{tot}}(z_0)\left[\mathbf{I} + \mathbf{A}\mathbf{P}^{-}_{\text{tot}}(z_0)\right]^{-1}. \tag{3.22}$$

Algorithm 3.2 Iterative 2D SRME

Choose $\mathbf{P}^{(0)} = \mathbf{P}_{\text{tot}}$.
for $0 \leq i < \infty$ **do**
Calculate $\mathbf{A}$ such that it minimizes $\|\mathbf{P}^{(i)} - \mathbf{P}^{(i)}\mathbf{A}\mathbf{P}_{\text{tot}}\|$

$$\mathbf{P}^{(i+1)} \leftarrow \mathbf{P}^{(i)} - \mathbf{P}^{(i)}\mathbf{A}\mathbf{P}_{\text{tot}}$$

end for

The previous equation constitutes the main result of SRME. The primary reflection can be written as a function of the total data $\mathbf{P}^{-}_{\text{tot}}(z_0)$ and the deconvolution operator $\mathbf{A}$. These quantities are all related to the surface z_0, and they involve no knowledge of the subsurface. However, the operator $\mathbf{A}$ is not known *a priori*. Nonetheless, as long as a suitable inverse for $\mathbf{S}^{+}(z_0)$ can be found and if $\mathbf{I} + \mathbf{A}\mathbf{P}^{-}_{\text{tot}}(z_0)$ has an inverse, $\mathbf{P}^{-}(z_0)$ can be obtained. Equation 3.22 stands as the two-dimensional version of Equation 3.15. From it, we can adapt the Algorithm 3.1 for the 2D case.

Algorithm 3.2 is based on the approach first shown in Verschuur et al. (1992) and detailed in Berkhout and Verschuur (1997). It is important to note that this algorithm can be extended to the three dimensional case (see Dragoset et al., 2010).

Chapter 4

Independent Component Analysis

Young man, in mathematics you don't understand things. You just get used to them.

John von Neumann

4.1 Introduction

4.1.1 Motivation

The method of independent component analysis was historically motivated by the desire to separate mixtures of independent signals (Herault and Jutten, 1987). That is, suppose that one has observed n random variables, $\{x_i\}_{i\in\{1,\ldots,n\}}$. Suppose also, that these random variables are linear combinations of statistically mutually independent random variables $\{s_i\}_{i\in\{1,\ldots,n\}}$, that is

$$\begin{aligned} x_1 &= a_{11}s_1 + \cdots + a_{1n}s_n \\ &\vdots \\ x_n &= a_{n1}s_1 + \cdots + a_{nn}s_n, \end{aligned} \tag{4.1}$$

which can be more succinctly written as:

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{4.2}$$

where $(\mathbf{x})_i = x_i$, $(\mathbf{A})_{ij} = a_{ij}$ and $(\mathbf{s})_i = s_i$, for $i \in \{1, \ldots, n\}$. While there exists results for rectangular matrices $\mathbf{A}$ (Comon, 1994; Taleb and Jutten, 1999), this work will concern itself with the square case. Without loss of generality, these random variables will be considered zero mean. Thus the problem ICC attempts to solve can be stated as follows.

Problem 1 (Separation). *Given a random vector* $\mathbf{x}$*, such that*

$$\mathbf{x} = \mathbf{A}\mathbf{s},$$

recover the random vector $\mathbf{s}$ *under the assumption that the components of* $\mathbf{s}$ *are statistically mutually independent.*

If the mixture matrix $\mathbf{A}$ was known and invertible, the problem could be easily solved by $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$. However, $\mathbf{A}$ is not known *a priori* — fact which will be remedied by assuming the mutual statistical independence of the components of $\mathbf{s}$.

The rest of this chapter will be concerned with developing the appropriate machinery to solve this problem efficiently. Some theorems regarding the "solvability" of the ICA problems (separability, identifiability and uniqueness) are well exposed and proved in Eriksson and Koivunen (2004), in which earlier work (Comon, 1994; Taleb and Jutten, 1999; Cao and Liu, 1996) is expanded. For our purposes, it suffices to know that Problem 1 has a solution under the mild assumptions that $\mathbf{A}$ is full column rank and at most one source s_j is Gaussian.

4.1.2 History

While the mathematical formulation of the problem as shown above was first given by Comon (1994), the method of ICA was known earlier. In particular, the works of Herault and Jutten were the first published results to blindly separate signals based on the criterion of independence. They worked on the problem in as early as 1983, culminating in their article Herault and Jutten (1987), where they expose an algorithm to perform ICA. The work was presented in 1986, and consisted in using an adaptive neural network to separate two mixtures of independent signals. It coined the term ICA, in connection with PCA (principal component analysis).

In 1987, identifiability, a key aspect of ICA, was developed and finally published in Giannakis et al. (1989). Throughout the late 1980s and early 1990s, efforts were made in optimizing the algorithm and establishing firmer theoretical grounds for it.

However, it was only in 1994 that Comon proposed a polynomial time algorithm for performing ICA, based on mutual information. In his aforementioned article, the

problem was formulated mathematically and criteria for its solution were established, extending previous results.

From then on, ICA garnered the attention of the signal processing community, and was established as a mature line of research. Currently, ICA is used in a myriad of different areas, from neuroscience, in the analysis of EEG (electroencephalography) signals, to geophysics, in the analysis of seismic signals.

4.2 ICA by maximizing nongaussianity

Problem 1 can be tackled in a number of different ways. As was described above, the first algorithm to perform ICA a neural network described in Herault and Jutten (1987). Subsequently, other approaches were developed, such as the infomax approach by Linkster (1992) that was based on optimizing mutual information. This optimization benefits from the use of other gradients which were first described independently by Cardoso (1996) and by Amari et al. (1997).

However, other approaches not based on information have surfaced. These include maximum likelihood algorithms (Gaeta and Lacoume, 1990; Pham et al., 1992), tensorial methods (Cardoso, 1989, 1990) (including the JADE algorithm first described in Cardoso and Souloumiac (1993)), and negentropy based methods. It is with this last class of methods that the following sections will be concerned with, following primarily Hyvärinen (1999) and Hyvärinen et al. (2001). They will be used because of their fast convergence, relatively simple theoretical basis, and attested competence.

4.2.1 Nongaussianity and the Central Limit Theorem

The idea behind using nongaussianity to maximize mutual independence is based on an interpretation of the Central Limit Theorem (CLT). While there are many CLTs, the following formulation is of interest.

Theorem 1 (Central Limit Theorem)**.** *Let $\{\eta_n\}_{n\in\mathbb{N}}$ be a sequence of independent random variables with finite variances. Define*

$$\zeta_n = \sum_{i=1}^{n} \eta_i \text{ and } \chi_n = \frac{\zeta_n - \mathrm{E}[\zeta_n]}{\mathrm{Var}[\zeta_n]}.$$

Then, if the Lindeberg condition[*] is satisfied, χ_n converges weakly[†] to a Gaussian distribution with zero mean, and unitary variance as $n \to \infty$.

In the application at hand, it is impossible to predict whether the random variables will satisfy the Lindeberg condition, therefore, the proof will be omitted. The reader can refer to Koralov and Sinai (2007) for more details.

The importance of the CLT lies on the fact that it states that sums of independent random variables tend to be "more Gaussian" than each random variable by itself. Therefore, an indirect measure of independence is nongaussianity, or how much a random variable deviates from being Gaussian. Important measures of nongaussianity include kurtosis and negentropy.

Before these measures are presented, it is important to expose how they will be used in practice. Recalling the ICA model, $\mathbf{x} = \mathbf{A}\mathbf{s}$. Assuming $\mathbf{A}$ is invertible, $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$, that is, s_i is given by a linear combination of x_i. Therefore, a linear combination of $\mathbf{s}$, $\mathbf{q}^T\mathbf{s}$, is a linear combination $\mathbf{b}^T\mathbf{x}$ of $\mathbf{x}$. Clearly $\mathbf{b}^T\mathbf{x} = \mathbf{b}^T\mathbf{A}\mathbf{s}$, and consequently, $\mathbf{q} = \mathbf{A}^T\mathbf{b}$. By the CLT, in general terms, the more nongaussian $\mathbf{q}^T\mathbf{s}$, the more independent that random variable will be. However, $\mathbf{q}^T\mathbf{s}$ will be the most independent when $\mathbf{q}^T\mathbf{s} \propto s_i$, since any linear combination of the s_i are more Gaussian than each s_i independently. Therefore, maximizing the nongaussianity of $\mathbf{q}^T\mathbf{s}$, one obtains a scaled version of one of the sources. The obtained random variable is called an independent component (IC). Also, since $\mathbf{q}^T\mathbf{s} = \mathbf{b}^T\mathbf{x}$, the right hand, which is known, can be used for the maximization of nongaussianity. The problem can be written as

$$\begin{aligned} &\max_{\mathbf{b}} \text{ nongaussianity of } \mathbf{b}^T\mathbf{x} \\ &\text{s. t. } \|\mathbf{q}\| = 1. \end{aligned}$$

[*] The Lindeberg condition is that

$$\lim_{n\to\infty} \frac{1}{\operatorname{Var}[\zeta_n]} \sum_{i=1}^{n} \int_{\left\{x\in\mathbb{R}:|x-\mathrm{E}[\eta_i]|\geq\varepsilon\sqrt{\operatorname{Var}[\zeta_n]}\right\}} (x-\mathrm{E}[\eta_i])^2 \,\mathrm{d}F_i(x) = 0$$

for every $\varepsilon > 0$. Intuitively, the Lindeberg condition states that $\dfrac{\eta_n - \mathrm{E}[\eta_n]}{\sqrt{\operatorname{Var}[\eta_n]}}$ become small as n increases.

[†] The sequence of probability measures P_n converges to P weakly if

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\,\mathrm{d}P_n(x) = \int_{\mathbb{R}} f(x)\,\mathrm{d}P(x)$$

for every continuous bounded function f.

The whole problem lies on accurately measuring nongaussianity of $\mathbf{b}^T\mathbf{x}$ and finding its global maxima, subject to the appropriate constraint.

4.2.2 Nongaussianity by kurtosis

Kurtosis is defined as the fourth cumulant of a random variable. For a zero mean random variable, it is given by[‡]

$$\text{kurt}[x] = \text{E}[x^4] - 3(\text{E}[x^2])^2.$$

The importance of kurtosis, lies in the fact that the kurtosis of a Gaussian random variable is zero. Since all random variables are assumed to be zero mean, this fact will stated and proved for a zero mean Gaussian variable, however it is still valid for any Gaussian variable.

Proposition 1. *Let x_g be a zero mean Gaussian variable such that* $\text{Var}[x_g] = \sigma^2$. *Then* $\text{kurt}[x_g] = 0$.

Proof. Let $\varphi(\omega)$ be the characteristic function of x_g, that is $\varphi(\omega) = \text{E}[\exp(i\omega x)]$. Then

$$\varphi(\omega) = \exp\left(-\frac{1}{2}\omega^2\sigma^2\right).$$

This stems directly from the fact that $\varphi(\omega)/2\pi$ is the inverse Fourier transform of $p_{x_g}(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\dfrac{-x^2}{2\sigma^2}\right)$. Now, let $M(t) = \varphi(-it)$ be the moment generating function of x_g. By definition, its Taylor series is given by

$$M(t) = \sum_{k=0}^{\infty} \text{E}[x^k]\frac{t^k}{k!}.$$

Therefore, in order to calculate the second and fourth moments of x_g one must find the second and fourth order terms of the Taylor expansion of $M(t)$. Since

$$M(t) = \exp\left(\frac{1}{2}t^2\sigma^2\right)$$

[‡]The general expression is (Hyvärinen et al., 2001, chap. 2)

$$\text{kurt}[x] = \text{E}[x^4] - 3(\text{E}[x^2])^2 - 4\,\text{E}[x^3]\,\text{E}[x] + 12\,\text{E}[x^2](\text{E}[x])^2 - 6(\text{E}[x])^4.$$

we have that

$$\begin{aligned} M(t) &= \sum_{k=0}^{\infty} \frac{(t^2\sigma^2/2)^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{\sigma^{2k}}{2^k} \frac{t^{2k}}{k!} \\ &= \sum_{k=0}^{\infty} \frac{\sigma^{2k} \cdot (2k)(2k-1)\cdots(k+1)}{2^k} \frac{t^{2k}}{(2k)!}. \end{aligned}$$

As was expected, $\mathrm{E}[x^2] = (\sigma^2 \cdot 2)/2 = \sigma^2$. Moreover, $\mathrm{E}[x^4] = (\sigma^4 \cdot 4 \cdot 3)/2^2 = 3\sigma^4$. Consequently, $\mathrm{kurt}[x_g] = 3\sigma^4 - 3(\sigma^2)^2 = 0$. □

In light of this, kurtosis is an attractive measure of nongaussianity. However, some caution must be had when using kurtosis for that purpose. Firstly, not all zero kurtosis random variables are Gaussian. These so called mesokurtic random variables do not appear much in practice, and can be safely ignored for the purposes of the applications at hand. Also, it is not wise to use kurtosis compare the nongaussianity of leptokurtic (positive kurtosis) random variables with platykurtic (negative kurtosis) random variables. This is because, while leptokurtic random variables can have arbitrarily large kurtosis, the kurtosis of platykurtic random variables are bounded by below.

Established the measure of nongaussianity, the problem can be stated as:

$$\begin{aligned} &\max_{\mathbf{b}} \ |\,\mathrm{kurt}[\mathbf{b}^T\mathbf{x}]| \\ &\text{s. t. } \|\mathbf{q}\| = 1. \end{aligned}$$

There is a problem, however. *Prima facie*, it is impossible to obtain $\|\mathbf{q}\|$ from $\mathbf{b}$, since $\mathbf{q} = \mathbf{A}^T\mathbf{b}$, and the mixing matrix $\mathbf{A}$ is unknown. Therefore, it is important to apply a whitening preprocessing to $\mathbf{x}$. Whitening $\mathbf{x}$ refers to applying a linear transformation $\mathbf{V}$ to $\mathbf{x}$, such that $\mathrm{E}[\mathbf{V}\mathbf{x}(\mathbf{V}\mathbf{x})^T] = \mathbf{I}$. Let $\mathbf{z} = \mathbf{V}\mathbf{x}$. This linear transformation can be given by $\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^T$, where $\mathrm{E}[\mathbf{x}\mathbf{x}^T] = \mathbf{E}\mathbf{D}\mathbf{E}^T$ is a spectral decomposition (Hyvärinen et al., 2001). Note that the ICA model and the ICs remain the same: $\mathbf{z} = (\mathbf{V}\mathbf{A})\mathbf{s}$, while the mixing matrix changes. Now, the function to be maximized is $|\,\mathrm{kurt}[\mathbf{w}^T\mathbf{z}]|$, subject to the same constraint as before. However, since $\mathbf{q} = \mathbf{A}^T\mathbf{b}$, and $\mathbf{b} = \mathbf{V}^T\mathbf{w}$,

$$\|\mathbf{q}\|^2 = (\mathbf{A}^T\mathbf{V}^T\mathbf{w})^T(\mathbf{A}^T\mathbf{V}^T\mathbf{w}) = \mathbf{w}^T\mathbf{V}\mathbf{A}\mathbf{A}^T\mathbf{V}^T\mathbf{w}. \tag{4.3}$$

Furthermore,

$$\begin{aligned}
\mathbf{I} &= \mathrm{E}[\mathbf{z}\mathbf{z}^T] \\
&= \mathrm{E}[(\mathbf{VAs})(\mathbf{VAs})^T], \text{since } \mathbf{z} = \mathbf{VAs}. \\
&= \mathrm{E}[\mathbf{VAss}^T\mathbf{A}^T\mathbf{V}^T] \\
&= \mathbf{VA}\,\mathrm{E}[\mathbf{ss}^T]\mathbf{A}^T\mathbf{V}^T, \text{since } \mathbf{s} \text{ has mutually independent components.} \\
&= \mathbf{VAA}^T\mathbf{V}^T. \qquad (4.4)
\end{aligned}$$

Substituting Equation 4.4 in Equation 4.3, one obtains $\|\mathbf{q}\| = \|\mathbf{w}\|$. The maximization problem becomes:

$$\begin{aligned}
&\max_{\mathbf{w}} \; |\operatorname{kurt}[\mathbf{w}^T\mathbf{z}]| \\
&\text{s. t. } \; \|\mathbf{w}\|^2 = 1.
\end{aligned} \qquad (4.5)$$

The whitening step is paramount to a practical solution via maximization of nongaussianity. Without it, it would be quite hard to algorithmically constrain $\mathbf{q}$ to lie on the unit sphere.

The first order necessary Karush–Kuhn–Tucker (KKT) conditions for this problem are the following (Nocedal and Wright, 2006, chap. 12). Let

$$\mathcal{L}(\mathbf{w}, \lambda) = -|\operatorname{kurt}[\mathbf{w}^T\mathbf{z}]| + \lambda(\|\mathbf{w}\|^2 - 1),$$

and $\mathbf{w}^*$ be a local minimum of the aforementioned maximization problem. Then there exists a λ^* such that

$$\nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w}^*, \lambda^*) = 0.$$

This is also known as the method of Lagrange multipliers, and is equivalent to

$$\left.\frac{\partial|\operatorname{kurt}[\mathbf{w}^T\mathbf{z}]|}{\partial\mathbf{w}}\right|_{\mathbf{w}=\mathbf{w}^*} = \lambda\left.\frac{\partial(\|\mathbf{w}\|^2 - 1)}{\partial\mathbf{w}}\right|_{\mathbf{w}=\mathbf{w}^*},$$

which yields

$$2\lambda\mathbf{w}^* = 4\operatorname{sign}(\operatorname{kurt}[\mathbf{w}^{*T}\mathbf{z}])(\mathrm{E}[\mathbf{z}(\mathbf{w}^{*T}\mathbf{z})^3] - 3\mathbf{w}^*\|\mathbf{w}^*\|^2),$$

or simply

$$\mathbf{w}^* = \tilde{\lambda}(\mathrm{E}[\mathbf{z}(\mathbf{w}^{*T}\mathbf{z})^3] - 3\mathbf{w}^*).$$

An iteration to solve this equation follows.

Algorithm 4.1 FastICA for kurtosis

Choose random vector $\mathbf{w}_0$.
while $\mathbf{w}_n^T \mathbf{w}_{n-1} \neq 1$ **do**

$$\mathbf{w}^+ \leftarrow \mathrm{E}[\mathbf{z}(\mathbf{w}_{n-1}^T \mathbf{z})^3] - 3\mathbf{w}_{n-1}$$

$$\mathbf{w}_n \leftarrow \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|}$$

end while

It is the basis for the FastICA algorithm, developed in Hyvärinen (1999), which shall be revisited in the following section.

4.2.3 Nongaussianity by negentropy

Another attractive measure of nongaussianity is negentropy, based on differential entropy. Differential entropy of a random variable x is defined as

$$h(x) = -\int_{\mathbb{R}} p_x(\xi) \log p_x(\xi)\, \mathrm{d}\xi. \tag{4.6}$$

It can be seen as a measure of randomness. For a more detailed discussion, see Cover and Thomas (2006). In the present discussion, it suffices to show an important property.

Proposition 2. *Let x_G be a zero mean Gaussian random variable. Then, for any zero mean random variable y with $\mathrm{Var}[y] = \mathrm{Var}[x_G] \equiv \sigma^2$, $h(x_G) \geq h(y)$. That is to say, a Gaussian random variable is the most entropic amongst random variables of the same variance.*

Proof. In order to prove this property, the Kullback-Leibler (KL) divergence will be introduced. Let p_u and p_v be the probability density functions (pdfs) of u and v, respectively. The KL divergence is defined as:

$$D_{\mathrm{KL}}(p_u, p_v) = \int_{\mathbb{R}} p_u(\xi) \log\left(\frac{p_u(\xi)}{p_v(\xi)}\right) \mathrm{d}\xi \tag{4.7}$$

for any p_u, p_v, $D_{\mathrm{KL}}(p_u, p_v) \geq 0$.

In fact, from Jensen's inequality (see Rudin, 1987, chap. 3), if φ is a strictly convex function, $\mathrm{E}[\varphi(z)] \geq \varphi(\mathrm{E}[z])$. Taking $\varphi(z) = -\log(z)$ and evaluating it at $z = p_v(\xi)/p_u(\xi)$,

it is true that

$$\begin{aligned} D_{\mathrm{KL}}(p_u, p_v) &= \mathrm{E}_u\left[\varphi\left(\frac{p_v(\xi)}{p_u(\xi)}\right)\right] \geq \\ &\geq \varphi\left(\mathrm{E}_u\left[\frac{p_v(\xi)}{p_u(\xi)}\right]\right) = \\ &= -\log\left(\int_{\mathbb{R}} p_u(\xi)\frac{p_v(\xi)}{p_u(\xi)}\,\mathrm{d}\xi\right) = \\ &= -\log\left(\int_{\mathbb{R}} p_v(\xi)\,\mathrm{d}\xi\right) = 0. \end{aligned}$$

Now, let x_G be the Gaussian random variable from the hypothesis, and y any zero mean random variable. Therefore

$$\begin{aligned} 0 \leq D_{\mathrm{KL}}(p_{x_G}, p_y) &= \\ &= \int_{\mathbb{R}} p_y(\xi)\log\left(\frac{p_y(\xi)}{p_{x_G}(\xi)}\right)\,\mathrm{d}\xi = \\ &= -h(y) - \int_{\mathbb{R}} p_y(\xi)\log(p_{x_G}(\xi))\,\mathrm{d}\xi. \end{aligned}$$

Substituting p_{x_G} for its value of $\frac{\exp\left(\frac{-\xi^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}$ and applying the logarithm rule yields

$$h(y) \leq -\int_{\mathbb{R}} p_y(\xi)\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)\,\mathrm{d}\xi - \int_{\mathbb{R}} p_y(\xi)\log(e)\left(\frac{-\xi^2}{2\sigma^2}\right)\,\mathrm{d}\xi. \tag{4.8}$$

The right-hand side, in turn, can be evaluated to

$$\begin{aligned} \frac{1}{2}\log(2\pi\sigma^2) + \frac{\log(e)}{2\sigma^2}\,\mathrm{E}_y[\xi^2] = \frac{1}{2}\log(2\pi\sigma^2) + \frac{\log(e)}{2} &= \\ = \frac{1}{2}\log(2\pi e\sigma^2) &= \\ = h(x_G). & \end{aligned} \tag{4.9}$$

The last passage stems from the fact that $h(x_G) = \frac{1}{2}\log(2\pi e\sigma^2)$. Combining Equation 4.8 and Equation 4.9 one obtains, as wanted, $h(y) \leq h(x_g)$. □

From entropy, one can build a measure of nongaussianity called negentropy. It is defined as

$$J(x) = h(x_G) - h(x), \tag{4.10}$$

where x_G is a Gaussian random variable with variance identical to that of x. In Section 4.2.2, the random variables were whitened before the maximization problem was treated numerically. Since the same whitening step will be performed in this case, all random variables will be assumed to be white, that is, zero mean and unit variance.

From the previous property of entropy, negentropy is nonnegative and is only zero when x is Gaussian. That is the reason negentropy is a very attractive measure of nongaussianity. By maximizing negentropy given a fixed variance, one obtains the most nongaussian random variable.

Note that in order to evaluate negentropy, one must know the pdf of the random variable, or, equivalently, all of its moments. In practice, these moments are unknown, and evaluating even the first higher order moments is both impractical and highly error-prone. Therefore, an approximation of J must be made. The classic approach by the use of cumulants yields the following approximation (first seen in Comon, 1994):

$$J_C(x) = \frac{\mathrm{E}[x^3]^2}{12} + \frac{\mathrm{kurt}[x]^2}{48}. \tag{4.11}$$

This expression results from expanding the pdf in the basis of Hermite polynomials, procedure detailed in Example 1 of Appendix A.

A latent problem with this expression is that it becomes unreliable as the data starts to contain outliers. The reason is because the conventional estimates of kurtosis are sensitive to isolated values that deviate too much from the mean.

A better method for approximating J is by using non-polynomial basis functions in the expansion of the pdf. This is explained in Appendix A. For the present purposes it suffices to choose a few of these functions G_i satisfying the following criteria (Hyvärinen et al., 2001, chap. 5):

1. Estimating $\mathrm{E}[G_i(x)]$ must be simple and robust;
2. $G_i(|x|)$ must not grow faster than quadratically;
3. G_i must measure aspects of the distribution related to entropy. In particular, choosing $G(x) = -\log p(x)$ would clearly be optimal if $p(x)$ was known, since $\mathrm{E}[G(x)] = J(x)$.

Common choices for these are $\log\cosh(x)$ and $-\exp\left(-\frac{x^2}{2}\right)$, but they are not exclusive (see Hyvärinen, 1999; Hyvärinen et al., 2001, chap. 5). It is important to note that while any number of functions can be chosen, it is wise to limit this number to one or two

functions. This is because the more functions are used, the harder it becomes to estimate the expected values $\mathrm{E}[G_i(x)]$, and the expression for the approximated negentropy.

Suppose only one function G is chosen. In this case, the nongaussianity cost function, that is, the approximated negentropy is given by

$$J_1(x) = \frac{1}{2}\,\mathrm{E}[G(x)]^2. \tag{4.12}$$

Thus, the problem after whitening becomes

$$\begin{aligned} &\max_{\mathbf{w}} \ \frac{1}{2}\,\mathrm{E}[G(\mathbf{w}^T\mathbf{z})]^2 \\ &\text{s.t.} \ \ \|\mathbf{w}\|^2 = 1. \end{aligned} \tag{4.13}$$

or

$$\begin{aligned} &\min_{\mathbf{w}} \ -\frac{1}{2}\,\mathrm{E}[G(\mathbf{w}^T\mathbf{z})]^2 \\ &\text{s.t.} \ \ \|\mathbf{w}\|^2 = 1. \end{aligned} \tag{4.14}$$

In this case, different from the solution of Equation 4.5, a quasi-Newton projection method shall be employed. Let $f(\mathbf{w}) = -\frac{1}{2}\,\mathrm{E}[G(\mathbf{w}^T\mathbf{z})]^2$. Therefore,

$$\frac{\partial f}{\partial \mathbf{w}} = -\,\mathrm{E}[G(\mathbf{w}^T\mathbf{z})]\,\mathrm{E}[G'(\mathbf{w}^T\mathbf{z})\mathbf{z}]$$

or

$$\frac{\partial f}{\partial w_i} = -\,\mathrm{E}[G(\mathbf{w}^T\mathbf{z})]\,\mathrm{E}[G'(\mathbf{w}^T\mathbf{z})z_i],$$

where G' is the derivative of the real valued function G. The Hessian is then given by

$$\begin{aligned} \frac{\partial^2 f}{\partial w_i \partial w_j} &= -\frac{\partial}{\partial w_j}\left[\mathrm{E}[G(\mathbf{w}^T\mathbf{z})]\,\mathrm{E}[G'(\mathbf{w}^T\mathbf{z})z_i]\right] \\ &= -\frac{\partial\,\mathrm{E}[G(\mathbf{w}^T\mathbf{z})]}{\partial w_j}\cdot\mathrm{E}[G'(\mathbf{w}^T\mathbf{z})z_i] - \mathrm{E}[G(\mathbf{w}^T\mathbf{z})]\cdot\frac{\partial\,\mathrm{E}[G'(\mathbf{w}^T\mathbf{z})z_i]}{\partial w_j} \\ &= \mathrm{E}[G'(\mathbf{w}^T\mathbf{z})z_j]\,\mathrm{E}[G'(\mathbf{w}^T\mathbf{z})z_i] - \mathrm{E}[G(\mathbf{w}^T\mathbf{z})]\,\mathrm{E}[G''(\mathbf{w}^T\mathbf{z})z_iz_j] \end{aligned}$$

or,

$$\nabla^2 f = -\,\mathrm{E}[G'(\mathbf{w}^T\mathbf{z})\mathbf{z}]\,\mathrm{E}[G'(\mathbf{w}^T\mathbf{z})\mathbf{z}]^T - \mathrm{E}[G(\mathbf{w}^T\mathbf{z})]\,\mathrm{E}[G''(\mathbf{w}^T\mathbf{z})\mathbf{z}\mathbf{z}^T].$$

Note that the second order information is contained in the last term, in G''. Therefore it

is justifiable to approximate $\nabla^2 f$ by

$$\begin{aligned}\nabla^2 f &\approx -\operatorname{E}[G(\mathbf{w}^T\mathbf{z})]\operatorname{E}[G''(\mathbf{w}^T\mathbf{z})\mathbf{z}\mathbf{z}^T] \\ &\approx -\operatorname{E}[G(\mathbf{w}^T\mathbf{z})]\operatorname{E}[G''(\mathbf{w}^T\mathbf{z})]\operatorname{E}[\mathbf{z}\mathbf{z}^T] \\ &\approx -\operatorname{E}[G(\mathbf{w}^T\mathbf{z})]\operatorname{E}[G''(\mathbf{w}^T\mathbf{z})]\mathbf{I}.\end{aligned}$$

Using $\mathbf{H} = -\operatorname{E}[G(\mathbf{w}^T\mathbf{z})]\operatorname{E}[G''(\mathbf{w}^T\mathbf{z})]\mathbf{I}$ is attractive because its inverse is simple to calculate. In fact, the corresponding search direction is given by (see Nocedal and Wright, 2006, chap. 3)

$$\begin{aligned}-\mathbf{H}^{-1}\nabla f &= -\frac{-1}{\operatorname{E}[G(\mathbf{w}^T\mathbf{z})]\operatorname{E}[G''(\mathbf{w}^T\mathbf{z})]}\mathbf{I}\left[-\operatorname{E}[G(\mathbf{w}^T\mathbf{z})]\operatorname{E}[G'(\mathbf{w}^T\mathbf{z})\mathbf{z}]\right] \\ &= -\frac{\operatorname{E}[G'(\mathbf{w}^T\mathbf{z})\mathbf{z}]}{\operatorname{E}[G''(\mathbf{w}^T\mathbf{z})]},\end{aligned}$$

giving rise to the following iteration (Hyvärinen, 1999).

Algorithm 4.2 FastICA for one nonlinearity

Choose random vector $\mathbf{w}_0$.
while $\nabla\mathbf{w}_n \neq 0$ **do**

$$\mathbf{w}^+ \leftarrow \operatorname{E}[\mathbf{z}G'({\mathbf{w}_n}^T\mathbf{z})] - \operatorname{E}[G''({\mathbf{w}_n}^T\mathbf{z})]\mathbf{w}_n$$

$$\mathbf{w}_{n+1} \leftarrow \frac{\mathbf{w}^+}{\|\mathbf{w}^+\|}$$

end while

Chapter 5

ICA and SRME

It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

Arthur Conan Doyle
A Scandal in Bohemia

5.1 Introduction

The SRME method is composed of two stages. First, there is the estimation of the multiple-only common shot gather. Secondly, there is the subtraction of the multiple. In the adaptive SRME, both steps are performed concurrently in each iteration. The ICA-based method explored in this work is different. Instead of subtracting the multiples from the gathers, the ICA based method will attempt to separate the two types of events.

In order to use ICA, the problem must be posed accordingly. Let $\mathbf{p}$ denote a full common shot gather, $\mathbf{p}_0$ denote the primaries and $\mathbf{m}$ denote the multiples, all as column vectors. The method shown in the Chapter 3 yields, from $\mathbf{p}$, an estimate of $\mathbf{m}$, which shall be denoted as $\tilde{\mathbf{m}}$. This estimate is then deconvolved to obtain a new estimate $\tilde{\mathbf{m}}_d$. This deconvolution can be done in many ways. One such way is by adaptive filtering, and can be performed by finding $\mathbf{f}^{(i)}$ such that

$$\mathbf{f}^{(i)} = \arg\min_{\mathbf{f}\in\mathbb{R}^{N_f}} \|\mathbf{p}^{(i)} - \mathbf{f} * \tilde{\mathbf{m}}^{(i)}\|,$$

where the i superscript refers to the ith window, and setting $\tilde{\mathbf{m}}_d^{(i)} = \mathbf{f}^{(i)} * \tilde{\mathbf{m}}^{(i)}$. Simply put, $\tilde{\mathbf{m}}_d$ is obtained using a window filter matching between $\mathbf{p}_0$ and $\tilde{\mathbf{m}}$. This window can be in time and space, and different authors have varying approaches (for a per trace approach see Verschuur and Berkhout (1997) and for a windowed approach, see Ventosa et al. (2012)). In this work, all of them will be tested, and the best method will be chosen for each separate dataset. Since the predicted deconvolved multiples do not match the true multiples exactly and the data is composed of only primaries and multiples, then there is bound to be residue of primaries in $\tilde{\mathbf{m}}_d$. The second equation stems from the fact that the total data is a linear combination of primaries and multiples.

$$\tilde{\mathbf{m}}_d = a_{21}\mathbf{p}_0 + a_{22}\mathbf{m} \tag{5.1}$$

$$\mathbf{p} = a_{11}\mathbf{p}_0 + a_{12}\mathbf{m} \tag{5.2}$$

These two equations give rise to a 2×2 system of equations

$$\mathbf{As} = \mathbf{x}, \tag{5.3}$$

where $\mathbf{s}^T = [\mathbf{p}_0^T \ \mathbf{m}^T]$, $\mathbf{x}^T = [\mathbf{p}^T \ \tilde{\mathbf{m}}_d^T]$ and $(\mathbf{A})_{ij} = a_{ij}$. Supposing that $\mathbf{m}$ and $\mathbf{p}_0$ are independent, one may apply ICA. This formulation is similar to that of Donno (2011), while Lu (2006) and Kaplan and Innanen (2008) use larger mixing matrices.

5.2 Single-reflector 1D models

In order to investigate the effectiveness of the ICA method for multiple removal it is important to use both simple and complex models. The simplest models one may devise are those of a flat-layered earth, with only one reflector. Two such models are Model 1, displayed in Figure 5.1, and Model 2. These first simple tests will serve as basis for the next, more complicated models.

5.2.1 Model 1

This model is essentially a single layered model, since the first reflector is the acquisition surface, and the model is deep enough (10 km, not shown) such that there are no reflections on the bottom of the model. The 2D SRME method requires sufficient spacial sampling of the source and receivers. It also requires close to zero-offset traces. In order to satisfy these requirements, the acquisition was done placing the receptors at distances

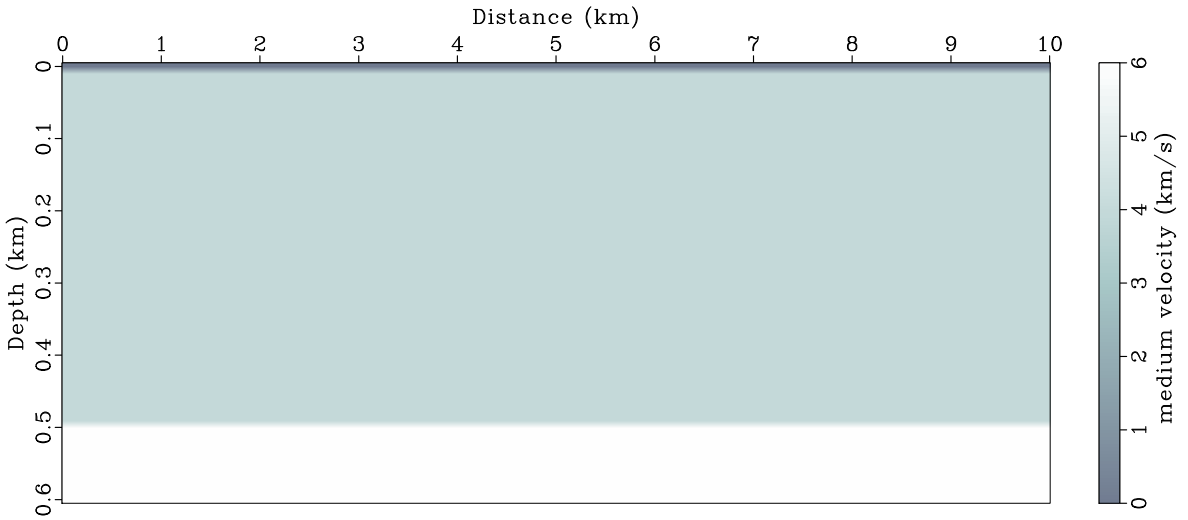
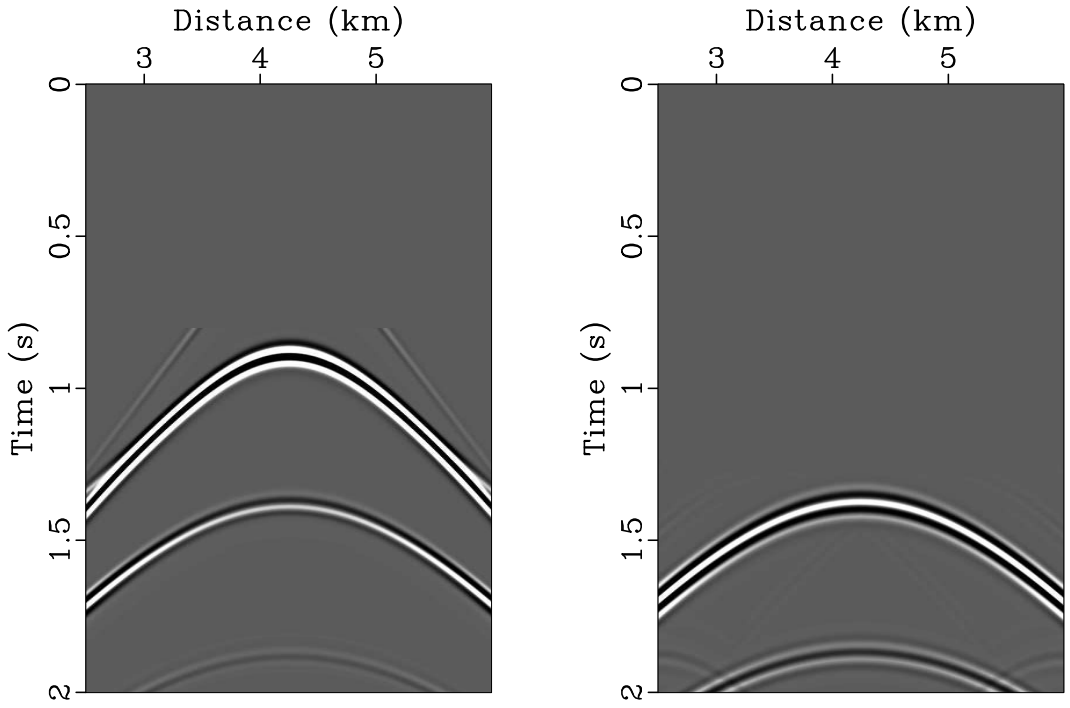


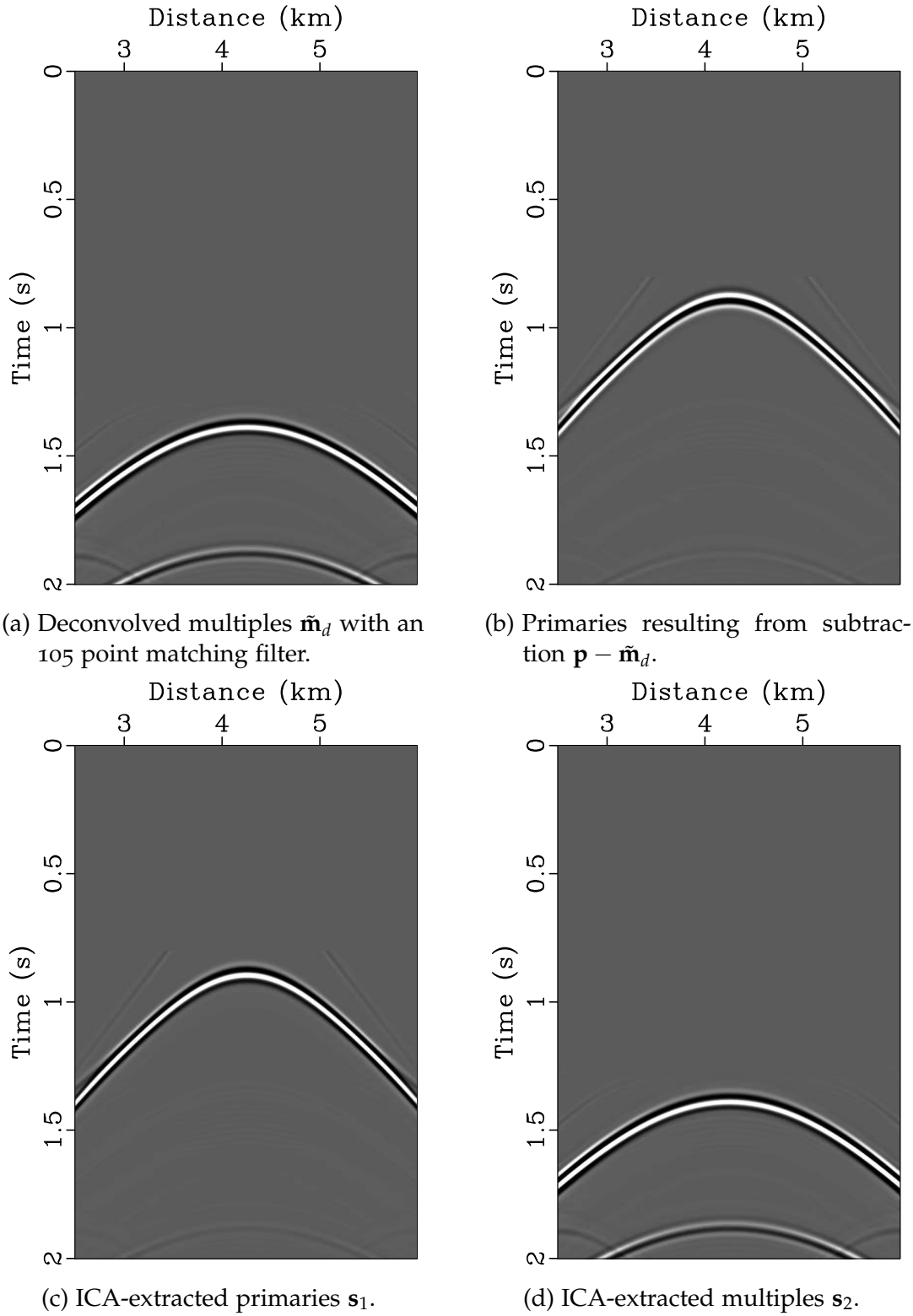
Figure 5.1: Velocity model. The reflectors are positioned at depths of 10 km and 500 m. The velocities in the layers are $0\,\mathrm{km\,s^{-1}}$, $4\,\mathrm{km\,s^{-1}}$ and $6\,\mathrm{km\,s^{-1}}$.



(a) Common shot with partially removed direct wave $\mathbf{p}$.

(b) Predicted multiples $\tilde{\mathbf{m}}$.

Figure 5.2: Common shot gather and its corresponding predicted (first) multiples. Notice the artifacts generated by the residual direct wave.



(a) Deconvolved multiples $\tilde{\mathbf{m}}_d$ with an 105 point matching filter.

(b) Primaries resulting from subtraction $\mathbf{p} - \tilde{\mathbf{m}}_d$.

(c) ICA-extracted primaries $\mathbf{s}_1$.

(d) ICA-extracted multiples $\mathbf{s}_2$.

Figure 5.3: Common shot gather and its multiples by LS subtraction and ICA.

from 0 km to 10 km, in steps of 10 m. Each shot simulation was equally ranged and spaced. All sources and receivers were placed at a depth of 10 m.

In Chapter 3, Figure 3.2 shows the modeled response of the earth from a source placed 5 km from the origin. In order to apply SRME, the direct wave must be removed. For this model, the direct wave was only be partially removed.

Figure 5.2 shows the common shot gather and its predicted first order multiples. Figures 5.2a and 5.2b show that further processing must be done before subtraction can be applied. This has been discussed in Section 5.1 where it was stated how the source signature needs to be deconvolved. The deconvolved multiples, $\tilde{\mathbf{m}}_d$, and the direct subtraction, $\mathbf{p} - \tilde{\mathbf{m}}_d$, are shown in Figures 5.3a and 5.3b, respectively. Figures 5.3d and 5.3c shows both ICs recovered applying FastICA (detailed in Algorithm 4.2) to the input data of $\mathbf{p}$ and $\tilde{\mathbf{m}}_d$.

	Number of filter points				
	1	35	70	105	140
ICA extraction	23.16%	3.76%	3.70%	3.41%	6.85%
LS subtraction	23.83%	3.78%	3.77%	3.53%	7.10%

Table 5.1: Error of the estimated primaries using differently sized filters. The error is given by $\|\mathbf{p}_0 - \mathbf{p}_0^{\text{est}}\| / \|\mathbf{p}_0\|$. Note that the length of the source is 70 samples.

The difference between Figure 5.3b is Figure 5.3c barely noticeable visually, if noticeable at all. This is expected. In the model, none of the primaries intersected the multiples. As such, there is to be very little "leakage" of primaries into the predicted multiples. Most errors stem from the artifacts generated by less than ideal data. Relative errors are shown in Table 5.1 for different filters.

5.2.2 Model 2

Model two is very similar to the previous one. The main differences will stem from how the data is processed. Figure 5.4 depicts the velocity model. The acquisition is done exactly as detailed above, with the same parameters.

The direct wave was correctly removed in the entire data, but the refractions were left in. The reason why they were not removed is evident in Figure 5.5b: the refraction of the multiple is predicted from the refraction of the primary. Also, in order to simulate a more marine geometry, the data was cut as in a split-spread geometry.

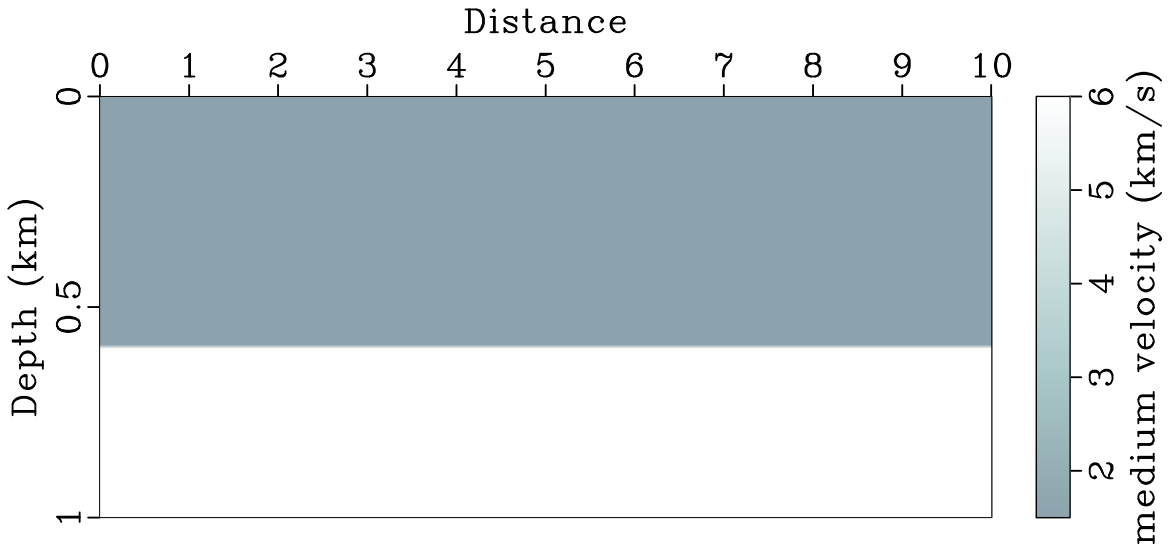
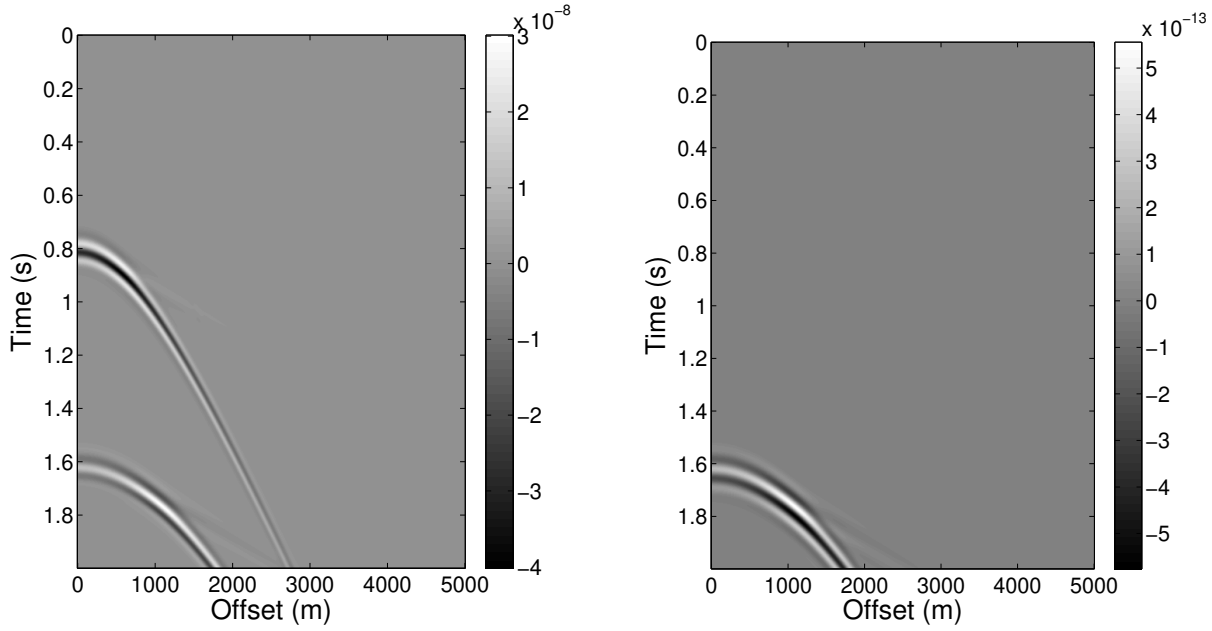


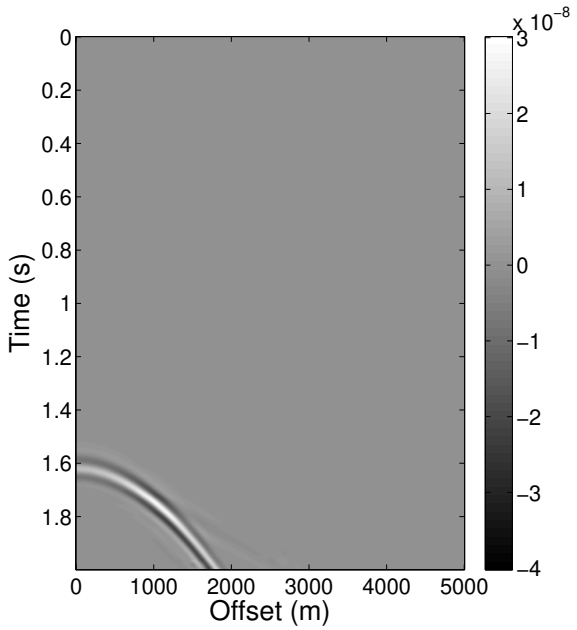
Figure 5.4: Velocity model. The reflectors are positioned at depths of 10 km and 600 m. The velocities in the layers are 1.5 km s^{-1} and 6 km s^{-1}.



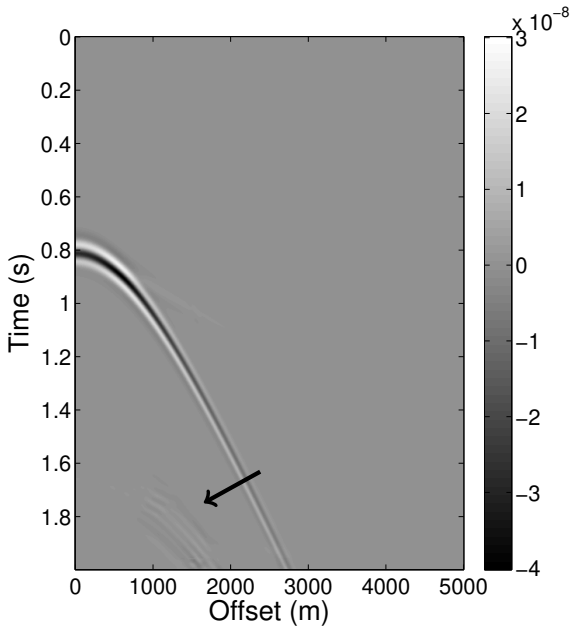
(a) Common shot with partially removed direct wave $\mathbf{p}$.

(b) Predicted multiples $\tilde{\mathbf{m}}$.

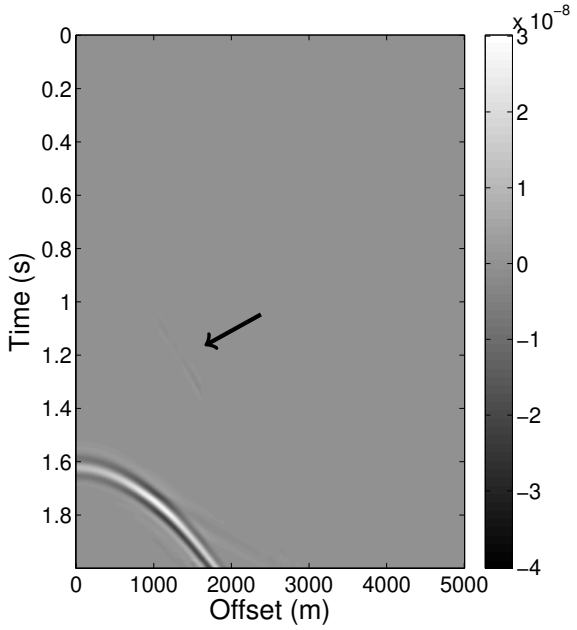
Figure 5.5: Common shot gather and its corresponding predicted multiple. The multiple prediction panel shows no artifacts.



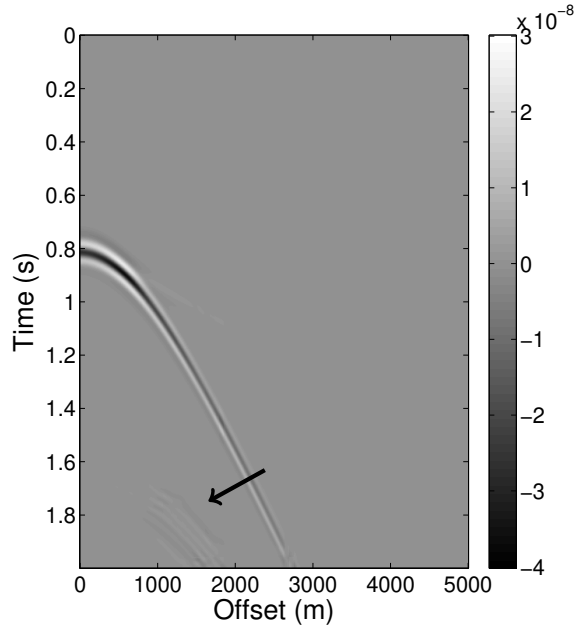
(a) Deconvolved multiples $\tilde{\mathbf{m}}_d$ with a 30 point matching filter.



(b) Primaries resulting from subtraction $\mathbf{p} - \tilde{\mathbf{m}}_d$.



(c) ICA-extracted multiples $\mathbf{s}_2$.



(d) ICA-extracted primaries $\mathbf{s}_1$.

Figure 5.6: Common shot gather and its multiples by LS subtraction and ICA. Differences indicated by the arrow.

Figure 5.5 depicts the common shot section that will be worked on and its predicted first multiple. One must notice how the first multiple in Figure 5.5b is on a much lower scale than that in Figure 5.5a. This evidences the need for a correction before subtraction.

In order to subtract the multiple, as was done in the previous model, different tests were performed. Since there are many parameters to set in each methods, only the one yielding the best results will be displayed. Following are the results of those experiments.

Firstly, the least squares subtraction method was performed using windows. The result displayed in Figure 5.6b is the one with the best performance, and it corresponds to that of a time window of 90 traces, a spacial window of 35 traces and 30 point filter. One can observe how the multiple was been well attenuated, but in a less than perfect way. The residuals are seen under the arrow.

This can be attributed to the fact that the wavelet occupies many samples, and would require a long filter to be correctly deconvolved. However, since the deconvolution process becomes unstable as the filter grows, the length of 30 samples was the largest filter that would not botch any window. Both the prediction of the multiples and the least squares subtraction were performed with software from the DELPHI consortium, which was generously provided by Eric Verschuur.

The ICA method is displayed in figures 5.6c and 5.6d. In this case, instead of doing in only one window as before, short spacial windows of 6 traces were used. The result is similar to what was seen before, the least squares subtraction in a simple model is hard to be improved by ICA. However, in this result we see something that was not seen before, namely, that primary leaked into the extracted multiple. This is marked with an arrow in Figure 5.6c. Nonetheless, the extracted primaries also have less multiples. This is seen by comparing the areas indicated by the arrows in figures 5.6b and 5.6d. The area in the ICA extracted primary panel is slightly more discolored than that in the panel corresponding to the LS subtraction.

Had we wanted to discriminate only the multiples it would be considered a regression in the result of the least squares subtraction scheme. However, since the objective is to obtain a panel with attenuated multiples, this can be seen as an improvement. With these results at hand, we may now attempt to separate multiples and primaries on a slightly more complicated model.

5.3 Double-reflector 1D model

This model is also a constant velocity model, however it has two horizontal reflectors. The acquisition, as previously, was done in a split-spread fashion, with the same spacing of 10 m in both the shot and the receiver directions.

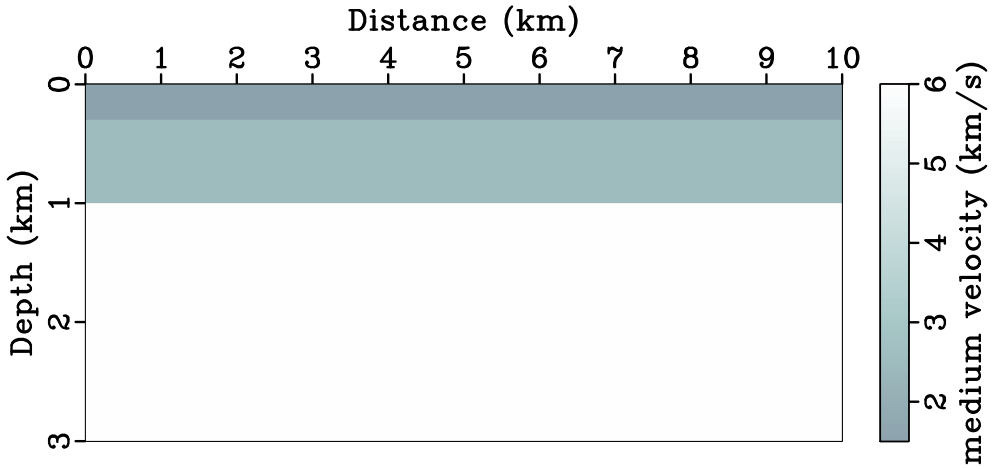


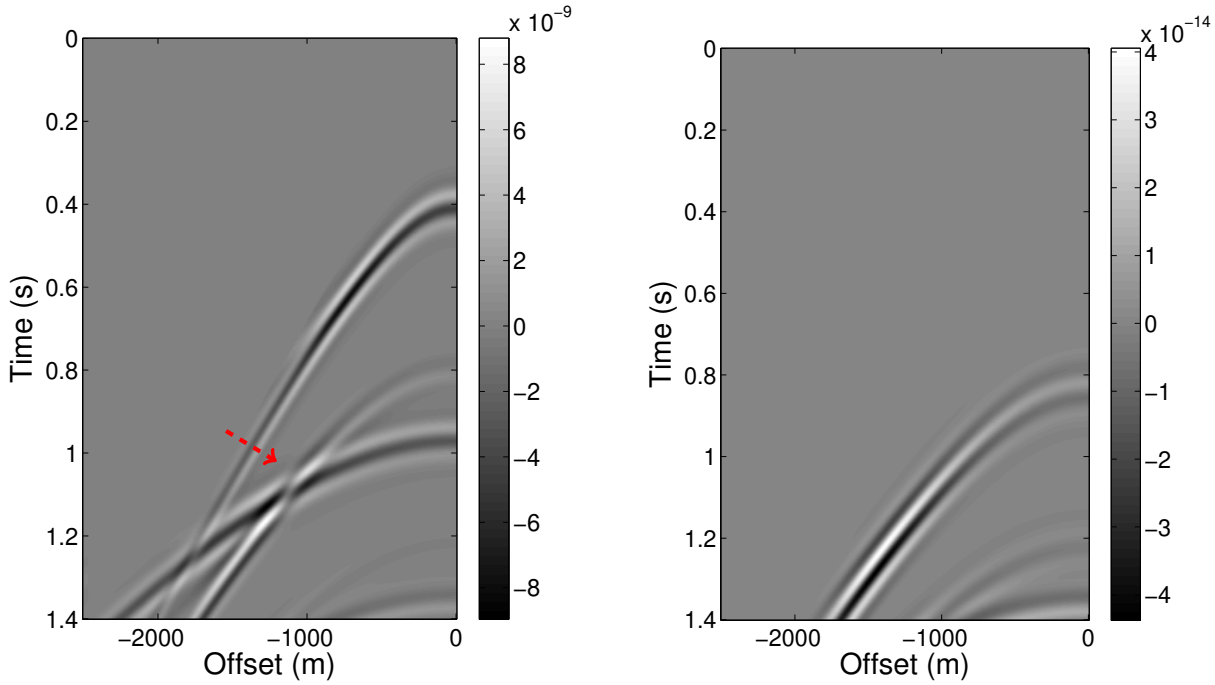
Figure 5.7: Velocity model. The reflectors are positioned at depths of 1 km and 300 m. The velocities in the layers are 1.5 km s^{-1}, 2.5 km s^{-1} and 6 km s^{-1}.

The presence of two reflectors allows for the primary reflection of the second reflector to overlap with the first multiple reflection of the second reflector on a common shot panel. This overlap is indicated in Figure 5.8a by the red dashed arrow. The preprocessing done to the data prior to the multiple prediction step was only the muting of the direct wave and the refraction. Since the offsets were not distant enough for the refraction of the multiple reflection to appear, it was also muted from the first primary.

One must note that there are in fact three multiples as is seen in Figure 5.8. The first two events are the first and second multiple of the first reflector, respectively. The final event is the first multiple of the second reflector.

The LS subtraction was performed in windows. The length in the time dimension was 701 samples, i.e. the full trace, and in the space direction was 100 samples. Smaller time and spacial windows made the process unstable, and presented zeroed windows in some cases. The filter length was of 31 points. The results of this subtraction are seen in Figure 5.9b.

A few things can be seen in the results of Figure 5.8. Regarding the multiples, it is evident that no primaries were leaked to the extracted multiples as happened in Model 2. This serves to show that while the leakage may occur, it will depend on the data,



(a) Common shot with muted direct wave and first refraction $\mathbf{p}$.

(b) Predicted multiples $\tilde{\mathbf{m}}$.

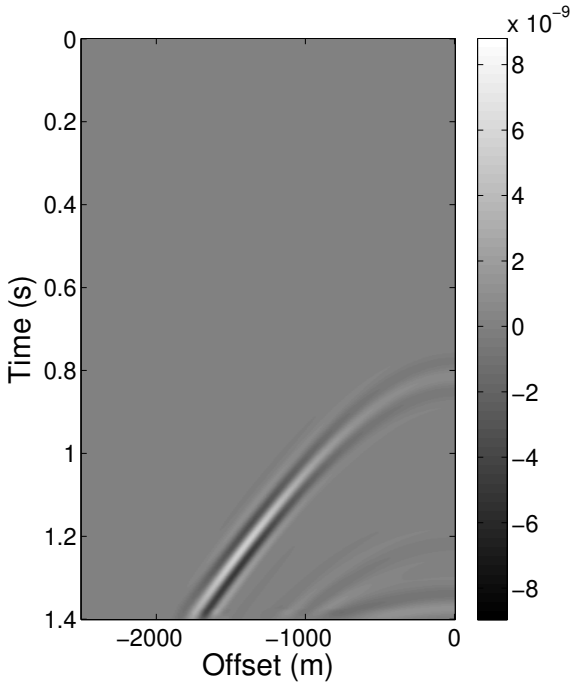
Figure 5.8: Common shot gather and its corresponding predicted multiple.

and even if it does happen, it will be of a low magnitude. Therefore, it can safely be ignored. In the area indicated by the black solid arrow in figures 5.9b and 5.9d, the LS subtraction shows a multiple that is not observed the panel of the ICA extraction. The same observation can be made of the area marked by the blue dotted arrow. These areas are of the first multiples of the first and second reflector, respectively. Therefore, the ICA extraction performed better in terms of multiple removal.

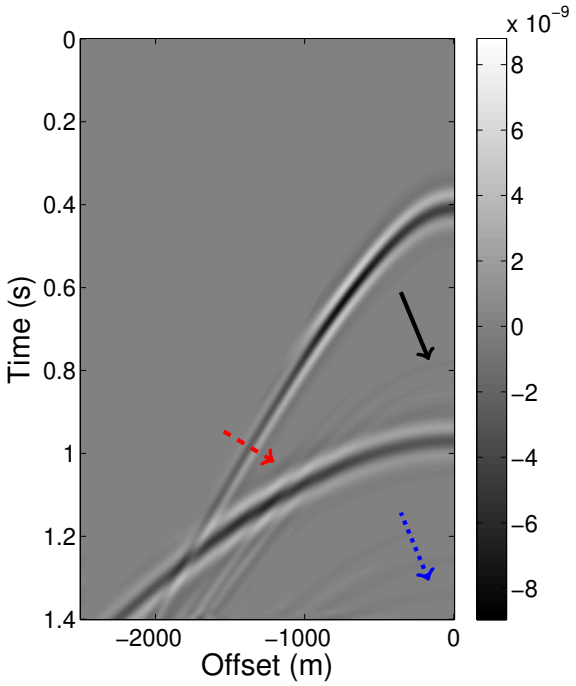
In objective terms, the sample at offset -120 m and time 0.784 s (under the black solid arrows) of $\mathbf{p} - \tilde{\mathbf{m}}_d$ had an amplitude of -1.135×10^{-10}, while the same sample in $\mathbf{s}_1$ had an amplitude of -4.776×10^{-11}.

In relation to the primaries, one can also see slight improvements in the ICA panel over the LS panel. Under the red dashed arrow, for example, on the sample at offset -1120 m and time 1.096 s the magnitude of the ICA panel is higher than that of the LS. Precisely, the sample of $\mathbf{s}_1$ has amplitude -5.105×10^{-9} while that of $\mathbf{p} - \tilde{\mathbf{m}}_d$ is -4.491×10^{-9}. It is a slight but positive difference, since it enhanced the primary reflection.

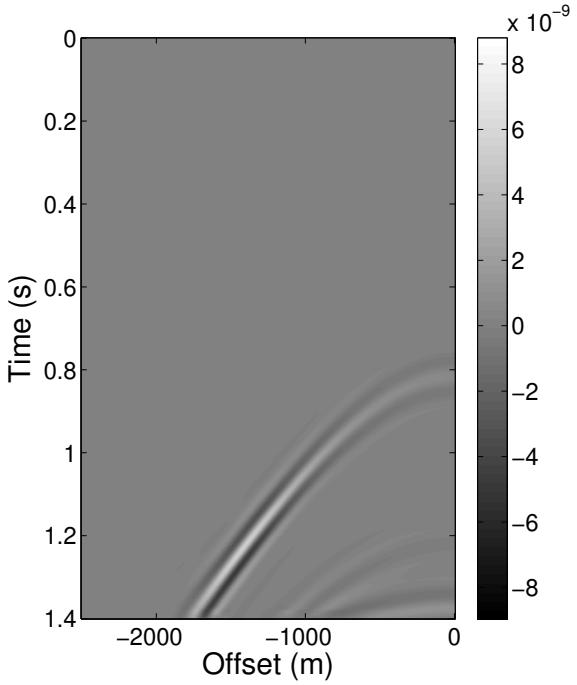
This model has shown that ICA can be a post-processing step after least squares subtraction, since it attenuated the multiples while enhancing the primaries.



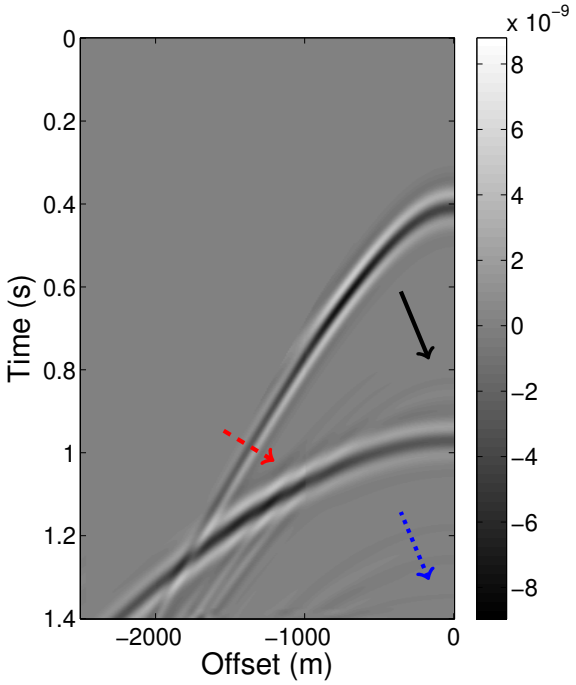
(a) Deconvolved multiples $\tilde{\mathbf{m}}_d$ with a 31 point matching filter.



(b) Primaries resulting from subtraction $\mathbf{p} - \tilde{\mathbf{m}}_d$.



(c) ICA-extracted multiples $\mathbf{s}_2$.

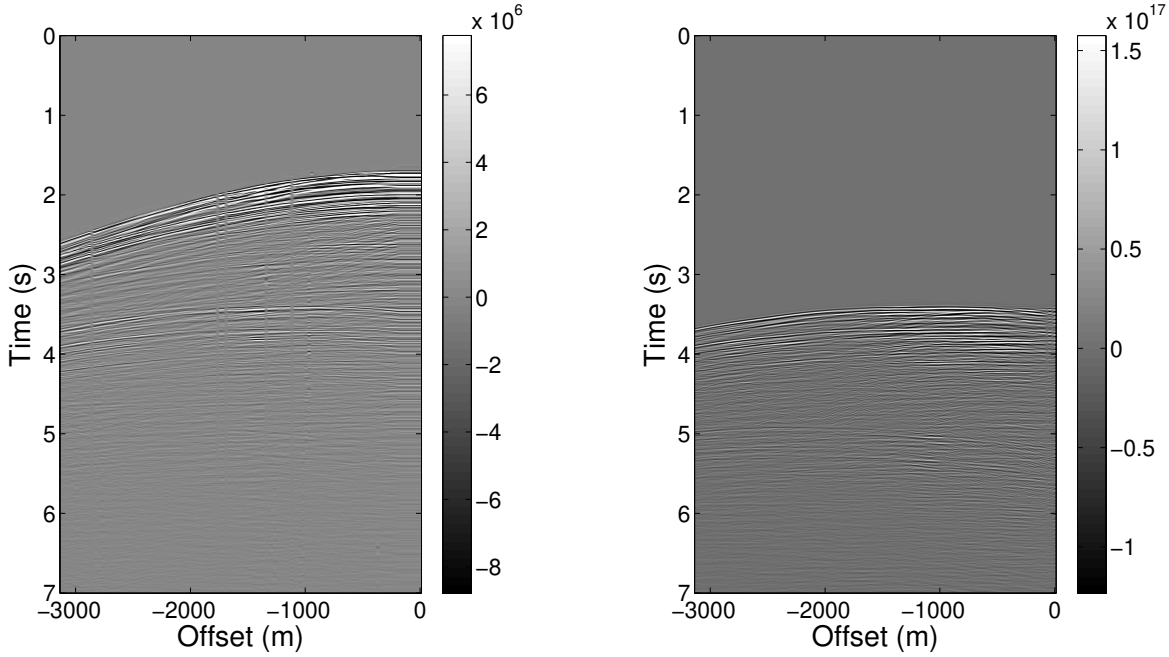


(d) ICA-extracted primaries $\mathbf{s}_1$.

Figure 5.9: Common shot gather and its multiples by LS subtraction and ICA. Differences indicated by the arrows.

5.4 Field dataset

Following the two synthetic models, the next experiment was performed on a real dataset acquired in the Jequitinhonha Basin which kindly ceded by Petrobras. The acquisition has a receiver spacing of 25 m, which is the same spacing that between shots. The minimum offset in a shot is -150 m and the maximum is -3125 m. The full data used features a total of 878 shots; one of them is shown in Figure 5.10a. The recording time is of 7 s: 1751 samples recorded at every 4 ms.

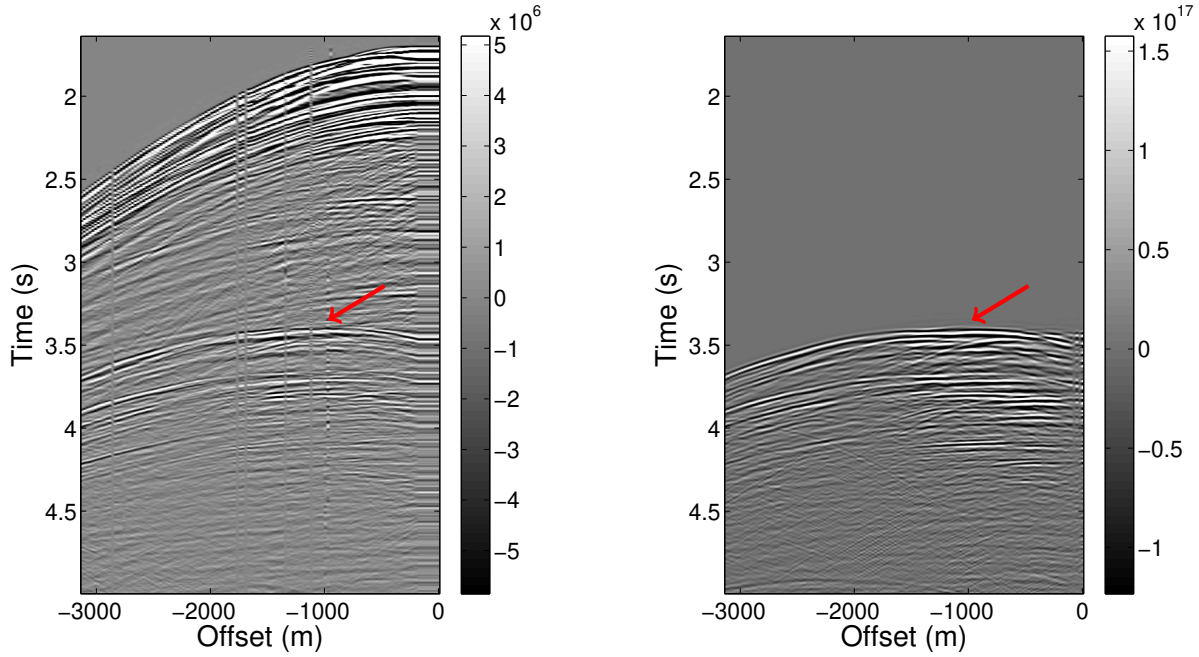


(a) Common shot panel $\mathbf{p}$ (clipped).

(b) Predicted multiples $\tilde{\mathbf{m}}$ (clipped).

Figure 5.10: Common shot gather and its corresponding predicted multiple.

After preparing the data and applying the prediction step of SRME, one obtains Figure 5.10b. Figure 5.11a shows a window of the common shot panel in Figure 5.10a, for a better view of the events. Analogously, Figure 5.11b shows the same window but of Figure 5.10b. In these figures one can observe, by aid of the arrows, the first and most prominent multiple reflection, as well as other multiples.



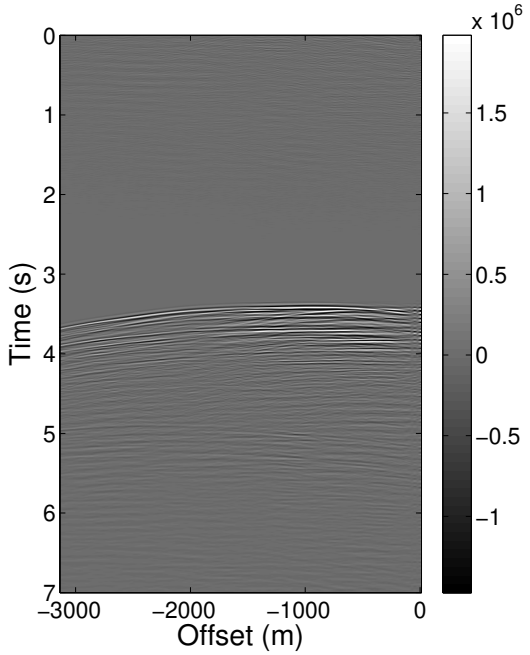
(a) Common shot panel $\mathbf{p}$ (clipped).

(b) Predicted multiples $\tilde{\mathbf{m}}$ (clipped).

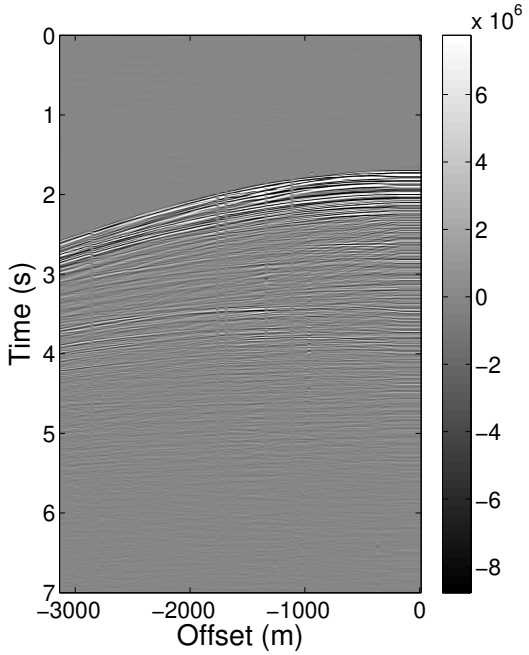
Figure 5.11: Windowed common shot gather and its predicted multiples.

Firstly, the horizontal small offsets in Figure 5.10a are a result of the technique used to fill the data up to zero offset, which consisted of copying the closest offset trace (in this case $-150\,\mathrm{m}$) to the offsets up to the zero offset. This technique is one of the three most common ones used, the other two being zero-padding and extrapolation by Radon transform.

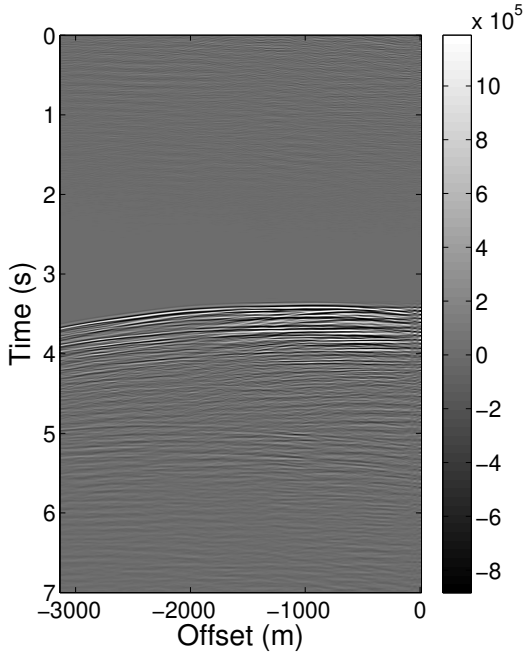
The zero-padding technique is the one with the worst results but arguably the fastest, while the Radon extrapolation provides the most accurate results at the cost of speed, since it requires extensive human interaction. Copying the smallest offset trace is an in between solution, since it is completely automated and provides reasonable results. Its shortcomings, as can be seen in Figure 5.11b, is that the multiple is poorly predicted close to the zero offset. Nonetheless, since the data that will be used for posterior processing should only contain the original offsets, this is not an unsurmountable problem. It should be noted, however, that for shallow water acquisitions, Radon extrapolation is recommended, since small offset data becomes valuable.



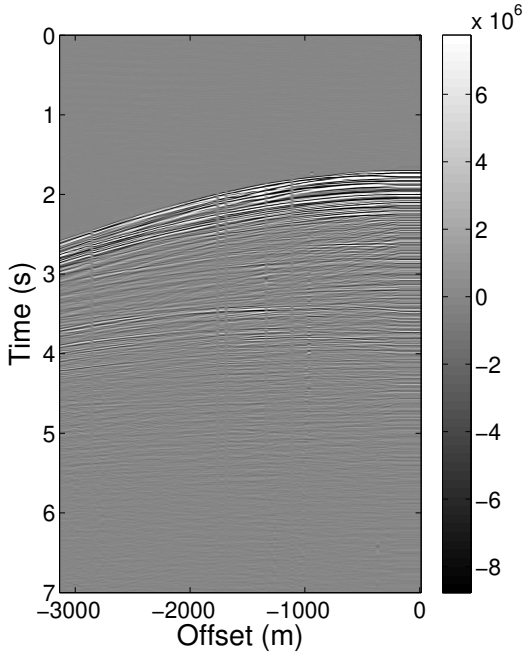
(a) Deconvolved multiples $\tilde{\mathbf{m}}_d$.



(b) Primaries resulting from $\mathbf{p} - \tilde{\mathbf{m}}_d$.

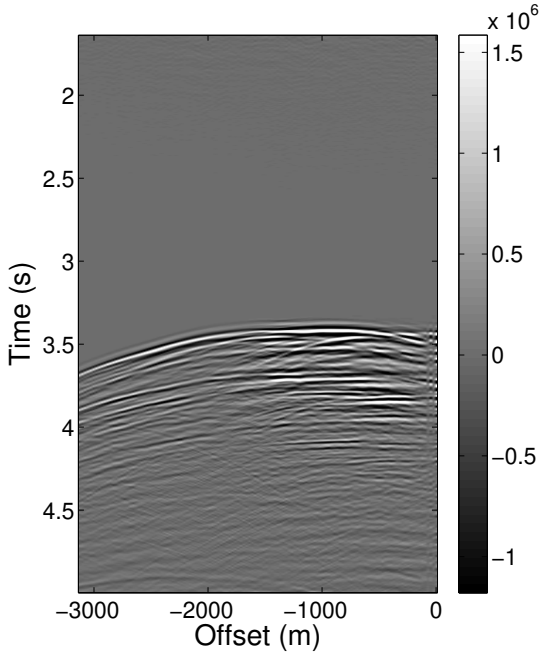


(c) ICA-extracted multiples $\mathbf{s}_2$.

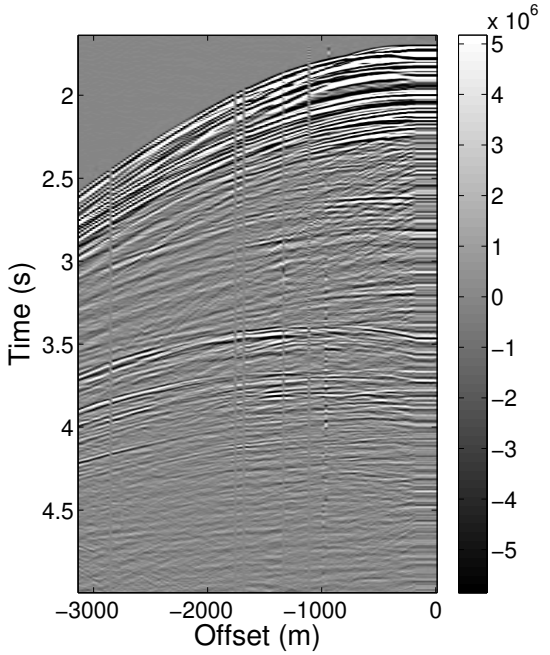


(d) ICA-extracted primaries $\mathbf{s}_1$.

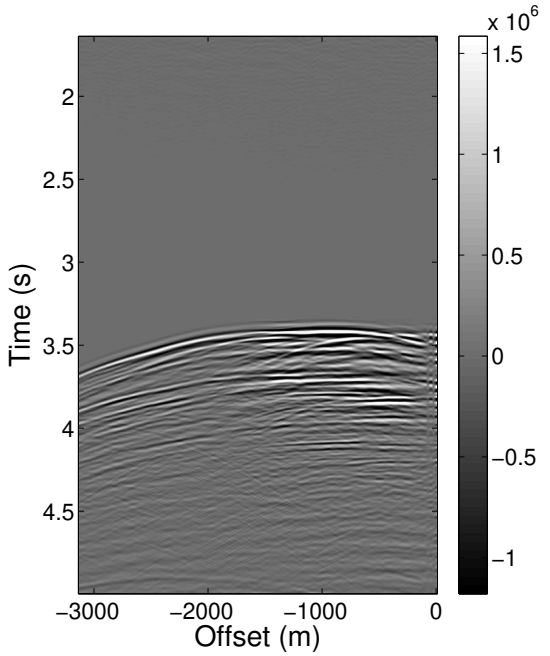
Figure 5.12: Common shot gather and its multiples by LS subtraction and ICA.



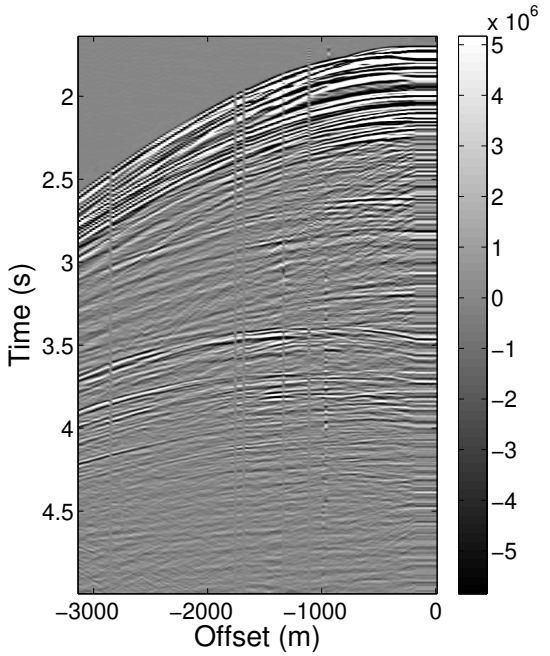
(a) Windowed deconvolved multiples $\tilde{\mathbf{m}}_d$.



(b) Windowed primaries resulting from subtraction $\mathbf{p} - \tilde{\mathbf{m}}_d$.



(c) Windowed ICA-extracted multiples $\mathbf{s}_2$.



(d) Windowed ICA-extracted primaries $\mathbf{s}_1$.

Figure 5.13: Windowed common shot gather and its multiples by LS subtraction and ICA.

After the prediction, a least squares filter was applied with the following parameters: time window of 700 samples and a spacial window of 100 samples. The process resulted in two panels, namely, the one with attenuated multiples, and another with the deconvolved multiple prediction. Following the filtering, ICA was performed using the deconvolved multiples and the original common shot panel, in order for the real multiples to be extracted.

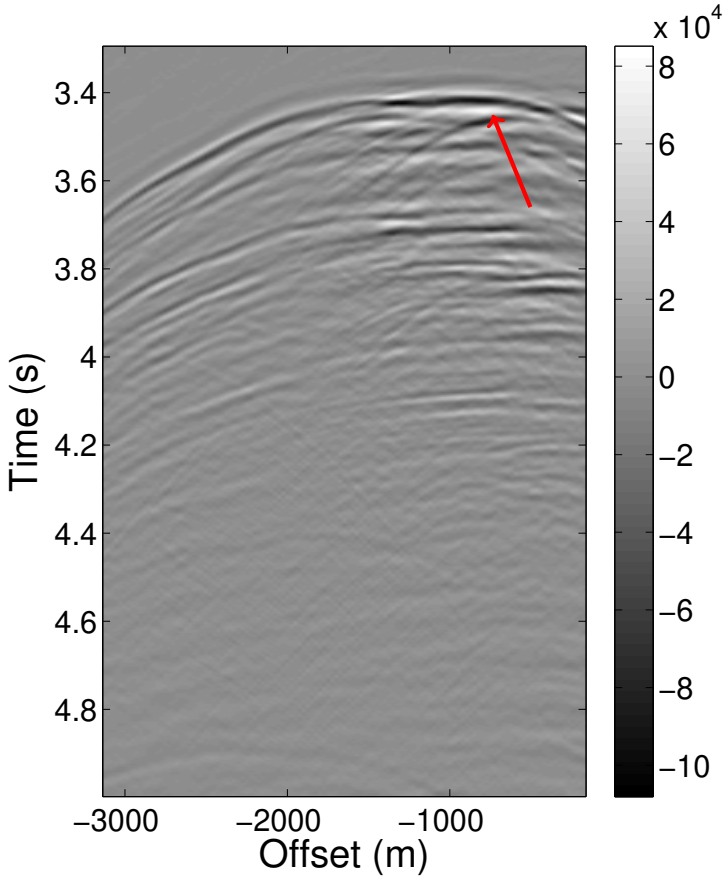


Figure 5.14: Difference panel $\mathbf{s}_1 - (\mathbf{p} - \tilde{\mathbf{m}}_d)$.

This is the same process that was applied to the synthetic datasets and its results can be seen in Figure 5.12. The windowed version of these panels are displayed in Figure 5.13. As in the previous model, thanks must be given to Eric Verschuur for allowing the use of the software for prediction and subtraction of the multiples.

The results show above are hard to compare with such images, as the differences between the two techniques are small. Therefore it is important to introduce a difference panel, such as in Figure 5.14. In it, one observes the difference $\mathbf{s}_1 - (\mathbf{p} - \tilde{\mathbf{m}}_d)$ in a time

window that focuses on the first multiple events. It is important to note that the ICA panel $\mathbf{s}_1$ has been renormalized to obtain a meaningful comparison. If the ICA panel attenuated the multiples better, than we should observe negative values on the multiples. This is because its absolute value is lower than that of the LS subtraction. Likewise for primaries, one should observe positive values for those. That is, in fact, what happens. Observing the first multiple reflection (under the red arrow), they exhibit a black color, which is characteristic of negative values. Those values above this event, which are primaries, are almost all light gray, showing them to be positive. Consequently, it can be said that the ICA extraction was successful, even though its improvement has about two orders of magnitude less than the primary events.

Chapter 6

Conclusion

Tudo acaba, leitor; é um velho truísmo, a que se pode acrescentar que nem tudo o que dura, dura muito tempo. Esta segunda parte não acha crentes fáceis; ao contrário, a ideia de que um castelo de vento dura mais que o mesmo vento de que é feito, dificilmente se despegará da cabeça, e é bom que seja assim, para que se não perca o costume daquelas construções quase eternas.

Machado de Assis
Dom Casmurro

Je n'ai fait celle-ci plus longue que parce que je n'ai pas eu le loisir de la faire plus courte.

Blaise Pascal
Lettres provinciales

In this monograph, two different techniques were studied separately and then applied together in order to better attenuate certain types of multiple reflections in seismic data. An overview of the problem was presented in the first section. Then, some basic theory on the propagation of acoustic waves was developed in order to serve as a firm mathematical basis for the technique of surface-related multiple elimination. Following the mathematical description of the physical phenomena, the technique of independent component analysis was formulated and described. It was then applied to the problem at hand, serving as a bridge between geophysics and signal processing. Results were

shown in Chapter 5. These results stemmed from four different datasets. The first three datasets were generated using simple synthetic models. They were first studied in order to gauge the performance of the algorithm and to gain experience for more elaborate data. The final dataset was real field data, of the Jequitinhonha Basin, Brazil.

The technique of ICA, applied after a least squares subtraction was shown to be beneficial to the data in general. Even though the gain were slight, the technique showed an improvement over the standard least squares elimination. This was seen particularly well in the synthetic models. In the field dataset, a subtraction panel shows that ICA did also improve the subtraction, however the nature of the data makes it hard to evaluate the performance for higher order multiples, since they are hard to identify in real data.

Nonetheless, there are downsides to ICA. First of all, it requires an optimization scheme that, though fast for one small common shot panel, may be a burden if applied to the full data. Secondly, it adds one more step whose parameters need to be carefully set for each different dataset, and maybe even different common shot panels of the same acquisition. Since the gains were not dramatic, it means that one may be adding an unnecessary amount of processing of prestack data for negligible differences in a poststack section.

The study done here has room for expansion. There are a few things that could have been done differently, that might also yield interesting results. First off, a lot of emphasis was put on the nongaussianity aspect of ICA, and the only technique to separate signals was FastICA. As is known, however, there are many complementary methods to the negentropy approach, such as likelihood methods and tensorial methods. Furthermore, more general forms of independent component analysis can be studied, such as one using a convolutional model, instead of an additive model.

Regarding the prediction of the multiples, any method that accurately predicts multiples could be used for an estimation of the multiples. For example, wave field extrapolation methods could be successfully used in this case. The three dimensional version of SRME could also be used, and since it obtains better results that the two dimensional version used, it is possible that ICA will perform even better in that case.

This work has established that this field of research can be a healthy addition to the plethora of multiple elimination and attenuation methods in seismic geophysics. The results were positive and have opened up new avenues of future research in the area.

Bibliography

Abramowitz, M., and I. A. Stegun, eds., 1972, Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, 10 ed.: National Bureau of Standards, volume **55** *of* National Bureau of Standards Applied Mathematical Series.

Amari, S., A. Cichocki, and H. H. Yang, 1997, A New Learning Algorithm for Blind Signal Separation: Advances in Neural Information Processing Systems, 757–763.

Berkhout, A. J., 1985, Seismic Migration: Imaging of Acoustic Energy by Wave Field Extrapolation, 3 ed.: Elsevier, volume **A** *of* Developments in Solid Earth Geophysics.

Berkhout, A. J., and D. J. Verschuur, 1997, Estimation of multiple scattering by iterative inversion, Part I: Theoretical considerations: Geophysics, **62**, 1582–1595.

Bleistein, N., 1984, Mathematical Methods for Wave Phenomena: Academic Press, Inc.

Cao, X.-R., and R. W. Liu, 1996, General approach to blind source separation: IEEE Transactions on Signal Processing, 562–571.

Cardoso, J.-F., 1989, Source separation using higher order moments: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2109–2112.

———, 1990, Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2655–2658.

———, 1996, Equivariant adaptive source separation: IEEE Transactions on Signal Processing, **44**, 3017–3030.

Cardoso, J.-F., and A. Souloumiac, 1993, Blind beamforming for non Gaussian signals: Radar and Signal Processing, IEE Proceedings F, **140**, 362–370.

Comon, P., 1994, Independent component analysis, A new concept?: Signal Processing, **36**, 287–314.

Cover, T. M., and J. A. Thomas, 2006, Elements of Information Theory: Wiley-Interscience.

Červený, V., 2001, Seismic Ray Theory: Cambridge University Press.

de Bruin, C. G. M., C. P. A. Wapenaar, and A. J. Berkhout, 1990, Angle-dependent reflectivity by means of prestack migration: Geophysics, **55**, 1223–1234.

Donno, D., 2011, Improving multiple removal using least-squares dip filters and independent component analysis: Geophysics, **76**, V91–V104.

Dragoset, B., E. Verschuur, I. Moore, and R. Bisley, 2010, A perspective on 3D surface-related multiple elimination: Geophysics, **75**, 75A245–75A261.

Eriksson, J., and V. Koivunen, 2004, Identifiability, Separability, and Uniqueness of Linear ICA Models: IEEE Signal Processing Letters, **11**.

Gaeta, M., and J.-L. Lacoume, 1990, Source separation without prior knowledge: the maximum likelihood solution: Proceedings of the European Signal Processing Conference, 621–624.

Giannakis, G., Y. Inouye, and J. Mendel, 1989, Cumulant based identification of multichannel moving-average models: Automatic Control, IEEE Transactions on, **34**, 783–787.

Herault, J., and C. Jutten, 1987, Space or time adaptive signal processing by neural network models: AIP Conference Proceedings 151 on Neural Networks for Computing, American Institute of Physics Inc., 206–211.

Hyvärinen, A., 1999, Fast and robust fixed-point algorithms for independent component analysis: IEEE Transactions on Neural Networks, **10**, 626–634.

Hyvärinen, A., J. Karhunen, and E. Oja, 2001, Independent Component Analysis: John Wiley & Sons, Inc.

Kaplan, S. T., and K. A. Innanen, 2008, Adaptive separation of free-surface multiples through independent component analysis: Geophysics, **73**, V29–V36.

Koralov, L. B., and Y. G. Sinai, 2007, Theory of Probability and Random Processes, second ed.: Springer. Universitext.

Kreyszig, E., 1978, Introductory Functional Analysis with Applications: John Wiley & Sons, Inc.

Linkster, R., 1992, Local synaptic learning rules suffice to maximise mutual information in a linear network: Neural Computation, **4**, 691–702.

Lu, W., 2006, Adaptive multiple subtraction using independent component analysis: Geophysics, **71**, S179–S184.

Lu, W., and L. Liu, 2009, Adaptive multiple subtraction based on constrained independent component analysis: Geophysics, **74**, V1–V7.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization, second ed.: Springer. Springer Series in Operations Research and Financial Engineering.

Pham, D.-T., P. Garrat, and C. Jutten, 1992, Separation of a mixture of independent sources through a maximum likelihood approach: Proceedings of the European Signal Processing Conference, 771–774.

Robinson, E. A., and S. Treitel, 2009, Geophysical Signal Analysis, seg ed.: Society of Exploration Geophysicists.

Rudin, W., 1987, Real and complex analysis: McGraw-Hill.

Sheriff, R. E., and L. P. Geldart, 1995, Exploration Seismology, 2 ed.: Cambridge University Press.

Sommerfeld, A., 1949, Partial Differential Equations in Physics: Academic Press.

Taleb, A., and C. Jutten, 1999, On undertermined source separation: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1445–1448.

Ventosa, S., S. L. Roy, I. Huard, A. Pica, H. Rabeson, P. Ricarte, and L. Duval, 2012, Adaptive multiple subtraction with wavelet-based complex unary Wiener filters: Geophysics, **77**, V183–V192.

Verschuur, D. J., 1991, Surface-related Multiple Elimination: An Inversion Approach: PhD thesis, Delft University of Technology.

———, 2006, Seismic multiple removal techniques: past, present and future: EAGE Publications bv. Education Tour Series.

Verschuur, D. J., and A. J. Berkhout, 1997, Estimation of multiple scattering by iterative inversion, Part II: Practical aspects and examples: Geophysics, **62**, 1596–1611.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177.

Yilmaz, Ö., 2001, Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data: Society of Exploration Geophysicists, volume **1** *of* Investigations in Geophysics.

Appendix **A**

Expansions of pdfs in functional bases

Except for boolean algebra (Section 1.2) there is no theory more universally employed in mathematics than linear algebra; and there is hardly any theory which is more elementary, in spite of the fact that generations of professors and textbook writers have obscured its simplicity by preposterous calculations with matrices.

Jean Dieudonné

Foundations of Modern Analysis

Suppose that one is able to write the pdf $p(x)$ of a random variable x in the following manner:

$$p(x) = \sum_{k=0}^{\infty} c_k F_k(x) \phi(x), \tag{A.1}$$

where $\phi(x) = \dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$ and the sum converges uniformly. Then one would have

$$\begin{aligned} \int_{\mathbb{R}} F_\ell(x) p(x)\, \mathrm{d}x &= \int_{\mathbb{R}} \left[F_\ell(x) \sum_{k=0}^{\infty} c_k F_k(x) \phi(x) \right] \mathrm{d}x \\ &= \sum_{k=0}^{\infty} c_k \left[\int_{\mathbb{R}} F_\ell(x) F_k(x) \phi(x)\, \mathrm{d}x \right]. \end{aligned}$$

Suppose now that $\{F_i\}_{i\in\mathbb{N}}$ is an orthogonal sequence with respect to the inner product $\langle f, g\rangle_\phi = \int_{\mathbb{R}} f(x)g(x)\phi(x)\,\mathrm{d}x$. Then

$$\begin{aligned}\int_{\mathbb{R}} F_\ell(x)F_k(x)\phi(x)\,\mathrm{d}x &= \langle F_\ell, F_k\rangle_\phi \\ &= \delta_{k\ell}.\end{aligned}$$

Consequently,

$$\begin{aligned}\int_{\mathbb{R}} F_\ell(x)p(x)\,\mathrm{d}x &= \sum_{k=0}^{\infty} c_k \left[\int_{\mathbb{R}} F_\ell(x)F_k(x)\phi(x)\,\mathrm{d}x\right] \\ &= \sum_{k=0}^{\infty} c_k\delta_{k\ell} \\ &= c_\ell.\end{aligned}$$

In conclusion, in order to obtain the coefficients c_ℓ of the expansion in Equation A.1, one simply needs to calculate the expected values $\mathrm{E}[F_\ell(x)]$. Note that the supposition of orthogonality is not restrictive, since any linearly independent set of functions can be made orthogonal by applying the Gram-Schmidt orthogonalization process (see Kreyszig, 1978, chap. 3). Therefore, for any sequence of functions, one can turn it into a orthonormal sequence.

Given a countable sequence of functions, $\{F_i\}_{i\in\mathbb{N}}$, one may make the following assumptions:

(a) $F_0(x) = 1$, so $c_0 = \mathrm{E}[1] = 1$;

(b) $F_1(x) = x$, with $c_2 = 0$, for a zero mean random variable;

(c) $F_2(x) = \frac{1}{\sqrt{2}}(x^2 - 1)$, with $c_3 = 0$, for a random variable with unit variance.

Again, these constraints are not restrictive. In a countable sequence of functions, the functions $1, x$, and x^2 can be introduced and the sequence may be reordered to let them be the first three functions. One can then apply a Gram-Schmidt orthogonalization to obtain $1, x$ and $x^2 - 1$. Finally, one may normalize them and relabel the sequence so as to have F_0, F_1 and F_2 as defined above.

Since any random variable can be redefined as to have zero mean, and renormalized to have unit variance (in case $\mathrm{E}[x^2] \neq 0$), the coefficients c_1 and c_2 can be assumed to be both zero. Using the expansion in Equation A.1, one may calculate the differential entropy of a certain random variable x.

$$\begin{aligned} h(x) &= -\int_{\mathbb{R}} p(x) \log p(x)\, \mathrm{d}x \\ &= -\int_{\mathbb{R}} \phi(x) \left(1 + \sum_{k=3}^{\infty} c_k F_k(x)\right) \log \left[\phi(x) \cdot \left(1 + \sum_{k=3}^{\infty} c_k F_k(x)\right)\right] \mathrm{d}x \end{aligned} \tag{A.2}$$

Letting $\varepsilon = \sum_{k=3}^{\infty} c_k F_k(x)$, Equation A.2 can be more succinctly developed.

$$\begin{aligned} h(x) &= -\int_{\mathbb{R}} \phi(x) \cdot (1+\varepsilon) \cdot [\log \phi(x) + \log(1+\varepsilon)]\, \mathrm{d}x \\ &= -\int_{\mathbb{R}} \phi(x) \log \phi(x) + \phi(x)\varepsilon \log \phi(x) + \phi(x)(1+\varepsilon)\log(1+\varepsilon)\, \mathrm{d}x \end{aligned} \tag{A.3}$$

Observe that $-\int_{\mathbb{R}} \phi(x) \log \phi(x) = h(x_G)$, where x_G is a Gaussian random variable with unit variance. The second term in Equation A.3 can be shown to equal zero. In fact,

$$\begin{aligned} -\int_{\mathbb{R}} \phi(x)\varepsilon \log \phi(x)\, \mathrm{d}x &= -\int_{\mathbb{R}} \phi(x) \left[\sum_{k=3}^{\infty} c_k F_k(x)\right] \cdot \left[-\frac{\log(2\pi)}{2} - x^2 \frac{\log(e)}{2}\right] \mathrm{d}x \\ &= \sum_{k=3}^{\infty} \left[\frac{\log(2\pi)}{2} c_k \int_{\mathbb{R}} F_k(x)\phi(x)\, \mathrm{d}x + \frac{\log(e)}{2} c_k \int_{\mathbb{R}} x^2 F_k(x)\phi(x)\, \mathrm{d}x\right] \\ &= \sum_{k=3}^{\infty} \left[\frac{\log(2\pi)}{2} c_k \langle 1, F_k\rangle_\phi + \frac{\log(e)}{2} c_k \langle x^2, F_k\rangle_\phi\right] \\ &= 0, \text{since } F_k \text{ is orthogonal to polynomials of degree} \leq 2. \end{aligned}$$

Therefore,

$$h(x) = -\int_{\mathbb{R}} \phi(x) \log \phi(x) + \phi(x)(1+\varepsilon)\log(1+\varepsilon)\, \mathrm{d}x.$$

Note also that, by a Taylor expansion, $(1+\varepsilon)\log(1+\varepsilon) = \varepsilon + \varepsilon^2/2 + O(\varepsilon^2)$.

$$
\begin{aligned}
-\int_{\mathbb{R}} \phi(x)(1+\varepsilon)\log(1+\varepsilon)\,\mathrm{d}x &= -\int_{\mathbb{R}} \phi(x)(\varepsilon + \varepsilon^2/2 + O(\varepsilon^2))\,\mathrm{d}x \\
&= -\int_{\mathbb{R}} \phi(x)\left[\sum_{k=3}^{\infty} c_k F_k(x) + \frac{1}{2}\left(\sum_{k=3}^{\infty} c_k F_k(x)\right)^2 + O(\varepsilon)\right]\mathrm{d}x \\
&= -\int_{\mathbb{R}} \phi(x)\left[\sum_{k=3}^{\infty} c_k F_k(x) + \frac{1}{2}\sum_{k=3}^{\infty} c_k^2 F_k(x)^2 - \sum_{\substack{k=3 \\ \ell=4 \\ k\neq\ell}}^{\infty} c_k c_\ell F_k F_\ell + O(\varepsilon)\right]\mathrm{d}x \\
&= -\sum_{k=3}^{\infty} c_k \langle 1, F_k\rangle_\phi + \sum_{\substack{k=3 \\ \ell=4 \\ k\neq\ell}}^{\infty} c_k c_\ell \langle F_k, F_\ell\rangle_\phi - \frac{1}{2}\sum_{k=3}^{\infty} c_k^2 \langle F_k, F_k\rangle_\phi - \int_{\mathbb{R}} \phi(x) O(\varepsilon)\,\mathrm{d}x \\
&= -\frac{1}{2}\sum_{k=3}^{\infty} c_k^2 - \int_{\mathbb{R}} \phi(x) O(\varepsilon)\,\mathrm{d}x, \text{since } \{F_i\}_{i\in\mathbb{N}} \text{ is an orthonormal sequence.}
\end{aligned}
$$

Hence, $h(x)$ may be approximated as

$$
h(x) \approx h(x_G) - \frac{1}{2}\sum_{k=3}^{\infty} c_k^2, \tag{A.4}
$$

which is obtained disregarding the terms $O(\varepsilon)$. This in turn, leads to

$$
J(x) \approx \frac{1}{2}\sum_{k=3}^{r} c_k^2, \tag{A.5}
$$

which is obtained substituting Equation A.4 in the definition of negentropy (Equation 4.10) and truncating to r terms the infinite sum of c_k^2.

Example 1 (Hermite polynomials). *Let $\{H_i\}_{i\in\mathbb{N}}$ be the Hermite polynomials. These polynomials are obtained applying the Gram-Schmidt procedure with respect to the internal product $\langle\cdot,\cdot\rangle_\phi$ to the sequence $\{x^i\}_{i\in\mathbb{N}}$. The first five (normalized) Hermite polynomials are (Abramowitz and Stegun, 1972, pg. 775):*

$$H_0(x) = 1 \qquad H_3(x) = \frac{1}{\sqrt{6}}(x^3 - 3x)$$
$$H_1(x) = x \qquad H_4(x) = \frac{1}{\sqrt{24}}(x^4 - 6x^2 + 3)$$
$$H_2(x) = \frac{1}{\sqrt{2}}(x^2 - 1)$$

Note that no modifications must be made to the order of the sequence, since the first three functions are exactly those required by the assumptions. Therefore, in this case,

$$J(x) \approx \frac{c_3^2 + c_4^2}{2},$$

where $c_3 = \mathrm{E}[H_3(x)]$ and $c_4 = \mathrm{E}[H_4(x)]$. Therefore,

$$\begin{aligned}
\mathrm{E}[H_3(x)] &= \mathrm{E}\left[\frac{1}{\sqrt{6}}(x^3 - 3x)\right] = \frac{1}{\sqrt{6}}(\mathrm{E}[x^3] - 3\,\mathrm{E}[x]) \\
&= \frac{\mathrm{E}[x^3]}{\sqrt{6}}, \textit{since } \mathrm{E}[x] = 0; \\
\mathrm{E}[H_4(x)] &= \mathrm{E}\left[\frac{1}{\sqrt{24}}(x^4 - 6x^2 + 3)\right] = \frac{\mathrm{E}[x^4] - 6 + 3}{\sqrt{6}}, \textit{since } \mathrm{E}[x] = 1 \\
&= \frac{\mathrm{kurt}[x]}{\sqrt{6}}, \textit{since } \mathrm{kurt}[x] = \mathrm{E}[x^4] - 3,
\end{aligned}$$

and finally, one obtains $J(x) \approx \dfrac{\mathrm{E}[x^3]^2}{12} + \dfrac{\mathrm{kurt}[x]^2}{48}$.