



FEDERAL UNIVERSITY OF PARÁ
GEOSCIENCES INSTITUTE
GRADUATE PROGRAM IN GEOPHYSICS

THESIS

Structural constraints for image-based inversion methods

JONATHAS DA SILVA MACIEL

BELÉM-PARÁ

2016

JONATHAS DA SILVA MACIEL

Structural constraints for image-based inversion methods

Thesis presented to Graduate Program in Geophysics of the Geosciences Institute at Federal University of Pará, in fulfillment of the requirements for obtaining the title of Doctor in Geophysics.

Supervisor: Prof. Dr. Jessé Carvalho Costa

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da UFPA

Maciel, Jonathas da Silva, 1984-
Structural constraints for image-based inversion
methods / Jonathas da Silva Maciel. - 2016.

Orientador: Jessé Carvalho Costa.
Tese (Doutorado) - Universidade Federal do
Pará, Instituto de Geociências, Programa de
Pós-Graduação em Geofísica, Belém, 2016.

1. Geofísica. 2. Equação de onda - Modelos
matemáticos. 3. Método de análise de velocidade.
4. Processamento de imagens - Morfologia. I.
Título.

CDD 22. ed. 550

JONATHAS DA SILVA MACIEL

Structural constraints for image-based inversion methods

Tese apresentada ao Programa de Pós-Graduação em Geofísica do Instituto de Geociências da Universidade Federal do Pará, em cumprimento às exigências para obtenção do título de Doutor em Geofísica.

Data de Aprovação: 22/04/2016

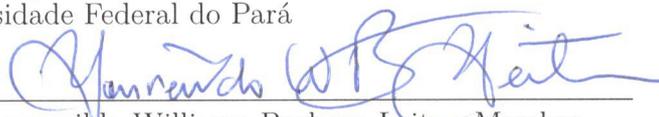
Conceito:

EXC (9,0)

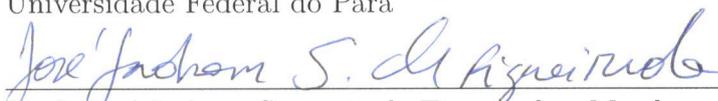
Banca Examinadora:



Prof. Jessé Carvalho Costa - Orientador
Doutor em Geofísica
Universidade Federal do Pará



Prof. Lourenildo Willame Barbosa Leite - Membro
Doutor em Geofísica
Universidade Federal do Pará



Prof. José Jadsom Sampaio de Figueiredo - Membro
Doutor em Geofísica
Universidade Federal do Pará



Prof. Jörg Dietrich Wilhelm Schleicher - Membro
Doutor em Geofísica
Universidade Estadual de Campinas



Prof^a. Maria Amélia Novais Schleicher - Membro
Doutora em Matemática Aplicada
Universidade Estadual de Campinas

I dedicate this work to my family.

ACKNOWLEDGMENTS

I would to thank my supervisor professor Dr. Jessé Carvalho Costa by the reliance, knowledge and experience passed to me, of which will be useful throughout of my professional life.

The current and previous coordinator of Graduate Program in Geophysics (GPG), Dr. Cicero Régis and Dra. Ellen Gomes, for the support and encouragement of participation of *Programa de Doutorado Sanduíche no Exterior* (CAPES/PDSE). Also the secretaries: Benildes Lopes, Lucibela Cardias and Suelen Buzaglo for transforming my academic life less complicated.

To my friends: Andrei Oliveira for technical support at GPG's computers, Carlos Alexandre Costa, Francisco Neto, Josafat Cardoso, Wildney Vieira and Williams Lima by technical discussions about inversion methods.

I would like to thank the professor Dr. Eric Verschuur and DELPHI consortium, by opportunity of working as a visiting researcher at Delft University of Technology, Kingdom of the Netherlands, as well as the secretary Margaret van Fessem by throughout aid in Delft.

I am grateful for the financial support provided by the *Compagnie Générale de Géophysique* (CGG Brazil), through the cooperation agreement for the training of doctors in geophysics, applied to exploration and production of hydrocarbons (CGG-UFPA). I also thank the MEC/CAPES by participation in PDSE program. Thank to *Instituto Nacional de Ciência e Tecnologia - Geofísica do Petróleo* (INCT-GP) by the travel allowance in international conferences in geophysics.

To committee members by the availability and suggestions that enrich this work.

And especially to my family: my father Elias Maciel (*in memoriam*) and my mother Amalia Maciel, which, with a powerful world vision, built a peaceful and stable home, producing a perfect environment for my complete dedication to studies. My siblings Eliseu Maciel and Késia Maciel, whose union has made us strong and gave us many victories.

*"The true sign of intelligence is not knowledge but imagination."
(Albert Einstein)*

RESUMO

Esta tese apresenta duas metodologias de regularização estrutural para os métodos de análise de velocidade com migração e inversão conjunta com migração: regularização gradiente cruzado e filtragem com operadores morfológicos. Na análise de velocidade com migração, a regularização de gradiente cruzado tem como objetivo vincular os contrastes de velocidade com o mapa de refletividade, através da paralelização dos vetores gradiente de velocidade com os vetores gradiente da imagem. Propõe-se uma versão com gradiente cruzado das funções objeto de minimização: *Differential Semblance*, *Stack Power* e *Partial Stack Power*. Combina-se a função *Partial Stack Power* com sua versão de gradiente cruzados, com o objetivo de aumentar gradativamente a resolução do modelo de velocidade, sem comprometer o ajuste das componentes de longo comprimento de onda do modelo de velocidade. Na inversão conjunta com migração propõe-se aplicar os operadores morfológicos de erosão e dilatação, no pré-condicionamento do modelo de velocidade em cada iteração. Os operadores usam o mapa de refletividade para delimitar as regiões com mesmo valor de propriedade física. Eles homogenizam a camada geológica e acentuam o contraste de velocidade nas bordas. Os vínculos estruturais não apenas irão reduzir a ambiguidade na estimativa do modelo de velocidade, mas tornará os métodos de inversão com migração mais estáveis, reduzindo artefatos, delineando soluções geologicamente plausíveis e acelerando a convergência da função objeto de minimização.

Palavras-chave: Imagem migrada. Vínculo estrutural. Análise de velocidade com equação de onda. Inversão conjunta com migração. Morfologia. Modelagem da equação de onda.

ABSTRACT

This thesis presents two methodologies of structural regularization for Wave-Equation Migration Velocity Analysis and Joint Migration Inversion: cross-gradient regularization and filtering with morphological operators. In Wave-Equation Migration Velocity Analysis, the cross-gradient regularization aims to constrain the velocity contrasts with the reflectivity map by parallelization of the velocity gradient vector and the image gradient vector. We propose a version with cross-gradient of the objective functions: Differential Semblance, Stack Power and Partial Stack Power. We combine the Partial Stack Power with its version of cross-gradient, in order to gradually increase the resolution of the velocity model without compromising the adjustment of the long wavelengths of the velocity model. In Joint Migration Inversion, we propose to apply morphological operators of erosion and dilation in the preconditioning of the velocity model in each iteration. Operators use the reflectivity map to mark the regions with the same value of physical property. They homogenize the geological layer and accentuate the velocity contrast at the edges. Structural constraints do not only reduce the ambiguity in estimating a velocity model, but also make the migration/inversion methods more stable, reducing artifacts, delineating geologically plausible solutions, and accelerating the convergence of the objective function.

Keywords: Migrated image. Structural constraint. Wave-Equation Migration Velocity Analysis. Joint Migration Inversion. Morphology. Wave-equation modeling.

LIST OF ILLUSTRATIONS

Figure 2.1 – Coordinate system.	22
Figure 2.2 – Reverse Time Migration (RTM) process: cross-correlation of forward wavefield (a) with backward wavefield (b), producing the image (c).	23
Figure 2.3 – ODCIG migrated with velocity values: (a) 20% greater, (b) exact and (c) 20% smaller.	23
Figure 2.4 – ADCIG migrated with velocity values: (a) 20% greater, (b) exact and (c) 20% smaller.	24
Figure 2.5 – Stack Power (SP), Differential Semblance (DS) and Partial Stack Power (PSP) penalty operators, in ADCIG (blue) and ODCIG (red) domains.	26
Figure 2.6 – The arrows show the eigenvalues of 2.9. Higher weights along structures than across them, and equal weights inside the layers.	27
Figure 2.7 – A cross-gradient regularization parallelize the gradient vectors ∇c and ∇I , minimizing (maximize) their outer (inner) product in a least-square sense.	28
Figure 2.8 – Gaussian window changing the phase analysis (large) to amplitude analysis (squeezed). Note the delay between the Partial Stack Power (continuous lines) and Cross-Gradient Partial Stack Power (dashed lines), which is reduced when the background is adjusted.	29
Figure 2.9 – Velocity model.	32
Figure 2.10 – Migrated image with:	33
Figure 2.11 – ODCIG migrated with true velocity.	34
Figure 2.12 – ODCIG migrated with initial velocity.	35
Figure 2.13 – ODCIG migrated with optimized Partial Stack Power velocity.	36
Figure 2.14 – ODCIG migrated with optimized Cross-Gradient Partial Stack Power velocity.	37
Figure 2.15 – ODCIG zero offset image traces $x_1 = (3000, 3600, 4200, 4800, 5400)$ m: true migration (blue), Partial Stack Power migration (red) and Cross-Gradient Partial Stack Power (black).	38
Figure 2.16 – ODCIG zero offset image traces $x_1 = (6000, 6600, 7200, 7800, 8400)$ m: true migration (blue), Partial Stack Power migration (red) and Cross-Gradient Partial Stack Power (black).	39
Figure 3.1 – Schematic representation of incoming and outgoing wavefields acting for each grid-point of depth level x_1^m	42

Figure 4.1	– Original image (a) and the results of binary erosion with (b) a square and (c) a circle structural element. The origin position of the structural element is represented as a dot, and the colors blue and red represent the match and unmatched criteria of equation (4.1), respectively. The binary eroded image (black) is created by setting the foreground pixels in all translations for which B is totally inside of A (i.e the gray area).	50
Figure 4.2	– Original image (a) and the result of binary dilation with (b) a square and (c) a circle structural element. The origin position of the structural element is represented as a dot, and the colors blue and red represent the match and unmatched criteria of equation (4.2), respectively. The binary dilation image (gray) is created by setting the foreground pixels in all translations for which B touches A (black).	50
Figure 4.3	– Erosion and dilation in a 1D gray-scale function. (a) Gray-scale function $f(\mathbf{x})$ and flat null structural function $g(\mathbf{x})$. The dot represent the origin of the $g(\mathbf{x})$'s coordinate system. (b) Erosion (blue) and (c) dilation (red).	52
Figure 4.4	– 2D morphological operators applied to a gray-scale image.	52
Figure 4.5	– Step formation: $f(x)$ masked (black), (a) Erosion (blue) and (b) dilation (red).	53
Figure 4.6	– 1D morphological operators applied to function $f(x) = \sin(x)$ (black curve): (a) Standard erosion (blue) and dilation (red) and the average function (magenta), for $k = 10$ (dotted line), $k = 30$ (dashed line) and $k = 100$ (continuous line). (b) Constrained curves $f(x)/H(x)$ (blue) and $f(x)H(x)$ (red). (c) Structure-oriented erosion and dilation with the same values of k as in (a).	55
Figure 4.7	– Structure-oriented morphological operator example in 2D. (a) True model, (b) true reflectivity, (c) mask function $M(\mathbf{x})$, (d) initial model $f(\mathbf{x})$, (e) eroded model $(f \ominus g)^{60}(\mathbf{x})$, (f) dilated model $(f \oplus g)^{60}(\mathbf{x})$ and (g) the final solution $A(\mathbf{x})$. Color bars are in m/s except the reflectivity model which is dimensionless.	56
Figure 4.8	– Subsurface model for JMI test. (a) True velocity model, (b) true reflectivity, (c) initial velocity model and (d) initial reflectivity. Color bars are in m/s except the reflectivity model which is dimensionless.	58
Figure 4.9	– Standard JMI results. (a) Reflectivity; (b) velocity; (c) profiles at $x = 1000$ m: velocity (left), true velocity (blue), initial velocity (green) and final velocity (red); reflectivity (right), true reflectivity (blue) and final reflectivity (red); (d) measured data (left) and the residual data (right). Color bars are in m/s except the reflectivity model which is dimensionless.	59

Figure 4.10 – JMI results with morphological constraints. (a) Reflectivity; (b) velocity; (c) profiles at $x = 1000$ m: velocity (left), true velocity (blue), initial velocity (green) and final velocity (red); reflectivity (right), true reflectivity (blue) and final reflectivity (red); (d) measured data (left), residual data (right). Color bars are in m/s except the reflectivity model which is dimensionless.	60
Figure 4.11 – Subsurface model for JMI test. (a) True velocity model, (b) True reflectivity, (c) initial velocity model and (d) initial reflectivity. Color bars are in m/s except the reflectivity model which is dimensionless.	62
Figure 4.12 – Standard JMI results. (a) Reflectivity; (b) velocity; (c) profiles at $x = 900$ m: velocity (left), true velocity (blue), initial velocity (green) and final velocity (red); reflectivity (right), true reflectivity (blue) and final reflectivity (red); (d) measured data (left), residual data (right). Color bars are in m/s except the reflectivity model which is dimensionless.	63
Figure 4.13 – JMI results with morphological constraints. (a) Reflectivity; (b) velocity; (c) profiles at $x = 900$ m: velocity (left), true velocity (blue), initial velocity (green) and final velocity (red); reflectivity (right), true reflectivity (blue) and final reflectivity (red); (d) measured data (left), residual data (right). Color bars are in m/s except the reflectivity model which is dimensionless.	64
Figure D.1 – 2D Erosion algorithm.	81
Figure D.2 – 2D Dilation algorithm.	82

LIST OF TABLES

Table 2.1 – WEMVA objective functions: Stack Power and Differential Semblance for angle domain (blue) and offset domain (red).	24
Table 2.2 – WEMVA objective functions constrained by cross-gradient regularization: Cross-Gradient Stack Power (CGSP) and Cross-Gradient Differential Semblance (CGDS), for angle domain (blue) and offset domain (red).	28
Table D.1 – The algorithm’s main parameters.	81

LIST OF ABBREVIATIONS AND ACRONYMS

ADCIG Angle Domain Common Image Gather

CGDS Cross-Gradient Differential Semblance

CGPSP Cross-Gradient Partial Stack Power

CGSP Cross-Gradient Stack Power

CIG Common Image Gather

DS Differential Semblance

FWI Full Waveform Inversion

FWM Full Wavefield Migration

FWMod Full Wavefield Modeling

JMI Joint Migration Inversion

ODCIG Offset Domain Common Image Gather

PSP Partial Stack Power

RTM Reverse Time Migration

SP Stack Power

WEMVA Wave-Equation Migration Velocity Analysis

SUMMARY

1	INTRODUCTION	16
1.1	Motivation	16
1.2	WEMVA with cross-gradient regularization	17
1.3	JMI method	18
1.4	High resolution velocity estimation with morphological operators . .	18
1.5	General conclusions and appendices	19
2	WEMVA WITH STRUCTURAL PRIOR INFORMATION	20
2.1	Introduction	20
2.2	WEMVA theory	22
2.3	Constraining the velocity model with the reflector map	26
2.4	Numerical experiments	30
2.5	Conclusions	40
3	JOINT MIGRATION INVERSION	41
3.1	Full Wavefield Modeling	41
3.2	Velocity perturbations	44
3.3	Reflectivity perturbations	45
3.4	Inversion	45
4	ENHANCING RESOLUTION IN IMAGE-BASED VELOCITY ESTIMATION USING MORPHOLOGICAL OPERATORS	48
4.1	Erosion and dilation for binary images	49
4.2	Erosion and dilation for gray-scale images	51
4.3	Enhancing the sharpness with modified erosion and dilation operators	53
4.4	Application of morphological operators in the JMI method	57
4.4.1	2D simple model with strong internal multiples	57
4.4.2	2D complex model with strong internal multiples	61
4.5	Conclusion	65
5	GENERAL CONCLUSIONS	66
	Bibliography	68
	APPENDICES	71
	APPENDIX A – OBJECTIVE FUNCTION GRADIENT	72

APPENDIX B – QUASI-NEWTON METHOD	77
APPENDIX C – B-SPLINES INTERPOLATION	80
APPENDIX D – EROSION AND DILATION PSEUDO CODE . . .	81

1 INTRODUCTION

1.1 Motivation

Tomographic methods based on ray tracing are well established in the oil industry as the main tool for velocity model building. The velocity model is estimated in order to optimize focusing of selected reflection events. As a consequence tomographic velocity models produce good images of the subsurface, but require intense human interaction for event selection and editing. Another well known limitation of tomographic methods results from the high-frequency approximation to describe wave propagation, which limits these methods to regions of mild geological complexity. Currently there are several efforts to develop velocity model building methods that do not have the limitations of ray methods and can take advantage of more accurate descriptions of the wavefield.

Wave-equation based inversion can be divided into three branches: inversion methods with objective function specified in the data domain as Full Waveform Inversion [FWI, (VIRIEUX; OPERTO, 2009)], inversion methods with objective function defined in the image domain as Wave-Equation Migration Velocity Analysis [WEMVA, (SAVA; BIONDI, 2004)] and hybrid inversion methods like Joint Migration Inversion [JMI, (BERKHOUT, 2014a; BERKHOUT, 2014b; BERKHOUT, 2014c)]. In the latter approach, the model parameterization is extended to include a background velocity and a reflectivity model. The velocity model is estimated from fitting the wavefield spectrum and reflectivity from least-squares migration. The objective function in FWI minimizes the difference between the modeled and recorded traces. WEMVA uses migration methods to generate Common Image Gathers (CIG), where an objective function will measure the focusing. JMI, like FWI, minimizes the differences between modeled data and recorded data, but differently from FWI, JMI separates the wavefield-phase fitting, to estimate the velocity model, and the wavefield-amplitude fitting to estimate the reflectivity of subsurface.

FWI has gained much attention recently because the method estimates velocity models with high resolution while possibly reducing the human effort in the velocity model building workflow. However, FWI has some challenges: the estimation of the source signature from the dataset, a strong dependence on an initial model and the requirement of large offsets to update the long wavelengths components of velocity model. WEMVA and JMI can produce better models than those obtained using ray tomography, although with low resolution when compared to successful applications of FWI. Additionally, WEMVA and JMI are more robust with respect to data quality and dependence on the initial model, and they can use more conventional data acquisition geometries to update the velocity field.

WEMVA and JMI, like many other non-linear inverse-problem methods, need preconditioning or regularization to ensure stability and reduce the ambiguity in inversion. Traditionally smoothness constraints, such as Tikhonov regularization (TIKHONOV; ARSENIN, 1977; KAPIO et al., 1999; WEIBULL; ARNTSEN, 2013), are used to stabilize inversion in wave-equation velocity-model building. WEMVA and JMI have in common the estimation of a migrated image or reflectivity model during the inversion. These images can be used as prior information to increase the resolution of the velocity model.

This thesis aims at developing techniques to increase the velocity model resolution of inversion methods in the image domain. The techniques of preconditioning and regularization do not only reduce the ambiguity of optimized solutions, but, as suggested by Williamson et al. (2011), the use of structural regularization also adds geologically meaningful information to the model. Moreover, preconditioning also speeds up the convergence of the method. Another aspect I explore is the variation of the regularization parameter during the iterations. At initial iterations smoothness constraints are enforced to help the recovering of long wavelength components of the velocity model. Structural information is gradually enforced at later iterations to avoid the inversion to be trapped prematurely in local minima far from the optimal solution. The following sections present the regularization techniques based in structural constraints and the organization of this thesis.

1.2 WEMVA with cross-gradient regularization

Chapter 2 describes the WEMVA method using reverse time migration (RTM). It revises the Claerbout's image condition (CLAERBOUT, 1971) and its extension to velocity analysis with the construction of common image gathers (CIG), in angle or subsurface offset domain (BIONDI; SYMES, 2004; RICKETT; SAVA, 2002; SAVA; FOMEL, 2006; YANG; SAVA, 2010). We show that the objective function traditionally used, Differential Semblance, has few local minima, but does not increase coherence for near offsets. To overcome this lack of information many authors use the concept of Stack Power or Semblance (CHAVENT; JACEWITZ, 1995; WEIBULL; ARNTSEN, 2013). However, this type of function suffers from cycle skipping and present many local minima. In order to avoid this, Zhang and Shan (2013) proposed combining the convergence properties of Differential Semblance with coherence sensitivity of Stack Power in a single objective function called Partial Stack Power. This avoids the formulation of Stack Power as a regularizing functional of the Differential Semblance. I also explore the cross-gradient (FREGOSO; GALLARDO, 2009; TRYGGVASON; LINDE, 2006) of the velocity field and the image field as regularization. The cross-gradient is a type of generalized Tikhonov regularization (KAPIO et al., 1999). In this work, the cross-gradient enforces velocity

contrast normal to reflectors, by aligning the velocity gradient vector with the migrated image-gradient vector in a least-square sense. The product of Partial Stack Power with cross-gradient regularization is proposed to enforce structural information. Using this multiplicative objective function, the structural regularization can be introduced gradually without compromising convergence. This new formulation of WEMVA is validated on Marmousoft data set.

1.3 JMI method

Chapter 3 presents the JMI method for isotropic media with no dependence on angle for the reflectivity model (STAAL; VERSCHUUR, 2014). JMI uses the one-way wave equations to separately modeling downgoing and upgoing wavefields. This feature enables a greater control over multiples, allowing JMI to generate, through least-squares migration, high-resolution reflectivity maps without preprocessing to attenuate multiples in the data. Another positive feature of one-way modeling is the separation of phase and amplitude of the wavefield. The velocity estimation is correlated with the perturbation of the wavefield extrapolation phase-shift operator, while the reflectivity model is estimated from wavefield amplitudes. This approach makes the problem much better conditioned. The chapter finalizes presenting the derivation of velocity and reflectivity gradients used in the optimization and the description of the JMI inversion algorithm.

1.4 High resolution velocity estimation with morphological operators

Chapter 4 presents a new strategy to constraint the velocity model with structural information derived from migrated images: preconditioning by morphological operators (MARAGOS, 2009). The operators of erosion and dilation, based on the theory of Mathematical Morphology (DOUGHERTY; LOTUFO, 2003), are initially presented for binary images and then extended to gray-scale images. This work introduces a modification of erosion and dilation operators by division and product of the current model with a mask function derived from the migrated image. The modified operators can build homogeneous layers with strong contrast at its borders. Initially, I illustrate the effects of the new morphological operator in simple in 1D and 2D models. Then, I present the numerical experiments applying the modified morphological operators to JMI for two synthetic data sets with different degrees of complexity. The proposed algorithm recovered high resolution velocity models and the reflectivity models in both cases.

1.5 General conclusions and appendices

The last chapter presents the general conclusions about the methods and techniques developed in this thesis. I also added five appendices which present mathematical derivations and definitions. In Appendix A the gradient of the objective function is derived using the adjoint state method. Appendix B describes the quasi-Newton optimization method and its update formula. Appendix C shows the B-splines interpolation function and, finally, Appendix D, details the pseudo code for the erosion and dilation operators applied in Chapter 4.

2 WEMVA WITH STRUCTURAL PRIOR INFORMATION

Wave-Equation Migration Velocity Analysis (WEMVA), like many other geophysical inversion methods, needs extra information to reduce the ambiguity in its solutions. However, unlike other methods, WEMVA has the advantage of having at its disposal the migrated image, from where we can extract structural information, allowing to get a well fitted model with subsurface geology. Using the concept of cross-gradient, this work proposed to couple the WEMVA objective function called Partial Stack Power, which combines the Differential Semblance and Stack Power, with its analogous version with cross-gradient, to precondition WEMVA convergence and increase the velocity model information.

2.1 Introduction

Methods using wave-equation modeling represent the best tools to image regions with complex geology, because they do not suffer from limitations in wave simulation, including information like refraction, diffraction, primaries and multiples. The evaluation of all these information is well computed in inversion methods in the data domain like Full Waveform Inversion [FWI, (VIRIEUX; OPERTO, 2009)]. FWI is gaining importance by producing better and detailed velocity models. Unfortunately FWI hardly converges if the low frequency, inside of the data and inside of the model, are not well fitted. However Wave-Equation Migration Velocity Analysis [WEMVA, (SAVA; BIONDI, 2004)], an image based inversion method, has less sensitivity to the initial model, also has the advantage of working with the frequency band conventionally acquired in seismic experiments, and it does not require large offsets to estimate the model in deeper regions as does FWI.

Improve the WEMVA method is a valid objective to enrich the toolbox of inversion methods in geophysics. For this purpose, it is necessary to highlight one the main negative features of the WEMVA method: the presence of strong artifacts in the objective function gradient (FEI; WILLIAMSON, 2010). These artifacts usually dominate the useful gradient and increase the optimization time.

Fei and Williamson (2010) saw that Differential Semblance (DS) (SHEN; SYMES, 2008) produces a 90° phase shift in its gradient, resulting in near-vertical stripe artifacts. To remove the artifacts, Fei and Williamson (2010) proposed to use a derivative of the residual image (LAMELOISE; CHAURIS; NOBLE, 2015; SHEN; SYMES, 2015). Besides that, as indicated by Shen and Symes (2008), Differential Semblance is limited and doesn't

have cover all offsets, in this case the zero-offset amplitude maximization is not evaluated. The Stack Power (SP) function can be included in Differential Semblance optimization to supply the zero-offset amplitude correction (CHAVENT; JACEWITZ, 1995; WEIBULL; ARNTSEN, 2013).

Zhang and Shan (2013) noted that the difference between Differential Semblance and Stack Power, is the kind of weight function that is convolved with the Common Image Gather (CIG). In the angle domain Differential Semblance is equivalent to the convolution of finite-difference operator with the CIG, while Stack Power represents the convolution of the CIG with the constant 1. A Radon transform (Fourier transform for plane waves) changes these convolutions to a multiplication with the offset-vector norm, $|\mathbf{h}|$, and the Dirac delta $\delta(\mathbf{h})$, respectively. This analysis suggests the possibility to evaluate Stack Power locally like Differential Semblance (MACIEL; COSTA; SCHLEICHER, 2012). The idea is to stack the image traces in a small Gaussian window, bringing to Stack Power better convergence proprieties as Differential Semblance, with advantage to compute the zero-offset amplitude residual. They named this new objective function as Partial Stack Power (PSP).

There is additional zero offset information which can help to conditioning WEMVA. It is the application of structure-oriented filters (FEHMERS; HÖCKER, 2003). Using an approximation of the anisotropic diffusion equation, as shown by Hale (2009), Williamson et al. (2011) smoothed the objective function gradient. They showed that in this way the velocities models recovered the main features of true velocity. Moreover this procedure accelerated the method's convergence. However it implied in resolving one more linear system per iteration.

It is also possible to obtain a structural-oriented velocity model, with less computational cost, by the concept of cross-gradient regularization (FREGOSO; GALLARDO, 2009; TRYGGVASON; LINDE, 2006). A cross gradient aims at aligning the gradient vectors of two parameters giving them the same structural map. Here, I propose an alternative strategy to couple the cross-gradient in WEMVA optimization, avoiding the regularization parameter adjustment and the solution of another linear system. It combines the Partial Stack Power function with its analogous version with cross-gradient, allowing to adjust the weight functions, to turn on the extra information when the velocity background is fitted well. I start at reviewing the WEMVA inversion method, then I show the cross-gradient regularization for WEMVA and finally I validate the algorithm by application to synthetic Marmousoft model (BILLETTE et al., 2003).

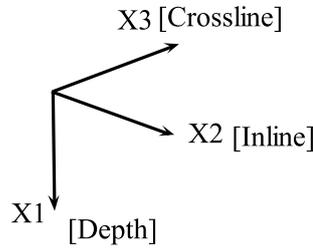
2.2 WEMVA theory

In this work, I assume that wave propagation is described by the acoustic equation with constant density,

$$\square P(\mathbf{x}, t) \equiv \left[\frac{1}{c^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} - \nabla^2 \right] P(\mathbf{x}, t) = F(\mathbf{x}, t), \quad (2.1)$$

where \square is the wave equation operator, $c(\mathbf{x})$ is the wave velocity, $P(\mathbf{x}, t)$ the pressure field, $F(\mathbf{x}, t)$ the seismic source, $\mathbf{x} = (x_1, x_2, x_3)$ a subsurface position (Figure 2.1) and t the time instant.

Figure 2.1 – Coordinate system.



Source: By the author

According to Claerbout (1971), a migrated image, $I(\mathbf{x})$, is made by the cross-correlation between the downward wave field $D(\mathbf{x}, t, s)$ (source field) with the upward wave field $U(\mathbf{x}, t, r)$ (receiver field), modeled with recorded traces injected from the last sample to the first,

$$I(\mathbf{x}) = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r(s)} \int_0^T dt D(\mathbf{x}, t, s) U(\mathbf{x}, t, r), \quad (2.2)$$

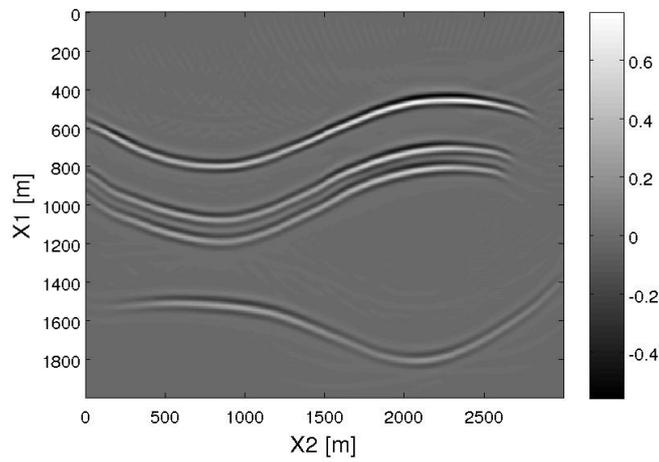
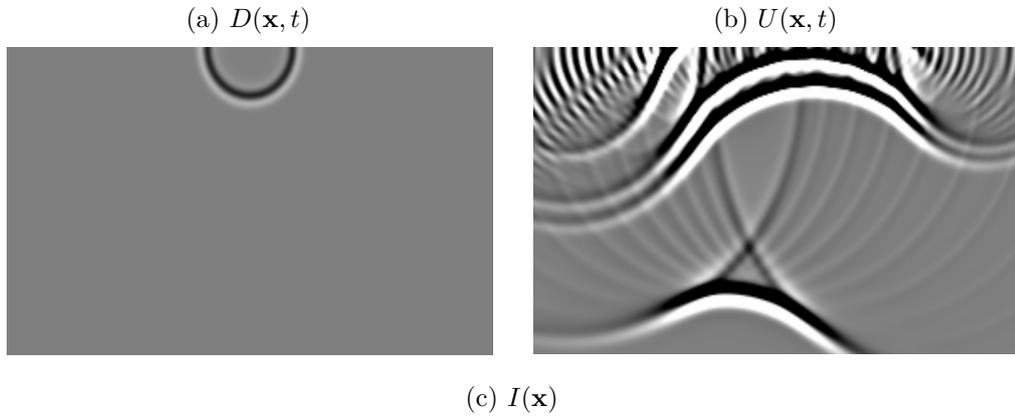
where N_s is the number of shots, $N_r(s)$ is the number of receivers per shot and T is the total recording time, see Figure 2.2.

If the velocity model is wrong, the Claerbout image condition, equation 2.2, will focus the energy in a wrong position. That energy can be related to a displacement of source and receiver in space (RICKETT; SAVA, 2002), in time (SAVA; FOMEL, 2006) or in space-time (YANG; SAVA, 2010). The collection of all these information is called the Offset Domain Common Image Gather (ODCIG),

$$I(\mathbf{x}, \mathbf{h}, \tau) = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r(s)} \int_0^T dt D(\mathbf{x} - \mathbf{h}, t - \tau, s) U(\mathbf{x} + \mathbf{h}, t + \tau, r), \quad (2.3)$$

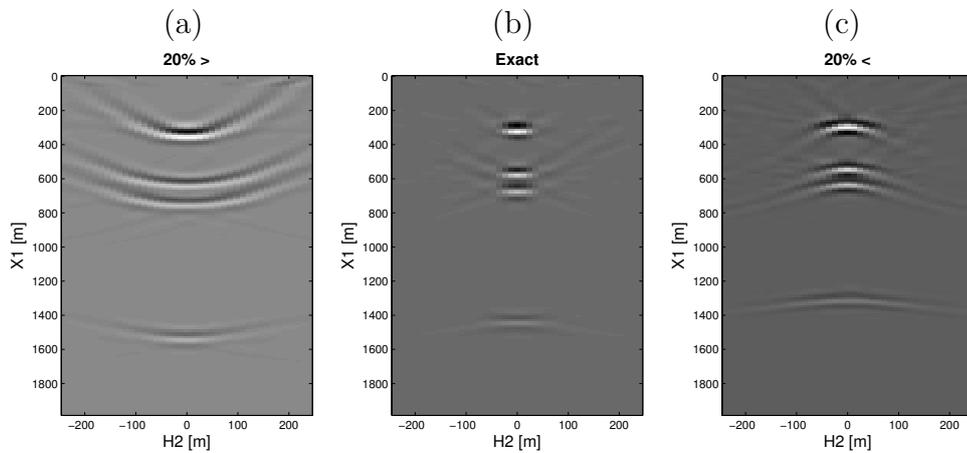
where $\mathbf{h} = (h_1, h_2, h_3)$ is the sub-surface offset and τ is the time shift. However, using the complete extension of the image condition represents a significant increasing of the computational cost (SAVA; VASCONCELOS, 2011). Usually only a horizontal space shift [$\mathbf{h} = (0, h_2, 0); \tau = 0$] is enough to estimate good 2D background velocities, see Figure 2.3.

Figure 2.2 – Reverse Time Migration (RTM) process: cross-correlation of forward wavefield (a) with backward wavefield (b), producing the image (c).



Source: By the author

Figure 2.3 – ODCIG migrated with velocity values: (a) 20% greater, (b) exact and (c) 20% smaller.

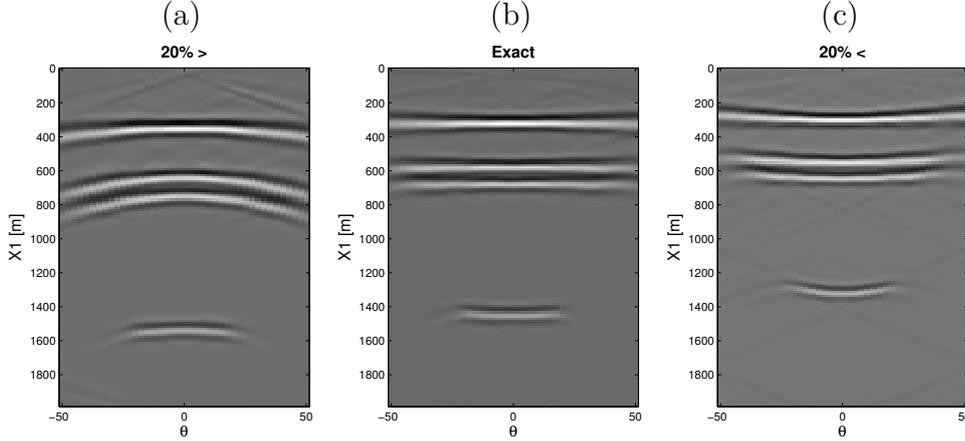


Source: By the author

The energy will be focused at zero offset, when the velocity is correct. A Radon

transform applied to this gather produces a flat horizontal event in another domain called Angle Domain Common Image Gather (ADCIG) (BIONDI; SYMES, 2004), see Figure 2.4.

Figure 2.4 – ADCIG migrated with velocity values: (a) 20% greater, (b) exact and (c) 20% smaller.



Source: By the author

In ADCIG domain, Differential Semblance optimization (DS) (SYMES; CARAZZONE, 1991) aims to measure the flatness events by minimization of a local derivative. Stack Power optimization (SP) aims to maximize the amplitude stacking. In ODCIG, Differential Semblance minimize all energy measured in non zero-offset image by linear offset penalty factor, and Stack Power aims to maximize the zero-offset image amplitude. The Table 2.1 show schematically the Stack Power and Differential Semblance objective functions in angle domain (blue) and in offset domain (red).

Table 2.1 – WEMVA objective functions: Stack Power and Differential Semblance for angle domain (blue) and offset domain (red).

	Stack Power (SP) (Maximization)	Differential Semblance (DS) (Minimization)
ADCIG	$\Phi(I) = \sum_{\mathbf{x}, \theta} [I(\mathbf{x}, \theta)]^2$	$\Phi(I) = \sum_{\mathbf{x}, \theta} \left[\frac{dI}{d\theta}(\mathbf{x}, \theta) \right]^2$
ODCIG	$\Phi(I) = \sum_{\mathbf{x}} [I(\mathbf{x}, \mathbf{0})]^2$	$\Phi(I) = \sum_{\mathbf{x}, \mathbf{h}} [hI(\mathbf{x}, \mathbf{h})]^2$

Like the objective function of FWI, also the Stack Power function suffers from a lot of local minima (VIRIEUX; OPERTO, 2009), which makes the function subject to cycle skipping. In contrast Differential Semblance, by its local analysis, has better convergence properties, and less local minima. In FWI, the wave-phase adjustment is related with

background velocity optimization and the amplitude fit with the high resolution of the velocity. Analogously in WEMVA we can say that Differential Semblance leads to a kind of phase driven optimization, whereas Stack Power leads to a kind of amplitude driven optimization.

Many authors (CHAVENT; JACEWITZ, 1995; WEIBULL; ARNTSEN, 2013) combine these two measures into a single objective function,

$$\Phi(I) = \frac{1}{2} \sum_{\mathbf{x}, \mathbf{h}} [\mathbf{h}I(\mathbf{x}, \mathbf{h})]^2 - \frac{\gamma}{2} \sum_{\mathbf{x}} [I(\mathbf{x}, \mathbf{0})]^2, \quad (2.4)$$

where the γ parameter needs to be adjusted like a regularization. The many local minima of Stack Power suggest to choose a γ parameter, that emphasizes the Stack Power function, when Differential Semblance amplitude has lost its sensibility. This is a hard task without previously knowing the minimum Differential Semblance value. Even with this information, the convergence is not guaranteed because Stack Power is not totally an amplitude optimization.

Zhang and Shan (2013) noted that the difference between Stack Power and Differential Semblance is the filter applied in each one. In Stack Power maximization the filter constant 1 is convolved with ADCIG map along angle coordinate. By a Radon transform, this process is equivalent an application of Dirac's delta $\delta(\mathbf{h})$ in ODCIG domain. Differential Semblance is a convolution of derivative operator in ADCIG domain. The equivalent in ODCIG domain is a offset linear penalty factor $|\mathbf{h}|$. Both operations can be generalized as:

$$\Phi(I) = \frac{1}{2} \sum_{\mathbf{x}, \theta} [g(\theta) * I(\mathbf{x}, \theta)]^2 \quad (2.5)$$

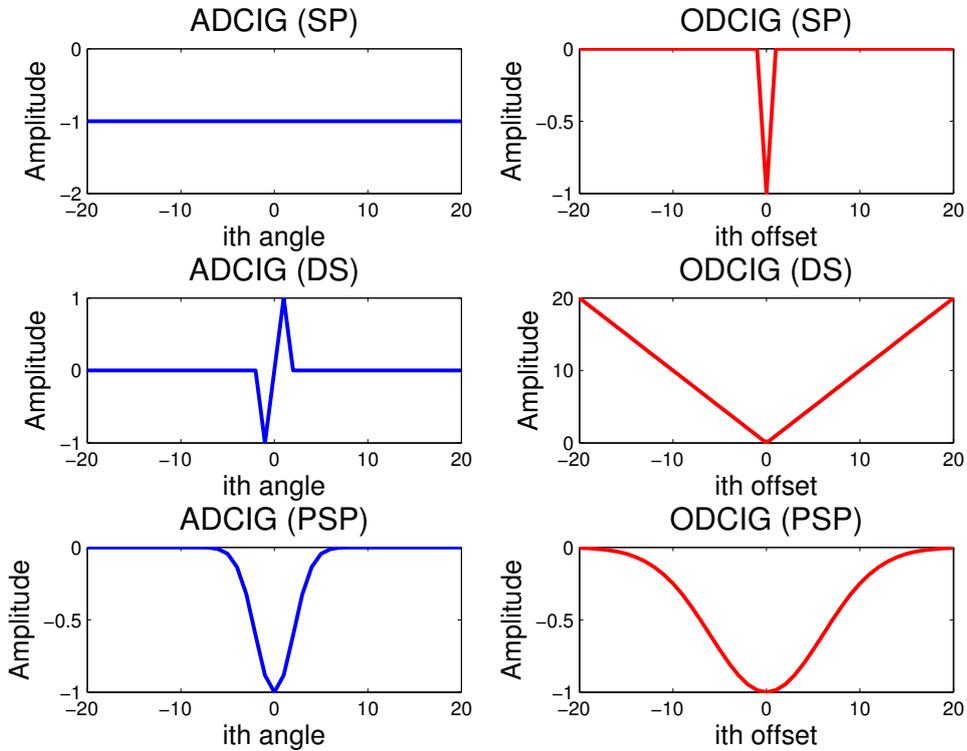
in an ADCIG and

$$\Phi(I) = \frac{1}{2} \sum_{\mathbf{x}, \mathbf{h}} [G(\mathbf{h})I(\mathbf{x}, \mathbf{h})]^2 \quad (2.6)$$

in an ODCIG, see Figure 2.5.

To combine the Differential Semblance and Stack Power optimization, Zhang and Shan (2013) proposed the maximization of Partial Stack Power, equations 2.5 and 2.6, where the weights $g(\theta)$ and $G(\mathbf{h})$ are represented by variable Gaussian windows. The Zhang and Shan (2013) idea is gradually transform the velocity model sensibility from Differential Semblance to Stack Power analysis, by squeezing the Gaussian window as the energy becomes more focused as proceed the inversion.

Figure 2.5 – Stack Power (SP), Differential Semblance (DS) and Partial Stack Power (PSP) penalty operators, in ADCIG (blue) and ODCIG (red) domains.



Source: By the author

2.3 Constraining the velocity model with the reflector map

It is very common to condition an ill-posed problem with the Tikhonov regularization (TIKHONOV; ARSENIN, 1977). The Tikhonov regularization minimizes the problem according to

$$\arg \min \quad \Phi_{reg} = \Phi + \lambda \|\Gamma(c - c_0)\|^2, \quad (2.7)$$

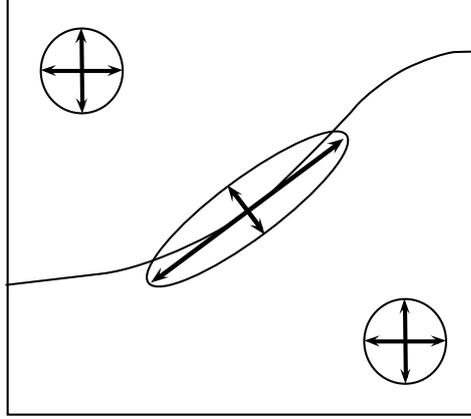
where Γ is the Tikhonov differential operator, c_0 is a prior velocity model and λ the regularization parameter. It is possible to impose different conditions to the solutions, as the minimum norm, $\Gamma \equiv 1$, the isotropic smoothness, $\Gamma \equiv \nabla$, or the gradient condition, $\Gamma \equiv \nabla^2$. Anisotropic smoothness conditions can be applied by general Tikhonov regularization,

$$R = \lambda [\nabla (c - c_0)]^T \mathbf{L} \nabla (c - c_0), \quad (2.8)$$

where \mathbf{L} is a symmetric weight matrix. For structural regularization according to Kaipio et al. (1999), \mathbf{L} must produce an anisotropic smoothness giving smaller penalties across discontinuities than along the faults and at places with no discontinuity the weight acts isotropically, see Figure 2.6. Kaipio et al. (1999) chosen as

$$\mathbf{L} = \mathbf{I} - (1 + |\nabla I|^2)^{-1} \mathbf{S}, \quad (2.9)$$

Figure 2.6 – The arrows show the eigenvalues of 2.9. Higher weights along structures than across them, and equal weights inside the layers.



Source: By the author

where \mathbf{I} the identity matrix and $\mathbf{S} = \nabla I (\nabla I)^T$ is the structure tensor.

Analyzing equation 2.9, removing the normalization and the isotropic term, I get

$$\mathbf{L} = |\nabla I|^2 \mathbf{I} - \mathbf{S} = \mathbf{M}^T \mathbf{M}, \quad (2.10)$$

with

$$\mathbf{M} = \begin{pmatrix} 0 & \frac{\partial I}{\partial x_3} & -\frac{\partial I}{\partial x_2} \\ -\frac{\partial I}{\partial x_3} & 0 & \frac{\partial I}{\partial x_1} \\ \frac{\partial I}{\partial x_2} & -\frac{\partial I}{\partial x_1} & 0 \end{pmatrix}. \quad (2.11)$$

Identity 2.10 is the cross-gradient factor in Kaipio et al. (1999) regularization,

$$(\nabla c)^T \mathbf{L} \nabla c = (\mathbf{M} \nabla c)^T (\mathbf{M} \nabla c) \equiv |\nabla I \times \nabla c|^2. \quad (2.12)$$

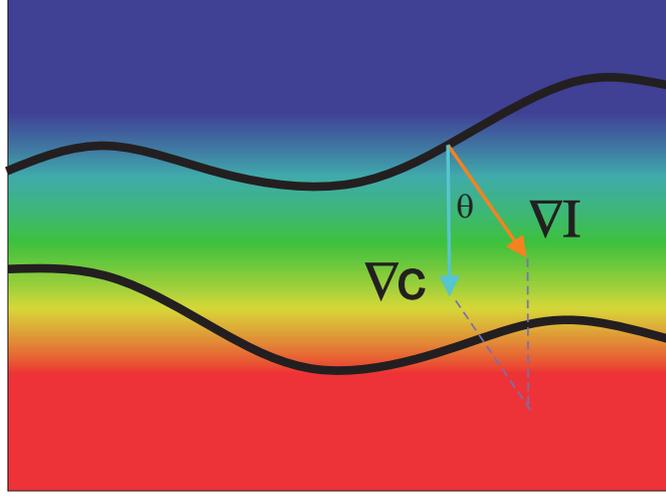
The last term of 2.10 is the cross-gradient version of the inner product,

$$(\nabla c)^T \mathbf{S} \nabla c = (\nabla c)^T \nabla I (\nabla I)^T \nabla c \equiv (\nabla I \cdot \nabla c)^2. \quad (2.13)$$

An interpretation of cross-gradient optimization is show in Figure 2.7. The figure shows two black curves indicating the image/reflectivity map embedded in a linear vertical gradient velocity. When the velocity model is almost adjusted, the WEMVA's image update has few changes. From now, the objective function decrease manly by changing the $\nabla c(\mathbf{x})$ direction towards the $\nabla I(\mathbf{x})$ direction. Parallelizing the gradient vectors can be done by minimization (maximization) of their outer (inner) product in a least-square sense.

We can redefine Stack Power and Differential Semblance in terms of cross-gradient constraint. The Table 2.2 shows the cross-gradient versions from Table 2.1.

Figure 2.7 – A cross-gradient regularization parallelize the gradient vectors ∇c and ∇I , minimizing (maximize) their outer (inner) product in a least-square sense.



Source: By the author

Table 2.2 – WEMVA objective functions constrained by cross-gradient regularization: Cross-Gradient Stack Power (CGSP) and Cross-Gradient Differential Semblance (CGDS), for angle domain (blue) and offset domain (red).

	CG Stack Power (CGSP) (Maximization)	CG Differential Semblance (CGDS) (Minimization)
ADCIG	$\Phi(I, c) = \sum_{\mathbf{x}, \theta} [\nabla I(\mathbf{x}, \theta) \cdot \nabla c(\mathbf{x})]^2$	$\Phi(I, c) = \sum_{\mathbf{x}, \theta} \left \frac{d}{d\theta} \nabla I(\mathbf{x}, \theta) \times \nabla c(\mathbf{x}) \right ^2$
ODCIG	$\Phi(I, c) = \sum_{\mathbf{x}} [\nabla I(\mathbf{x}, \mathbf{0}) \cdot \nabla c(\mathbf{x})]^2$	$\Phi(I, c) = \sum_{\mathbf{x}, \mathbf{h}} \mathbf{h} ^2 \nabla I(\mathbf{x}, \mathbf{h}) \times \nabla c(\mathbf{x}) ^2$

The same construction can be done for the Partial Stack Power. Following the normalization proposed by Zhang and Shan (2013), the Cross Gradient Partial Stack Power in an ODCIG is defined as:

$$\psi(I, c) = \frac{1}{2} \sum_{x_2} \frac{\sum_{x_1, h_2} [\tilde{G}(\mathbf{h}) \nabla I(\mathbf{x}, \mathbf{h}) \cdot \nabla c(\mathbf{x})]^2}{\sum_{x_1, h_2} [\nabla I(\mathbf{x}, \mathbf{h}) \cdot \nabla c(\mathbf{x})]^2}. \quad (2.14)$$

A cross-gradient objective function optimization brings high-frequency information to the velocity. It is not appropriated to apply it if the background velocity model is not solved yet. To include the cross-gradient from the beginning, it is necessary to couple it with a standard optimization function. Here, I combine the Zhang and Shan (2013) function

$$\varphi(I) = \frac{1}{2} \sum_{x_2} \frac{\sum_{x_1, h_2} [G(\mathbf{h}) I(\mathbf{x}, \mathbf{h})]^2}{\sum_{x_1, h_2} [I(\mathbf{x}, \mathbf{h})]^2} \quad (2.15)$$

with the Cross-Gradient Partial Stack Power, equation 2.14. From now on, this work will

use for the Cross-Gradient Partial Stack Power function the definition

$$\Phi(I, c) = \varphi(I)\psi(I, c). \quad (2.16)$$

The gradient of equation 2.16 is calculated in Appendix A.

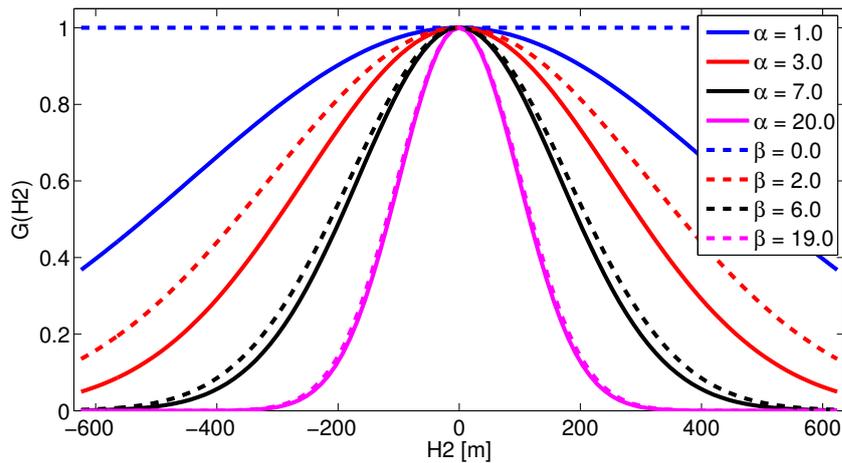
The Gaussian weights are defined as:

$$G(\mathbf{h}) = \exp(-\alpha h_2^2/H^2), \quad (2.17a)$$

$$\tilde{G}(\mathbf{h}) = \exp(-\beta h_2^2/H^2), \quad (2.17b)$$

where H is the maximum sub-surface offset, and α and β are parameters for squeezing the Gaussian. Mathematically \tilde{G} is independent of G , and in principal, choosing α and β is as undetermined as λ in Tikhonov regularization. However some conditions need to be taken into account: α starts with a small value [wide $G(\mathbf{h})$] to include the most unfocused energy. As the inversion proceeds, and the gathers become more focused, α is gradually increased [small $G(\mathbf{h})$] to approach the Stack Power analysis (ZHANG; SHAN, 2013). For the first iterations, the β value needs to give priority to φ in order to estimate estimate the background velocity. As the inversion proceeds, the weights for φ and ψ must approach each other asymptotically (see Figure 2.8). One suggestion is to take $\beta = \alpha(1.0 - \alpha_0/\alpha)$, where α_0 ($\alpha \geq \alpha_0$) is the first α selected.

Figure 2.8 – Gaussian window changing the phase analysis (large) to amplitude analysis (squeezed). Note the delay between the Partial Stack Power (continuous lines) and Cross-Gradient Partial Stack Power (dashed lines), which is reduced when the background is adjusted.



Source: By the author

2.4 Numerical experiments

To validate the Cross-Gradient Partial Stack Power I used synthetic data from a smoothed version of the Marmousi data (VERSTEEG; GRAU, 1991) called Marmousoft (BILLETTE et al., 2003). It is a simulated marine acquisition with 261 shots and a streamer with 96 receivers. The data was modeled using the Born approximation, which means that it contains no multiples or refracted waves. The pulse is a Ricker wavelet with peak frequency of 12 Hz.

The model possesses 243 samples in the x_1 and 767 in the x_2 direction, both spaced at 12 m. The velocities vary between 1500 m/s and 4750 m/s. Figure 2.9a shows the true velocity model.

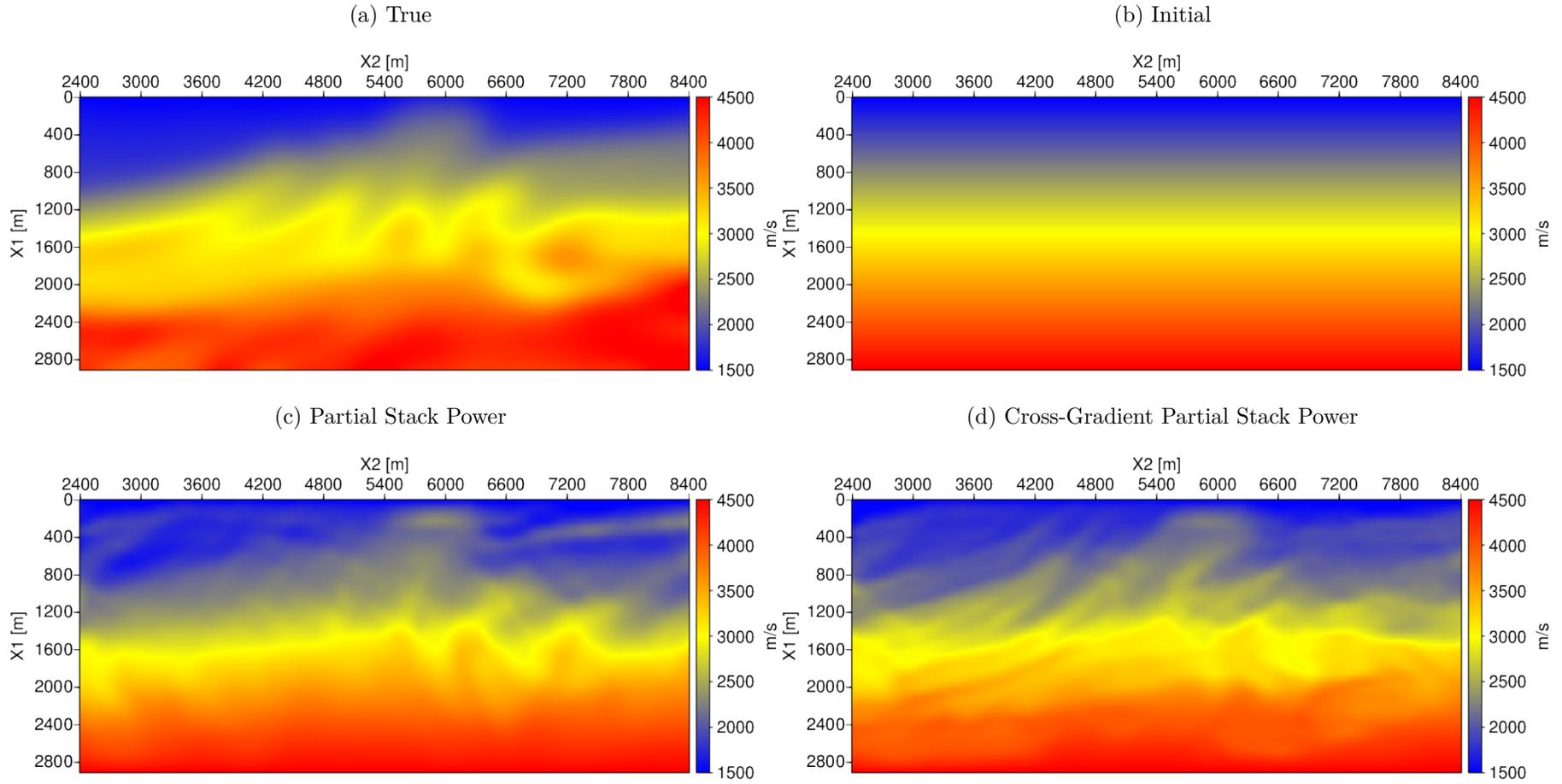
The objective function gradient was calculated by adjoint state method (PLESSIX, 2006), see Appendix A. For optimization process I used the quasi-Newton algorithm L-BFGS-B (BYRD et al., 1995), see Appendix B. The numbers of parameters was reduced by B-splines interpolation, see Appendix C. The interpolated model has 37 samples in the x_1 and 97 samples in the x_2 direction, spaced at 81 m and 95.875 m respectively. The initial model is a linear vertical gradient with velocity varying between 1500 m/s and 4500 m/s (see Figure 2.9b). The maximum subsurface offset is equal to 624 m. The optimization took 4 steps with 15 iterations each. At each step, α was increased empirically, taking the values 1.0, 3.0, 7.0 and 20.0 (Figure 2.8).

Figure 2.9c show the model resulting from Partial Stack Power optimization. It is a model with little change with respect of the initial model, but enough to produce a well focused image (Figure 2.10c). Figure 2.9d show the velocity model build with Cross-Gradient Partial Stack Power optimization. We can note the cross-gradient regularization gives a structural information similar founded in true velocity model (Figure 2.9a), and less artifacts when we compare the region above $x_1 = 800$ m between $x_2 = 6600$ m and $x_2 = 8400$ m, in Figures 2.9c and 2.9d. The correspondent PSP and CGPSP migrated images are show in Figures 2.10c 2.10d respectively. In both cases we can see good convergence, when we compare the migrated image using the true velocity model (Figure 2.10a), and initial model (Figure 2.10b), but for CGPSP we note a better definition of the layers on the region above $x_1 = 800$ m between $x_2 = 6600$ m and $x_2 = 8400$ m.

Figures 2.11 to 2.14 show selected CIGSs (3000, 3600, 4200, 4800, 5400, 6000, 6600, 7200, 7800 and 8400 m) as obtained with the four velocity models. By the energy focused at zero-offset we evaluate how the velocity background is fitted well. In both cases, PSP and CGPSP (Figures 2.13 and 2.14 respectively), we observe no correlation for far subsurface offsets. Note the convergence comparing the true migration at Figure 2.11, and the CIG migrated with initial model at Figure 2.12.

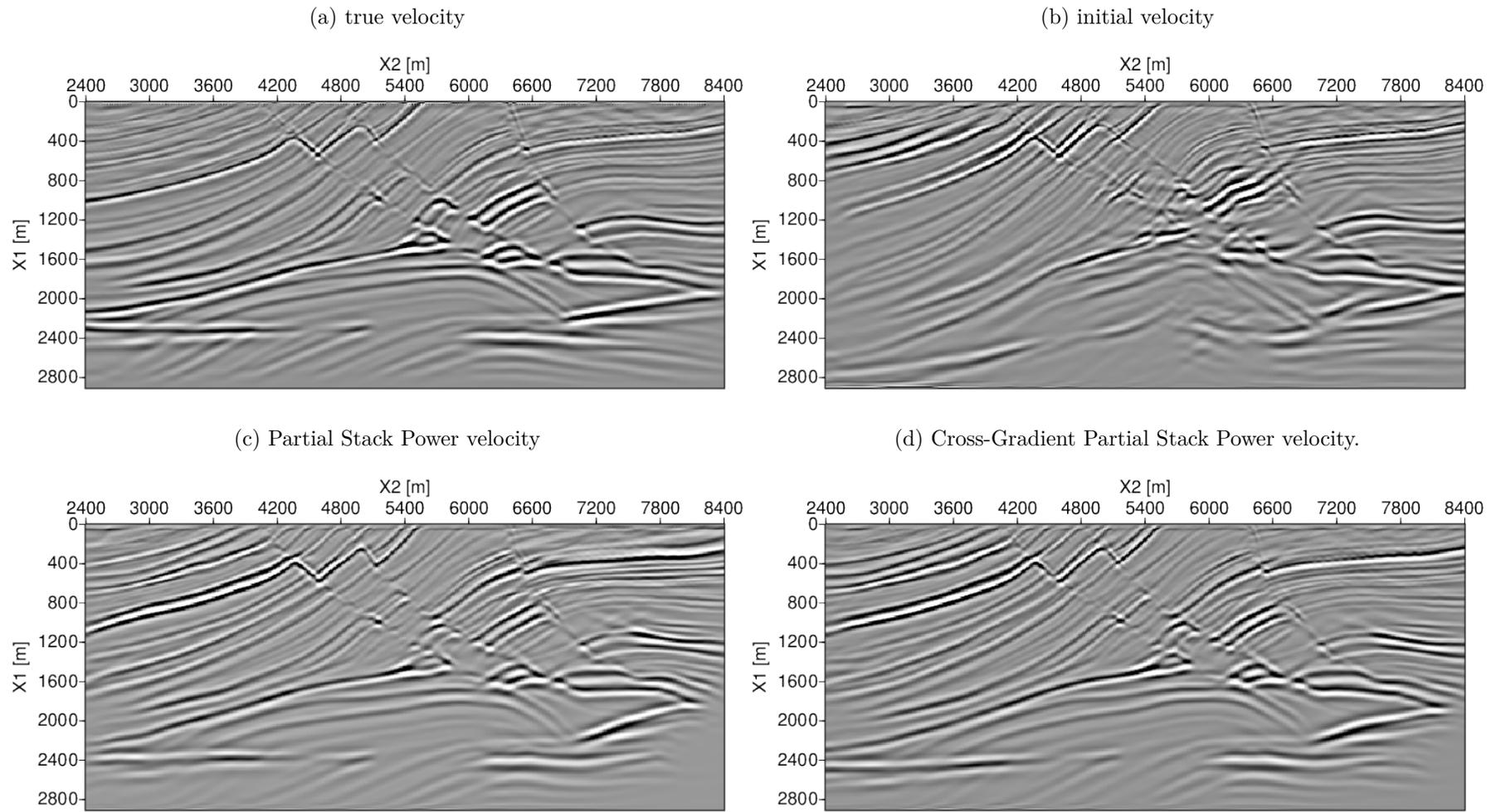
The short wavelength components of the velocity model are related with amplitude maximization in zero-offset image. The Figures 2.15 and 2.16 show the zero-offset image trace took at the CIG position mentioned above. Particularly at the positions around the target, $x_2 = 6000$ m, $x_2 = 6600$ m and $x_2 = 7200$ m (the target depth is $x_1 = 2400$ m). We can see the Cross-Gradient Partial Stack Power has better results in sense of avoiding cycle skipping and amplitude fitting than regular Partial Stack Power (intervals 1000 to 2000 in CIG 6000, first 800 m in CIG 6600). Unfortunately the interweave of parameters spread in 2D model, has nonlinearity to makes a local cycle skipping, see the interval 1000 to 2000 in CIG 7200, even with a correct time shift on the rest of the trace.

Figure 2.9 – Velocity model.



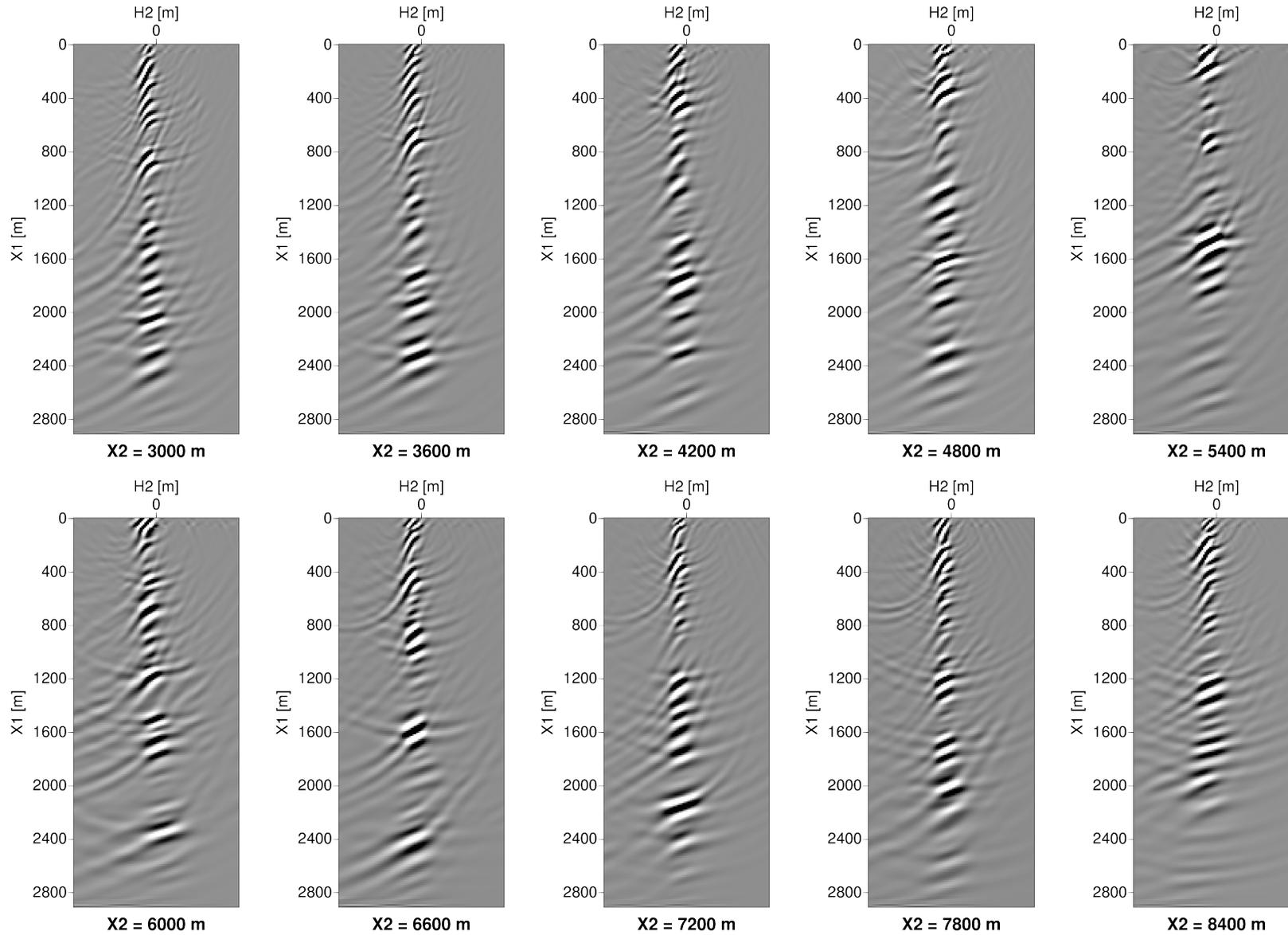
Source: By the author

Figure 2.10 – Migrated image with:



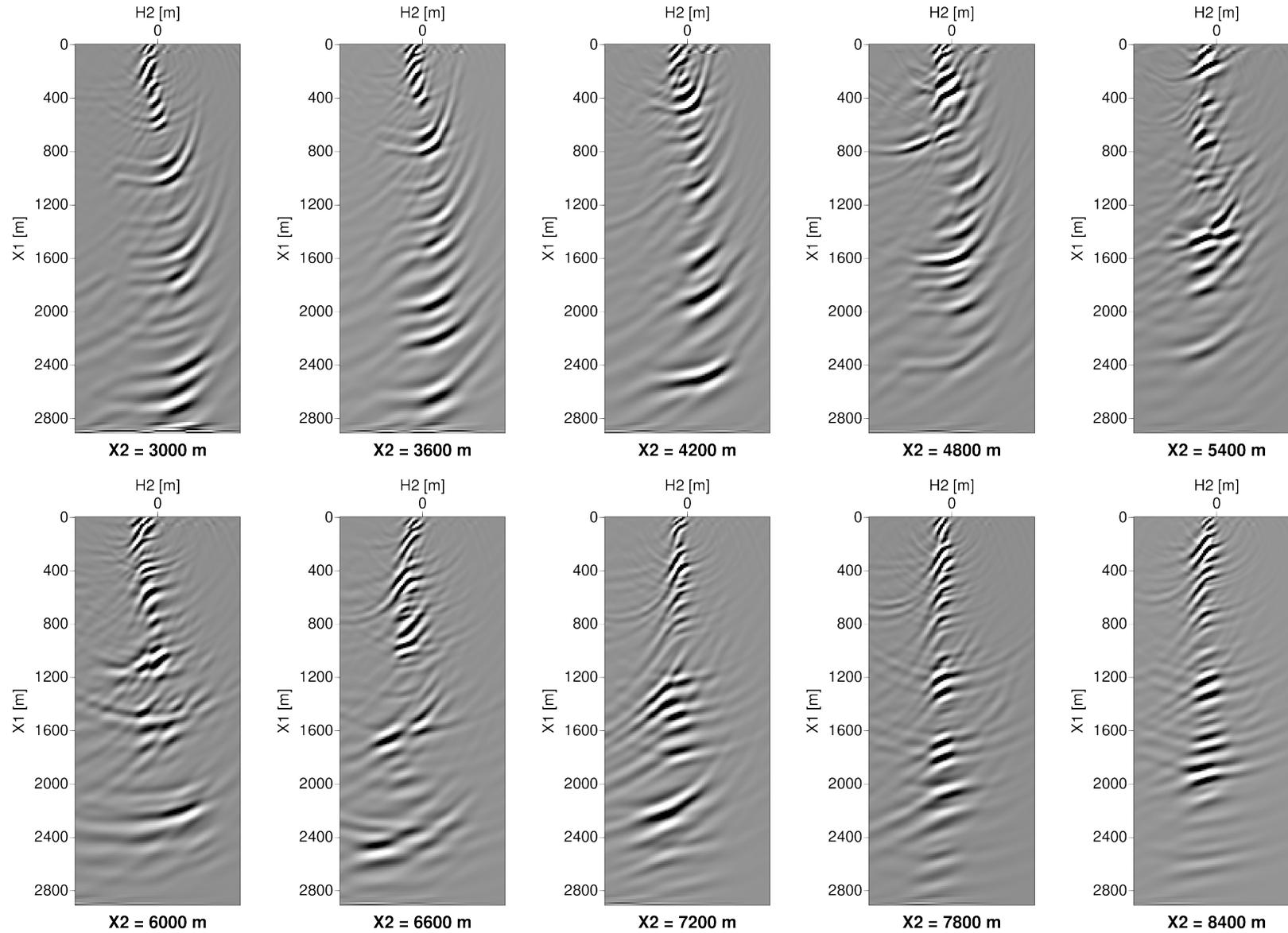
Source: By the author

Figure 2.11 – ODCIG migrated with true velocity.



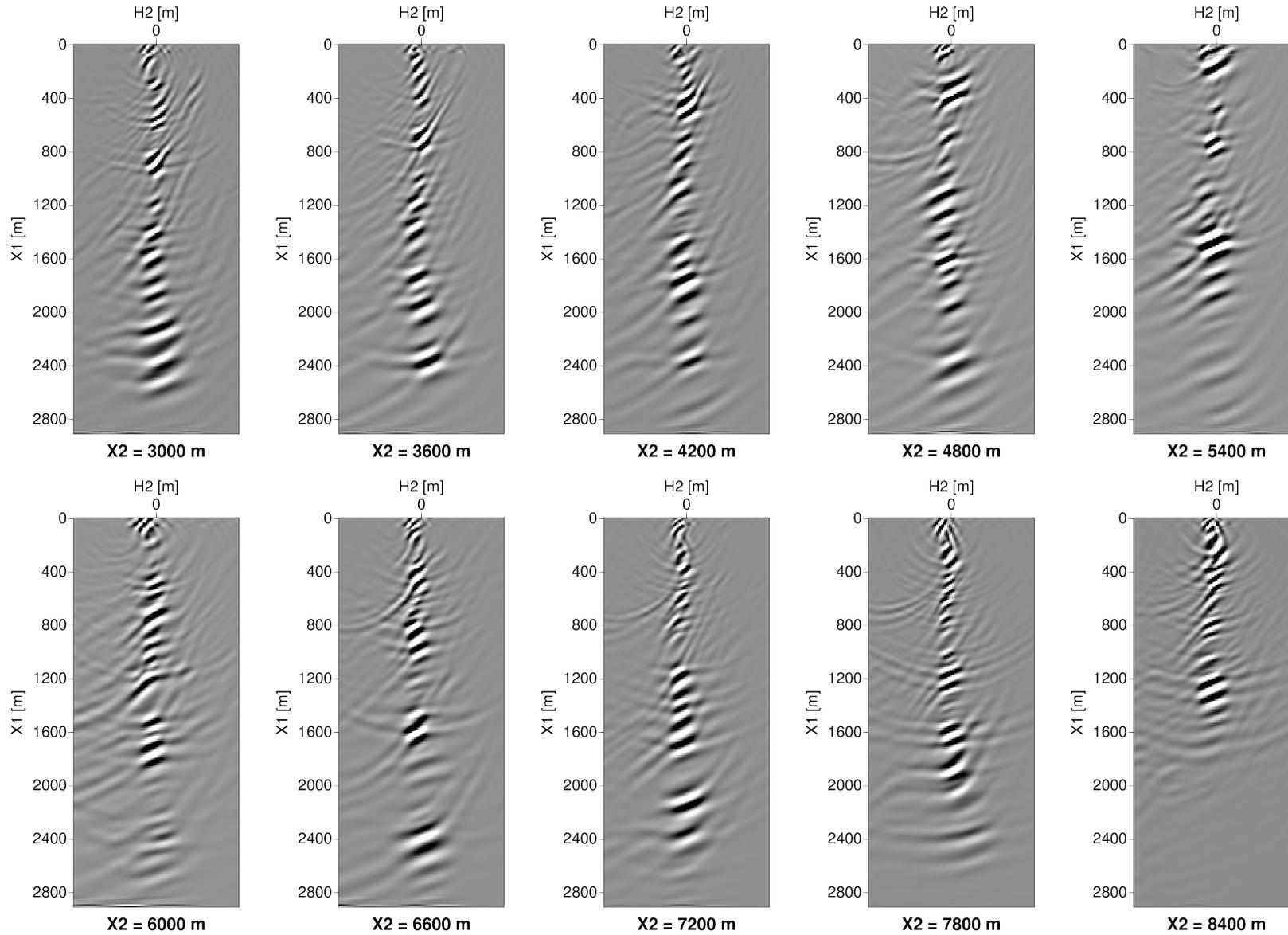
Source: By the author

Figure 2.12 – ODCIG migrated with initial velocity.



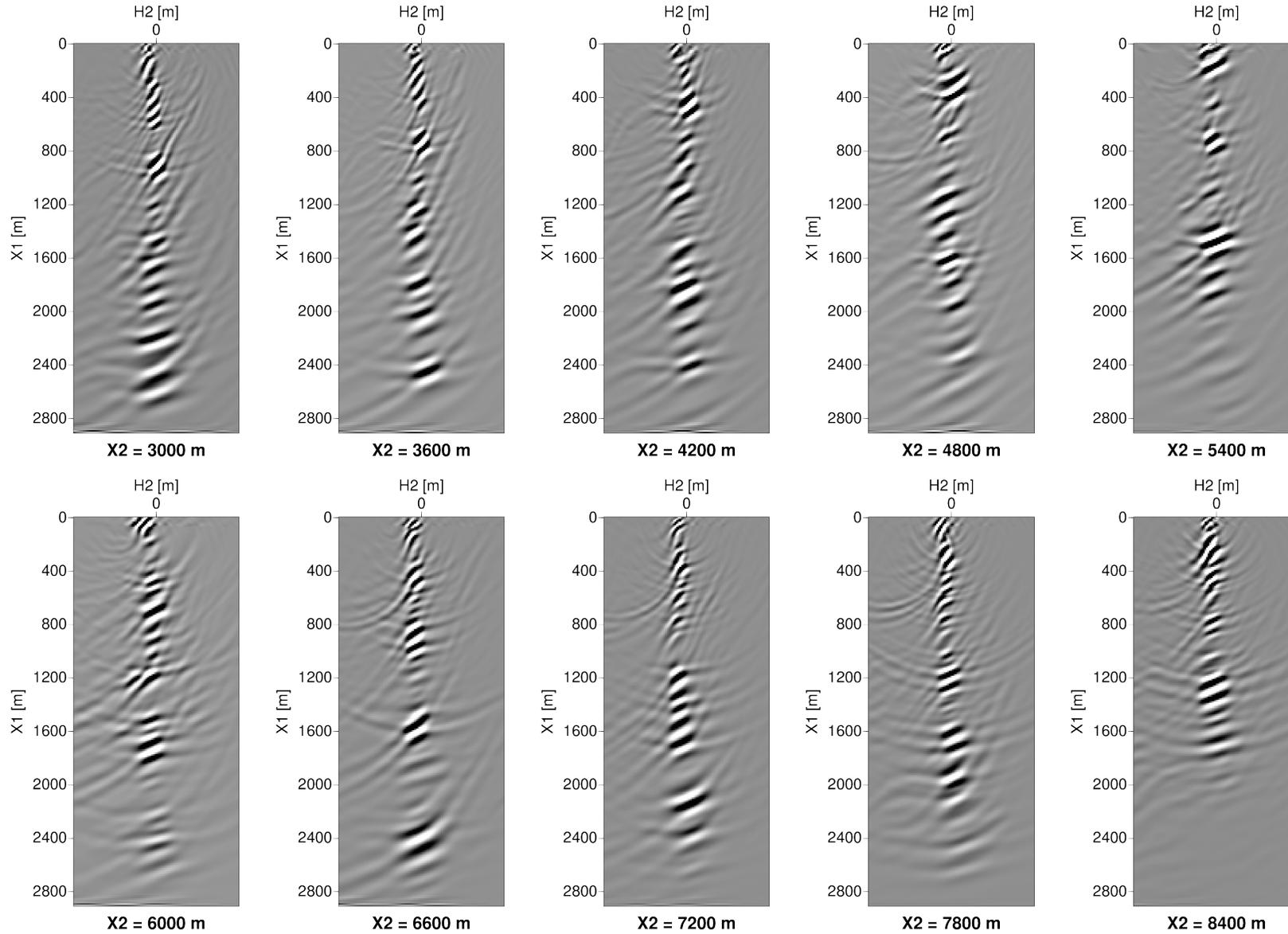
Source: By the author

Figure 2.13 – ODCIG migrated with optimized Partial Stack Power velocity.



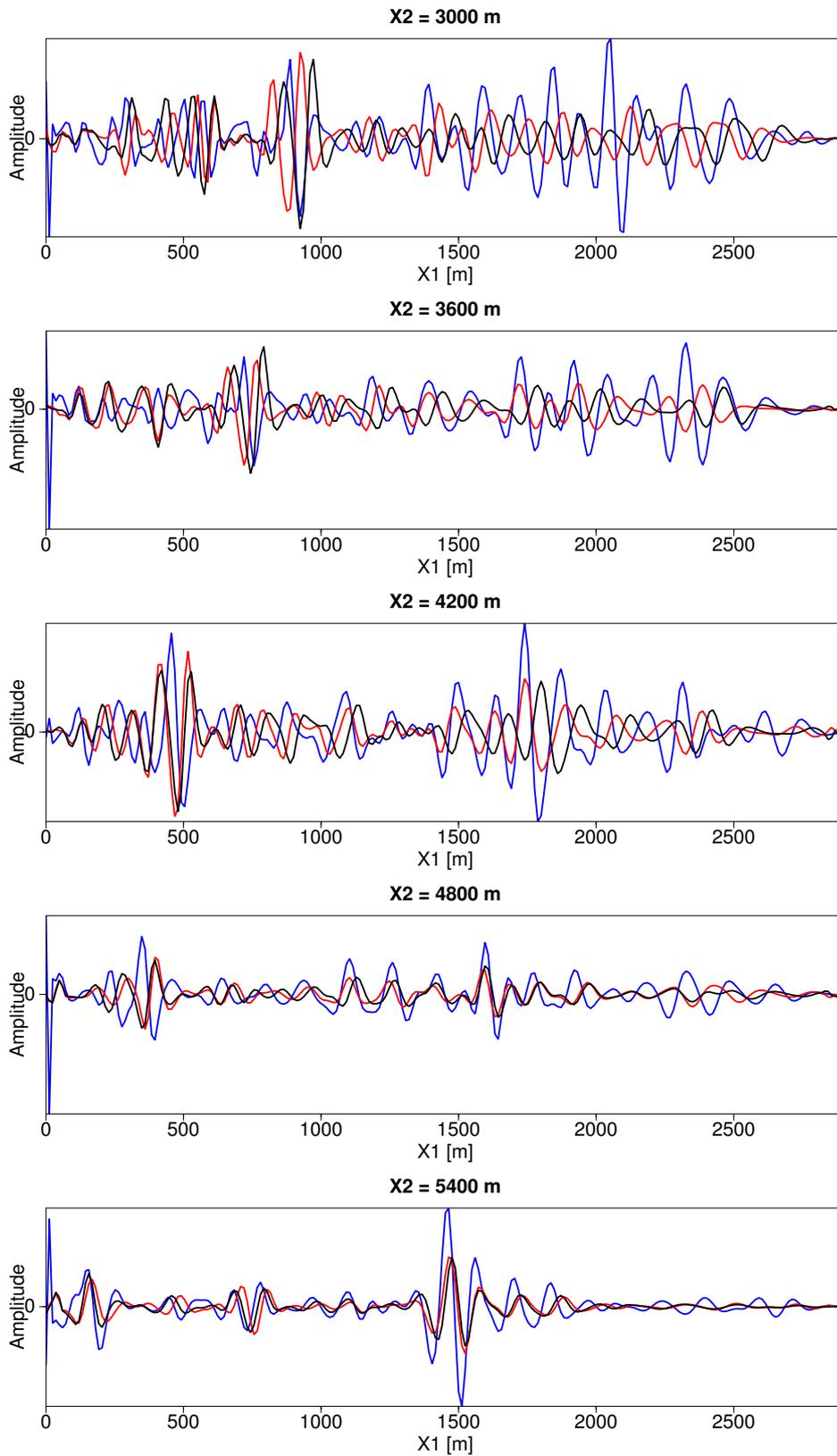
Source: By the author

Figure 2.14 – ODCIG migrated with optimized Cross-Gradient Partial Stack Power velocity.



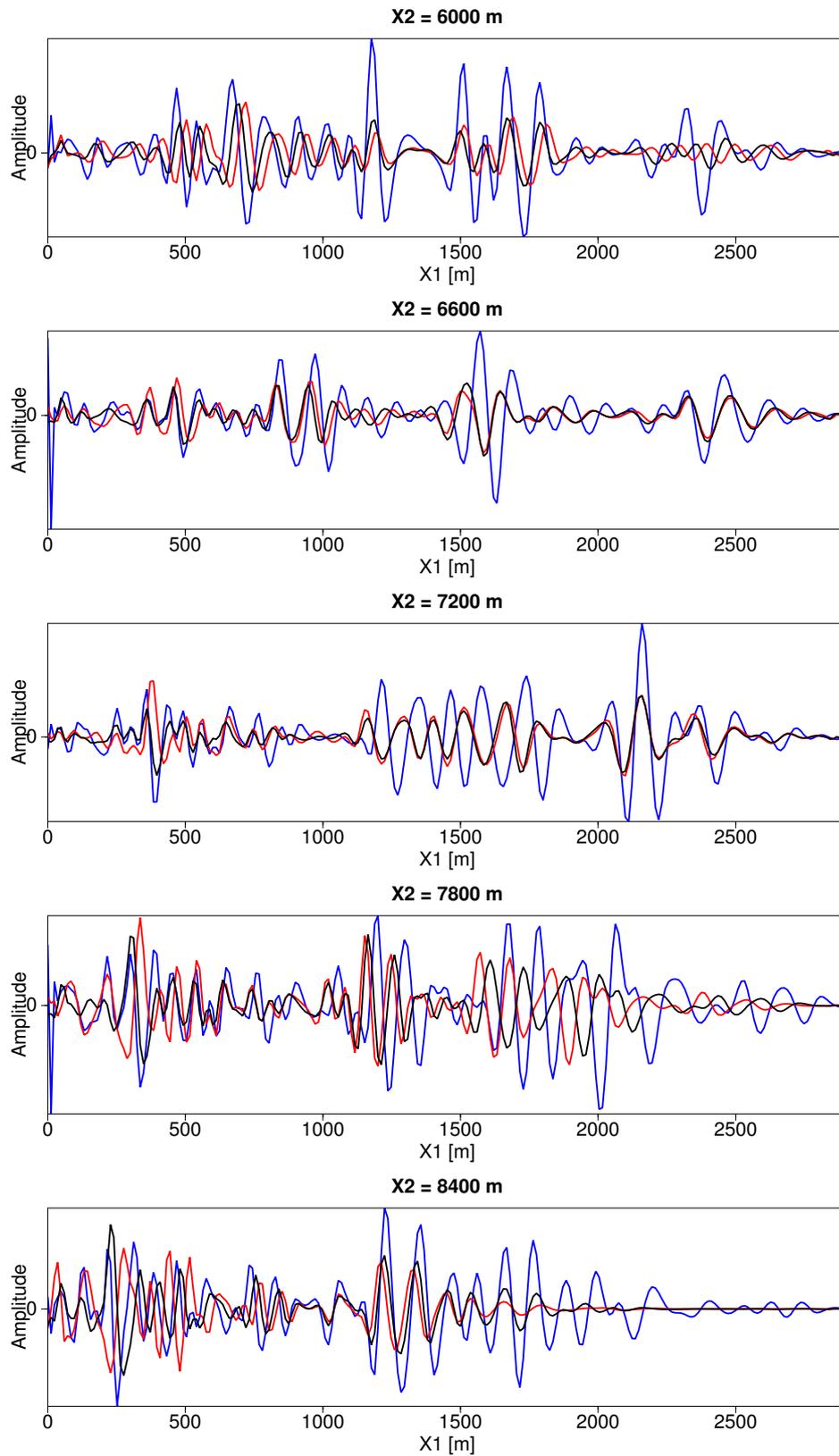
Source: By the author

Figure 2.15 – ODCIG zero offset image traces $x_1 = (3000, 3600, 4200, 4800, 5400)$ m: true migration (blue), Partial Stack Power migration (red) and Cross-Gradient Partial Stack Power (black).



Source: By the author

Figure 2.16 – ODCIG zero offset image traces $x_1 = (6000, 6600, 7200, 7800, 8400)$ m: true migration (blue), Partial Stack Power migration (red) and Cross-Gradient Partial Stack Power (black).



Source: By the author

2.5 Conclusions

In this Chapter I discussed preconditioning WEMVA adding structural information extracted from stacked migrated images. In order to enforce the structural information we require that the gradient of the velocity field and the gradient of the image intensity are collinear. This regularization was applied as a multiplicative factor on the partial stack power objective function. The structural regularization is gradually enforced during the quasi-Newton iterations in order to avoid premature convergence to local minima. This modification resulted in velocity models correlated with the seismic image and improved the migration focusing.

The new objective function is very sensitive to high frequency noise. Preprocessing is required to attenuate multiples and coherent noises. Additionally, dynamic preconditioning is required to decrease the nonlinearity and improve image focusing. In our numerical experiments we used a Gaussian filter to improve the robustness when numerically computing of the gradients of image and velocity. A Gaussian filter is used also to remove high frequency artifacts in the objective function gradient. B-splines parameterization of the velocity model reduces the number of model parameters and also enforces smoothness preconditioning in the gradient of the objective function.

Our numerical experiments on the Marmousoft data set indicate that the proposed method preconditioning the convergence of the method, resulting in focused common image gathers and a velocity model correlated with the image. The RTM image computed with the estimated velocity model is very close to the image using the exact velocity model except in regions compromised by lack of illumination.

3 JOINT MIGRATION INVERSION

Joint Migration Inversion (JMI) (BERKHOUT, 2014a; BERKHOUT, 2014b; BERKHOUT, 2014c) is a data and image based inversion method that uses the Full Wavefield Migration (FWM) to simultaneously estimate the reflectivity and velocity models. It is a data driven method in sense that it minimizes the squared error, between the observed and modeled data, and it is an image driven method in sense that the migration operators are iteratively estimated and back projected to a velocity update. JMI is a new strategy for seismic inversion, because it has the advantage of decoupling the reflectivity and velocity, reducing the nonlinearity, which usually appears, in the first iterations in traditional methods like FWI or WEMVA. This chapter, based on Staal and Verschuur (2014), will describe schematically the JMI method, introducing the concepts of Full Wavefield Modeling (FWMod) and how it is includes in reflectivity estimation (by FWM) and extended to velocity estimation (JMI). This chapter will give a JMI background for a better understanding of the next chapter, where we will combine the JMI method with morphological operators.

3.1 Full Wavefield Modeling

Full Wavefield Modeling uses integral operators to evaluate separately the phase changes, for the estimation of kinematically accurate velocity models, and the amplitude changes, for the estimation of reflectivity maps. The wavefield modeling use the Rayleigh II integral,

$$P(x_1^m, x_2^m) = 2 \int_{-\infty}^{+\infty} \frac{\partial G}{\partial x_1}(x_1^m, x_2^m; x_1^n, x_2) Q(x_1^n, x_2) dx_2, \quad (3.1)$$

where $P(x_1^m, x_2^m)$ is the monochromatic extrapolated wavefield to the position (x_1^m, x_2^m) , $Q(x_1^n, x_2)$ is the previous known wavefield, measured at position (x_1^n, x_2) . $\frac{\partial G}{\partial x_1}$ is vertical derivative of Green's function. In discrete form, equation 3.1 is a matrix-vector equation:

$$\vec{P}(x_1^m) = \mathbf{W}(x_1^m, x_1^n) \vec{Q}(x_1^n), \quad (3.2)$$

where $\vec{P}(x_1^m)$ is the extrapolated wavefield measured at depth level x_1^m , $\vec{Q}(x_1^n)$ is the wavefield that we want to extrapolate, measured at depth level x_1^n , and matrix \mathbf{W} is

$$\mathbf{W}_{i,j}(x_1^m, x_1^n) = 2 \frac{\partial G}{\partial x_1}(x_1^m, x_2^j; x_1^n, x_2^i). \quad (3.3)$$

The JMI approach assumes that the velocity is locally homogeneous, in other words the columns of matrix \mathbf{W} use the local velocity $c(x_1^n, x_2^j)$ and by the Toeplitz structure, we can relate each column to a phase shift operator in k_{x_2} domain. Then each column j of

\mathbf{W} is written as

$$\vec{W}_j(x_1^m, x_1^n) = \mathcal{F}_x^{-1} \left[e^{-ik_{x_1} \Delta x_1} e^{-ik_{x_2} x_2^j} \right] \quad (3.4)$$

where $\mathcal{F}_{x_2}^{-1}$ is the inverse Fourier transform in x_2 , $k_{x_1} = \sqrt{k^2 - k_{x_2}^2}$ with $k = \omega/c$, x_2^j is the source position of Green's function and $\Delta x_1 = |x_1^n - x_1^m|$. If we consider the upgoing extrapolation operator \mathbf{W}_j^- equal to the equation 3.4, to maintain the reciprocity between source and receiver, as in the homogeneous case, the downgoing extrapolation operator must satisfy:

$$\mathbf{W}^+(x_1^{n+1}, x_1^n) = \left[\mathbf{W}^-(x_1^n, x_1^{n+1}) \right]^T. \quad (3.5)$$

When the wavefield encounters a sharp velocity contrast, a part of the energy will scatter and a part will be transmitted. The incident wavefield at depth level x_1^n can be written as:

$$\vec{Q}^-(x_1^n) = \mathbf{R}^\cup(x_1^n) \vec{P}^+(x_1^n) + \mathbf{T}^\cap(x_1^n) \vec{P}^-(x_1^n), \quad (3.6a)$$

$$\vec{Q}^+(x_1^n) = \mathbf{T}^\cup(x_1^n) \vec{P}^+(x_1^n) + \mathbf{R}^\cap(x_1^n) \vec{P}^-(x_1^n), \quad (3.6b)$$

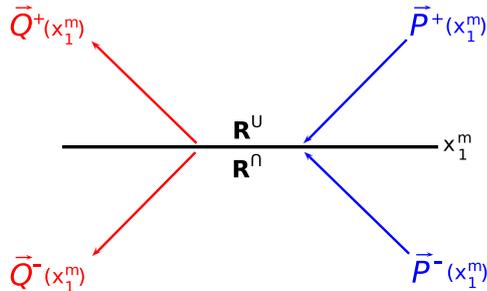
where $P^-(x_1^n)$ and $P^+(x_1^n)$ are the upgoing and downgoing wavefields arriving at level x_1^n . $Q^-(x_1^n)$ and $Q^+(x_1^n)$ are the upgoing and downgoing wavefields leaving level x_1^n . The operators \mathbf{R}^\cup and \mathbf{R}^\cap are the reflectivity coefficients acting on incident waves from above and from below respectively (see Figure 3.1). The operators \mathbf{T}^\cup and \mathbf{T}^\cap are the analogous transmission operators. In acoustic approximation, the transmission coefficients are related by

$$\mathbf{R}^\cup = -\mathbf{R}^\cap = \mathbf{R} \quad (3.7a)$$

$$\mathbf{T}^\cup = \mathbf{I} + \mathbf{R} \quad (3.7b)$$

$$\mathbf{T}^\cap = \mathbf{I} - \mathbf{R}. \quad (3.7c)$$

Figure 3.1 – Schematic representation of incoming and outgoing wavefields acting for each grid-point of depth level x_1^m .



Source: By the author

Then using equations 3.7 in 3.6, we find

$$\vec{Q}^-(x_1^n) = \vec{P}^-(x_1^n) + \delta\vec{P}(x_1^n) \quad (3.8a)$$

$$\vec{Q}^+(x_1^n) = \vec{P}^+(x_1^n) + \delta\vec{P}(x_1^n), \quad (3.8b)$$

where $\delta\vec{P} = \mathbf{R}(\vec{P}^+ - \vec{P}^-)$ is the scattered wavefield.

Full Wavefield Modeling consists of recursive downgoing and upgoing extrapolation. The source wavefield $\vec{S}(x_1^0)$ is previously known, and the fields outside the grid is assumed to be zero. Schematically the recursive process is:

1. Modeling the downgoing wavefield ($m = 0, 1, 2, \dots, M - 1$)

- At the surface, x_1^0 , add the source and the scattered wavefields:

$$\vec{Q}^+(x_1^0) = \vec{P}^+(x_1^0) + \delta\vec{P}(x_1^0) + \vec{S}(x_1^0), \quad (3.9)$$

for $m \neq 0$ add only the scattering:

$$\vec{Q}^+(x_1^m) = \vec{P}^+(x_1^m) + \delta\vec{P}(x_1^m). \quad (3.10)$$

- Then extrapolate the wavefield with downgoing extrapolation operator \mathbf{W}^+ :

$$\vec{P}^+(x_1^{m+1}) = \mathbf{W}^+(x_1^{m+1}, x_1^m)\vec{Q}^+(x_1^m). \quad (3.11)$$

2. Modeling the upgoing wavefield ($m = M, M - 1, M - 2, \dots, 1$)

- Add the scattered field:

$$\vec{Q}^-(x_1^m) = \vec{P}^-(x_1^m) + \delta\vec{P}(x_1^m). \quad (3.12)$$

- Extrapolate the wavefield with the upgoing extrapolation operator \mathbf{W}^- :

$$\vec{P}^-(x_1^{m-1}) = \mathbf{W}^-(x_1^{m-1}, x_1^m)\vec{Q}^-(x_1^m). \quad (3.13)$$

When each round trip (downgoing extrapolation followed by upgoing extrapolation) is complete, it adds one order of two-way scattering. It is equivalent to the evaluation of the Neumann series. In the first round trip, $\vec{P}^+(x_1^m)$ and $\vec{P}^-(x_1^m)$ are assumed as zero wavefields for every depth level x_1^m . Then we can summarize this process, according to equations 3.11 and 3.13, as:

$$\vec{P}^+(x_1^m) = \mathbf{W}^+(x_1^m, x_1^0)\vec{S}(x_1^0) + \sum_{n=0}^{m-1} \mathbf{W}^+(x_1^m, x_1^n)\delta\vec{P}(x_1^n), \quad (3.14a)$$

$$\vec{P}^-(x_1^m) = \sum_{n=m+1}^M \underline{\mathbf{W}}^-(x_1^m, x_1^n) \delta \vec{P}(x_1^n), \quad (3.14b)$$

where M is the total samples in depth and

$$\underline{\mathbf{W}}^+(x_1^m, x_1^n) = \prod_{k=n+1}^m \mathbf{W}^+(x_1^k, x_1^{k-1}), \quad (3.15a)$$

$$\underline{\mathbf{W}}^-(x_1^m, x_1^n) = \prod_{k=n-1}^m \mathbf{W}^-(x_1^k, x_1^{k+1}). \quad (3.15b)$$

The user selects how many round trips will be evaluated.

3.2 Velocity perturbations

In JMI the phase of the wavefield only depends on the velocity model, which is related to the extrapolation operator \mathbf{W} . JMI applies perturbation theory to the extrapolation operator to extract information on the velocity perturbation, which will be used in a gradient descent scheme to minimize the difference between the modeled and recorded data in the least-squares sense. The algorithm is written in terms of the contrast parameter $\zeta(x_1, x_2)$, defined as:

$$\zeta(x_1, x_2) = 1 - \frac{c_0^2(x_1, x_2)}{c^2(x_1, x_2)}, \quad (3.16)$$

where $c_0(x_1, x_2)$ is the background velocity and $c(x_1, x_2)$ the current velocity model. A perturbation in the upgoing extrapolation operator is

$$\Delta \mathbf{W}^-(x_1^m, x_1^n) = \mathbf{W}^-(x_1^m, x_1^n) - \mathbf{W}_0^-(x_1^m, x_1^n), \quad (3.17)$$

where \mathbf{W}_0^- is the background upgoing extrapolation operator. According equation 3.4 the columns of $\Delta \mathbf{W}^-$ can be expressed as:

$$\Delta \vec{W}_j^-(x_1^m, x_1^n) = \mathcal{F}_{x_2}^{-1} \left[(e^{-ik_{x_1} \Delta x_1} - e^{-ik_0 \Delta x_1}) e^{-ik_{x_2} x_2^j} \right], \quad (3.18)$$

where $k_{x_1} = \sqrt{k_0^2(1 - \zeta) - k_{x_2}^2}$ with $k_0 = \frac{\omega}{c_0}$. Equation 3.18 is the approximation of the derivative in ζ , so that

$$\Delta \vec{W}_j^-(x_1^m, x_1^n) \approx \frac{\partial \vec{W}_j^-}{\partial \zeta} \Bigg|_{\zeta=\zeta_0} \Delta \zeta(x_1^n, x_2^j) \approx \vec{G}_{0j}^-(x_1^m, x_1^n) \Delta \zeta(x_1^n, x_2^j), \quad (3.19)$$

with

$$\vec{G}_{0j}^-(x_1^m, x_1^n) = \mathcal{F}_{x_2}^{-1} \left[i \frac{\Delta z}{2} \frac{k_0^2}{k_{x_1}} e^{-ik_{x_1} \Delta x_1} e^{-ik_{x_2} x_2^j} \right]_{\zeta=\zeta_0}, \quad (3.20)$$

$\zeta_0 = 0$ and $\Delta\zeta = \zeta$. For $k_{x_1} = 0$ (i.e at 90° propagation angle), expression 3.20 becomes unstable. To stabilize the division, a stabilization parameter ε can be used in 3.20:

$$\vec{G}_{0j}^-(x_1^m, x_1^n) = \mathcal{F}_{x_2}^{-1} \left[i \frac{\Delta z}{2} k_0^2 \frac{k_{x_1}^*}{k_{x_1}^* k_{x_1} + \varepsilon} e^{-ik_{x_1} \Delta z} e^{-ik_{x_2} x_2^j} \right]_{\zeta=\zeta_0}, \quad (3.21)$$

where $*$ denotes the complex conjugation. The complete matrix $\Delta\mathbf{W}_0^-(x_1^m, x_1^n)$ is written as:

$$\Delta\mathbf{W}_0^-(x_1^m, x_1^n) = \mathbf{G}_0^-(x_1^m, x_1^n) \boldsymbol{\zeta}(x_1^n), \quad (3.22)$$

where $\boldsymbol{\zeta}(x_1^n)$ is a diagonal matrix with $diag[\boldsymbol{\zeta}(x_1^n)] = \vec{\zeta}(x_1^n)$ being the velocity perturbations at depth level x_1^n . Individually, all velocity perturbations are carried to the surface by

$$\underline{\mathbf{G}}_0^-(x_1^0, x_1^n) = \underline{\mathbf{W}}_0^-(x_1^0, x_1^{n-1}) \mathbf{G}_0^-(x_1^{n-1}, x_1^n), \quad (3.23)$$

and these effects are measured as a scattered wavefield

$$\Delta\vec{P}_\zeta^-(x_1^0) = \sum_{n=1}^N \underline{\mathbf{G}}_0^-(x_1^0, x_1^n) \boldsymbol{\zeta}(x_1^n) \vec{Q}^-(x_1^n). \quad (3.24)$$

3.3 Reflectivity perturbations

This work didn't take account the angle-dependency of reflectivity. The reflectivity operators are diagonal matrix, where $diag[\mathbf{R}_0^\cup(x_1^n)] = \vec{r}(x_1^n)$ are the reflectivities at depth level x_1^n . The effect of perturbation is derived directly from equation 3.6a, where only the reflected amplitude, $\mathbf{R}^\cup P^+(x_1^n)$, matters for the receivers detection at the surface, so that the scattered wavefield at the reflectivity perturbation can be written, in analogy to equation 3.24, as

$$\Delta\mathbf{P}_{\Delta r}^-(x_1^0) = \sum_{n=1}^N \underline{\mathbf{W}}_0^-(x_1^0, x_1^n) \Delta\mathbf{R}(x_1^n) \vec{P}^+(x_1^n). \quad (3.25)$$

3.4 Inversion

JMI estimates the velocity and reflectivity models by minimizing the difference between the observed and modeled data as indicated by the objective function

$$J = \sum_{\omega_1}^{\omega_2} \|\mathbf{P}_{obs}^-(x_1^0) - \mathbf{P}_{mod}^-(x_1^0)\|^2 = \sum_{\omega_1}^{\omega_2} \|\Delta\mathbf{P}^-(x_1^0)\|^2, \quad (3.26)$$

where $[\omega_1, \omega_2]$ is the frequency bandwidth of modeling and the matrix $\mathbf{P}(x_1^0)$ is the collection of all shot measurements $\vec{P}(x_1^0)$. According to equations 3.24 and 3.25, the gradients for velocity and reflectivity, calculated at depth level x_1^n , are:

$$\nabla\vec{\zeta}(x_1^n) = diag \left(\sum_{\omega_1}^{\omega_2} [\underline{\mathbf{G}}_0^-(x_1^n, x_1^0)]^H \Delta\mathbf{P}(x_1^0) [\mathbf{Q}^-(x_1^n)]^H \right) \quad (3.27a)$$

$$\nabla \vec{r}(x_1^n) = \text{diag} \left(\sum_{\omega_1}^{\omega_2} [\mathbf{W}^-(x_1^n, x_1^0)]^H \Delta \mathbf{P}(x_1^0) [\mathbf{P}^+(x_1^n)]^H \right), \quad (3.27b)$$

where the operators \mathbf{W} and \mathbf{G} are evaluated for the current velocity and reflectivity models. The wavefield perturbation predicted by these gradients are:

$$\Delta \vec{P}_{\nabla r}^-(x_1^0) = \sum_{n=1}^N \mathbf{W}^-(x_1^0, x_1^n) \nabla \mathbf{R}(x_1^n) \vec{P}^+(x_1^n) \quad (3.28a)$$

$$\Delta \vec{P}_{\nabla \zeta}^-(x_1^0) = \sum_{n=1}^N \mathbf{G}_0^-(x_1^0, x_1^n) \nabla \zeta(x_1^n) \vec{Q}^-(x_1^n), \quad (3.28b)$$

where $\text{diag}[\nabla \mathbf{R}(x_1^n)] = \nabla \vec{r}(x_1^n)$ and $\text{diag}[\nabla \zeta(x_1^n)] = \vec{\zeta}(x_1^n)$. Finally the update formulas for reflectivity and velocity are:

$$r_{new}(x_1, x_2) = r_{old}(x_1, x_2) + \alpha_r \nabla r(x_1, x_2) \quad (3.29a)$$

$$c_{new}(x_1, x_2) = \frac{c_{old}(x_1, x_2)}{\sqrt{1 - \alpha_\zeta \nabla \zeta(x_1, x_2)}}, \quad (3.29b)$$

with

$$\alpha_\zeta = \arg \min_{\alpha_\zeta} \left(\sum_{\omega_1}^{\omega_2} \|\Delta \mathbf{P}^-(x_1^0) - \alpha_\zeta \Delta \mathbf{P}_{\nabla \zeta}^-(x_1^0)\|^2 \right) \quad (3.30a)$$

$$\alpha_r = \arg \min_{\alpha_r} \left(\sum_{\omega_1}^{\omega_2} \|\Delta \mathbf{P}^-(x_1^0) - \alpha_r \Delta \mathbf{P}_{\nabla r}^-(x_1^0)\|^2 \right). \quad (3.30b)$$

Schematically a single iteration of JMI is summarized as follows:

1. Update the reflectivity for each depth level:

- Update the wavefields $\mathbf{P}^+(x_1^m)$ and $\mathbf{P}^-(x_1^m)$ according the equations 3.14a and 3.14b for the most recent velocity and reflectivity estimates.
- Calculate the reflectivity gradient $\nabla \vec{r}(x_1, x_2)$, equation 3.27b.
- Calculate the wavefield perturbation $\Delta \mathbf{P}_{\nabla r}^-(x_1^0)$, equation 3.28b, and α_r , equation 3.30b
- Update the reflectivity, equation 3.29a

2. Update the velocity for each depth level:

- Update the wavefields $\mathbf{P}^+(x_1^m)$ and $\mathbf{P}^-(x_1^m)$ according the equations 3.14a and 3.14b for the most recent velocity and reflectivity estimated (one round trip).
- Calculate the velocity gradient $\nabla \vec{\zeta}(x_1, x_2)$, equation 3.27a.
- Calculate the wavefield perturbation $\Delta \mathbf{P}_{\nabla \zeta}^-(x_1^0)$, equation 3.28a, and α_ζ , equation 3.30a
- Update the velocity, equation 3.29b

3. Repeat 1 and 2 steps until the objective function J gets minimum or if the maximum iterations has been achieved.

JMI needs to specify the frequency range $[\omega_1, \omega_2]$. Normally the strategy is to choose a low frequency for ω_1 and to increase the bandwidth as optimization proceeds. This strategy guarantees the stability of the algorithm and decreases the computational efforts in the optimization.

4 ENHANCING RESOLUTION IN IMAGE-BASED VELOCITY ESTIMATION USING MORPHOLOGICAL OPERATORS

Seismic inversion methods like Wave-Equation Migration Velocity Analysis (WEMVA) or Joint Migration Inversion (JMI), are powerful tools to estimate velocity models in complex areas. They have the advantage of estimating reflectivity information that can be used in velocity/gradient preconditioning or with structure-oriented regularizations. Structural information can speed up the convergence of the inversion process while the velocity model gets more geological meaning. In this work we use nonlinear filters from the field of morphological image processing to enhance the contrast of JMI velocity solution and, thereby, also improve the final image. First, we briefly describe morphological operators of erosion and dilation in gray-scale, and second, we propose a modified version of them, which incorporates the structural information. Finally, we show the method's effectiveness with some synthetic examples for the 2D case.

Introduction

Edge preserving filtering is a branch of digital image processing theory used to remove random noises without blurring the main features of gray-scale images. Some of these techniques have been applied in seismic image processing to improve the interpretation of the layers, such as applications of the anisotropic diffusion equation by Fehmers and Höcker (2003) and Hale (2009), or bilateral filtering by Hale (2011).

Williamson et al. (2011) also proposed to use the reflectivity image as constraint in Wave-Equation Migration Velocity Analysis [WEMVA, (SAVA; BIONDI, 2004)]. They used the migrated image as prior information to constrain the gradient of the objective function, obeying the geometry of the layers. They show that with this methodology it was possible to speed up the convergence of the minimization process and it also constrained the velocity estimate towards a better geological meaning. On the other hand, the strategy of Williamson et al. (2011) has a drawback. They need to resolve a partial differential equation by a conjugate gradient method, which implies smoothness for the input and output. In other words, the presence of the differential operator still imposes a residual blurring on the velocity enhancement.

Another class of nonlinear filters, derived from mathematical morphology, can improve the methodology proposed by Williamson et al. (2011). Mathematical morphology, as the name suggest, studies the shapes in digital images, using concepts from set theory,

like union and intersection for binary images, and their counterparts in lattice theory called supremum and infimum for gray-scale images. Developed by George Matheron (1975) and Jean Serra (1983), mathematical morphology uses a probe with different sizes or shapes to scan the geometrical structure of a digital image, looking for patterns that match with the morphological operator's prior criterion. The results can enhance or eliminate these patterns and at the same time preserve important features of the original image.

Mathematical morphology is based on two operators called erosion and dilation. These operators have been used in a toggle operator to sharpen the edges in gray-scale images (KRAMER; BRUCKNER, 1975; MEYER; SERRA, 1989). To construct a structure-oriented contrast sharpening operator for seismic models, we followed a different way without a toggle operator. We propose to modify the original model, dividing/multiplying it with a mask function (which has the information of the reflectivity geometry), to make some gaps, which will be used to constrain the erosion (dilation) operation to produce a lower (upper) bound staircase function of the original signal. An average between these two signals produces our desired approximation of the original signal by a staircase function, according to prior information on the step distribution.

Joint Migration Inversion [JMI, Berkhout (2014a), Berkhout (2014b), Berkhout (2014c)], like WEMVA, has the structural information available that can be used as prior information. However, differently from WEMVA, JMI can handle all reflection data including primaries, surface multiples and internal multiples. This makes the method more suitable to imaging a salt basin and consequently produce better reflectivity maps. Because of these advantages, we apply the structure-oriented morphological operators in the JMI scheme, not via preconditioning the gradient of the objective function like Williamson et al. (2011) did, but applying it directly to the estimated velocity.

First, we present the theory of the binary morphological operators, to provide insight into gray-scale morphology theory. We modify the gray-scale operators combining them with a migrated image which carries all structural information to constrain the tomographic models to have more geological meaning. Next, we successfully show the application in JMI to a 2D synthetic dataset.

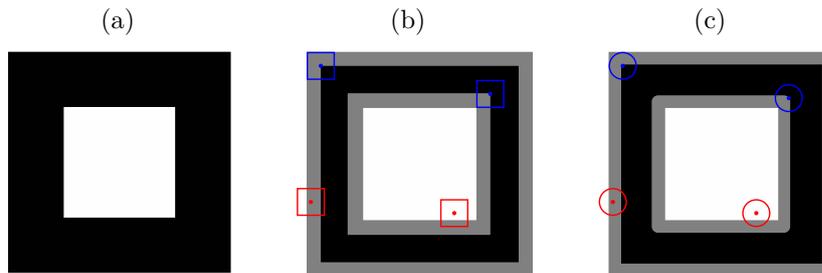
4.1 Erosion and dilation for binary images

According to Maragos (2009), the basic idea of the morphological operators consist of probing an image A using another pixel set called structural element B , to quantify how the structural element fits or does not fit on to image A . If the matching criteria consist of an intersection of B translations, $B_x = \{b + x : b \in B\}$, with image A , written as

$$A \ominus B = \{x : B_x \subset A\} = \bigcap_{b \in B} A_{-b}, \quad (4.1)$$

we have the erosion binary operation, also called Minkowski subtraction (DOUGHERTY; LOTUFO, 2003). The effect of the operator is shrinking the areas of the foreground pixels and enlarge the holes inside these areas, see Figure 4.1.

Figure 4.1 – Original image (a) and the results of binary erosion with (b) a square and (c) a circle structural element. The origin position of the structural element is represented as a dot, and the colors blue and red represent the match and unmatched criteria of equation (4.1), respectively. The binary eroded image (black) is created by setting the foreground pixels in all translations for which B is totally inside of A (i.e the gray area).



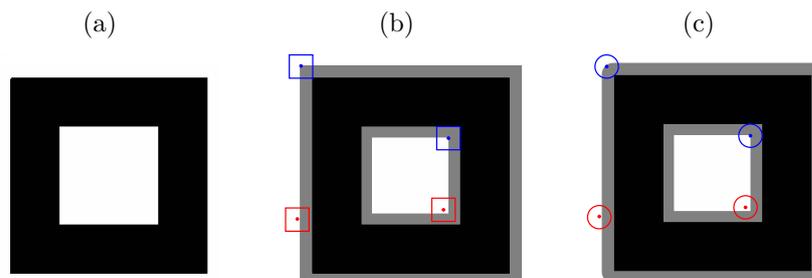
Source: By the author

If the match consist of the union of all B translations with image A , we have the erosion dual operation, called dilation,

$$A \oplus B = \{x : (B^s)_x \cap A \neq \emptyset\} = \bigcup_{a \in A} B_a = \bigcup_{b \in B} A_b, \quad (4.2)$$

where $B^s = \{b : -b \in B\}$ is symmetric to B . As equation 4.2 shows, the dilation is a commutative operation equivalent to the Minkowski addition (MARAGOS, 2009). The binary dilation effect is to enlarge the areas of the foreground pixels and shrink the holes inside these areas, see Figure 4.2.

Figure 4.2 – Original image (a) and the result of binary dilation with (b) a square and (c) a circle structural element. The origin position of the structural element is represented as a dot, and the colors blue and red represent the match and unmatched criteria of equation (4.2), respectively. The binary dilation image (gray) is created by setting the foreground pixels in all translations for which B touches A (black).



Source: By the author

4.2 Erosion and dilation for gray-scale images

To extend the erosion and dilation to gray-scale, we need to identify the binary counterparts in gray-scale. In binary the sets image A and probe B become real functions $f(\mathbf{x})$ and $g(\mathbf{x})$ respectively, where \mathbf{x} is the vector position in an n -dimensional domain \mathbb{E}^n . The intersection \cap and union \cup operations become the infimum \wedge and supremum \vee operation respectively. Then, when a subtraction (addition) between $f(\mathbf{x})$ and $g(\mathbf{x})$ is done in $g(\mathbf{x})$'s domain \mathbb{E} , the smallest (largest) value is selected producing the erosion (dilation) operation. In a complete lattice theory the erosion and dilation operation are given by:

$$(f \ominus g)(\mathbf{x}) = \bigwedge_{\mathbf{y} \in \mathbb{E}^n} f(\mathbf{x} + \mathbf{y}) - g(\mathbf{y}), \quad (4.3a)$$

$$(f \oplus g)(\mathbf{x}) = \bigvee_{\mathbf{y} \in \mathbb{E}^n} f(\mathbf{x} - \mathbf{y}) + g(\mathbf{y}). \quad (4.3b)$$

In fact, for practical purposes we need to consider that domain \mathbb{E} is finite, so that infimum and supremum become the minimum and maximum operators respectively. In this work we also consider flat null structural functions (i.e $g(x) = 0$ for all $x \in \mathbb{E}$ and $g(x) = -\infty$ for $x \notin \mathbb{E}$), and sequential applications of erosion and dilation to amplify their effects (DOUGHERTY; LOTUFO, 2003). Thus, equations 4.3a and 4.3b become

$$(f \ominus g)^k(\mathbf{x}) = \min_{\mathbf{y} \in \mathbb{E}^n} f(\mathbf{x} + \mathbf{y}), \quad (4.4a)$$

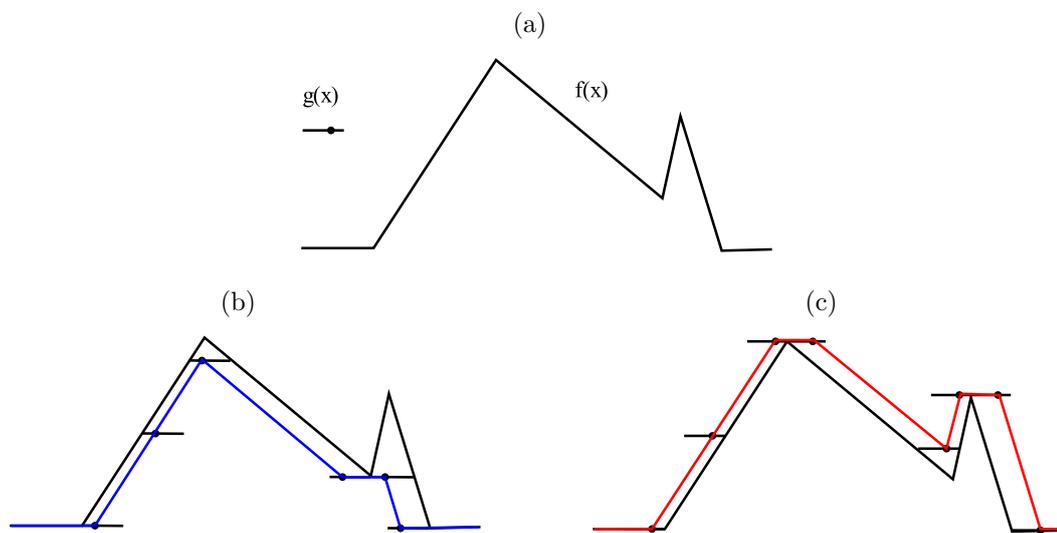
$$(f \oplus g)^k(\mathbf{x}) = \max_{\mathbf{y} \in \mathbb{E}^n} f(\mathbf{x} - \mathbf{y}), \quad (4.4b)$$

where $(f \ominus g)^k(\mathbf{x}) = \underbrace{(((f \ominus g) \ominus g) \ominus g \dots \ominus g)}_{k \text{ times}}$ and $(f \oplus g)^k(\mathbf{x}) = \underbrace{(((f \oplus g) \oplus g) \oplus g \dots \oplus g)}_{k \text{ times}}$ are the k th sequential application.

Figure 4.3 shows the effect of erosion and dilation for a flat null structural function, where the erosion (dilation) operation flattens the valleys (peaks) and expands the minima (maxima) (MARAGOS, 2009).

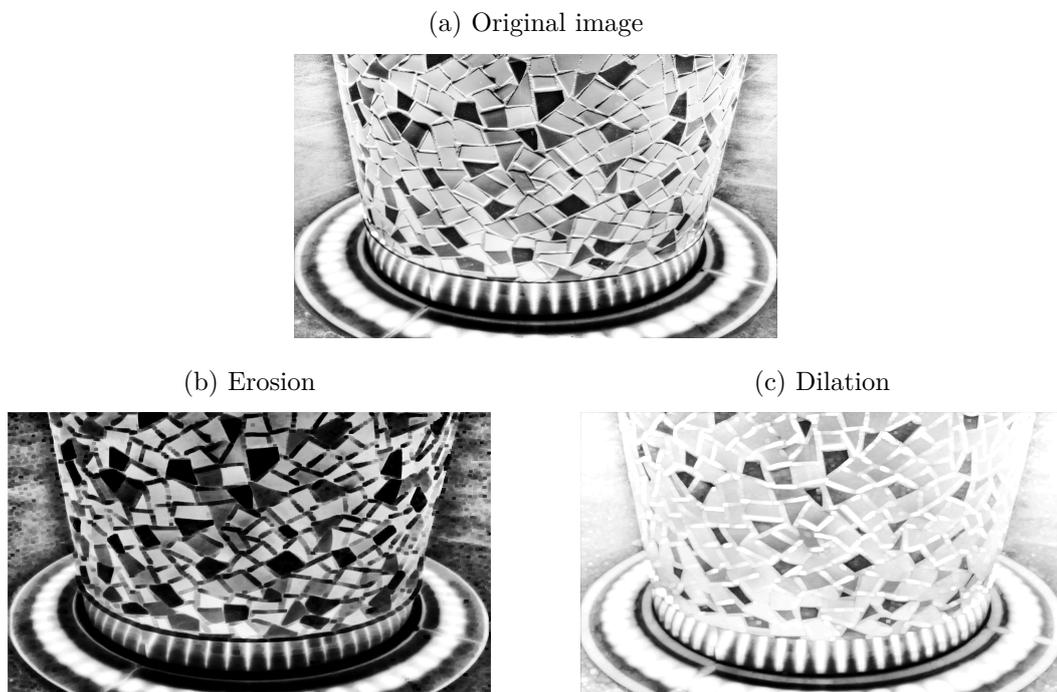
Figure 4.4 shows the standard erosion and dilation operations for a 2D gray-scale image, using a flat null structural function with size 5×5 and $k = 7$. Note that the morphological operation preserves the geometrical shapes of the image.

Figure 4.3 – Erosion and dilation in a 1D gray-scale function. (a) Gray-scale function $f(\mathbf{x})$ and flat null structural function $g(\mathbf{x})$. The dot represent the origin of the $g(\mathbf{x})$'s coordinate system. (b) Erosion (blue) and (c) dilation (red).



Source: By the author

Figure 4.4 – 2D morphological operators applied to a gray-scale image.



Source: By the author

4.3 Enhancing the sharpness with modified erosion and dilation operators

We propose to modify equations 4.4a and 4.4b for our application to construct a staircase function. We divide/multiply the original signal $f(\mathbf{x})$ by another function, called mask $M(\mathbf{x})$, with $0 < M(\mathbf{x}) \leq 1$. $M(\mathbf{x})$ contains information on the layers's edge that will be used to identify the step locations for a staircase function. The modified erosion and dilation are defined as:

$$(f \ominus g)^k(\mathbf{x}) = \min_{\mathbf{y} \in \mathbb{E}^n} [f(\mathbf{x} + \mathbf{y})/H(\mathbf{x} + \mathbf{y})], \quad (4.5a)$$

$$(f \oplus g)^k(\mathbf{x}) = \max_{\mathbf{y} \in \mathbb{E}^n} [f(\mathbf{x} - \mathbf{y})H(\mathbf{x} - \mathbf{y})], \quad (4.5b)$$

with

$$H(\mathbf{x}) = \begin{cases} M(\mathbf{x}), & \text{if } f(\mathbf{x}) \geq 0 \\ M^{-1}(\mathbf{x}), & \text{otherwise} \end{cases}. \quad (4.6)$$

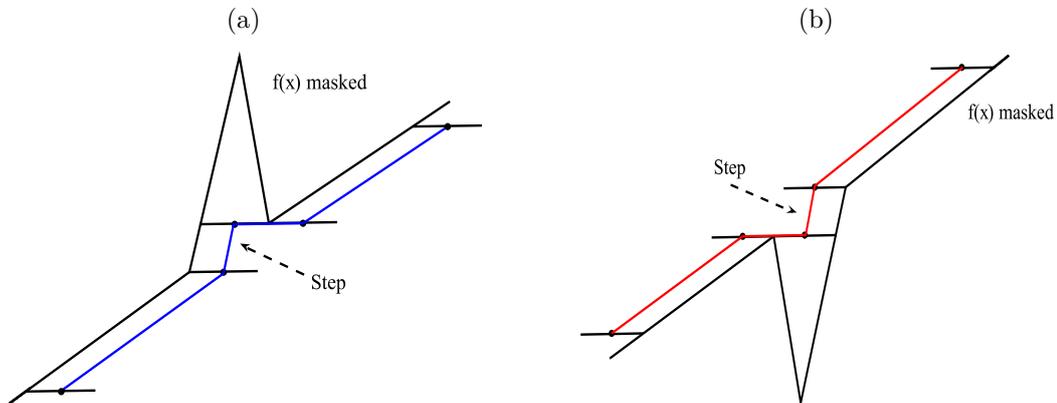
The definition of the mask is not unique, but all of them must consider $M(\mathbf{x}) = 1$ on regions where we desire the velocity/ gradient models to be constant and $0 < M(\mathbf{x}) < 1$ on an impedance change, based on our image. One choice is

$$M(\mathbf{x}) = \exp \left\{ - \left[\frac{R(\mathbf{x})}{\max |R(\mathbf{x})|} \right]^2 \right\}, \quad (4.7)$$

where $R(\mathbf{x})$ is the prior information, the migrated image.

Dividing/ Multiplying the model by the mask will produce gaps in $f(\mathbf{x})$. When the erosion and dilation is applied to these modified models, the structural function follows the deformation highlighting the velocity contrast between the values before and after the gap, producing the steps of the staircase function as shown in Figure 4.5.

Figure 4.5 – Step formation: $f(x)$ masked (black), (a) Erosion (blue) and (b) dilation (red).



Source: By the author

After several applications the steps of the staircase function become sharper and their values move away from the original model. To return to values close to the original model, the solution is to take the average between the lower bound and upper bound staircase function. In other words, the solution we are looking for is

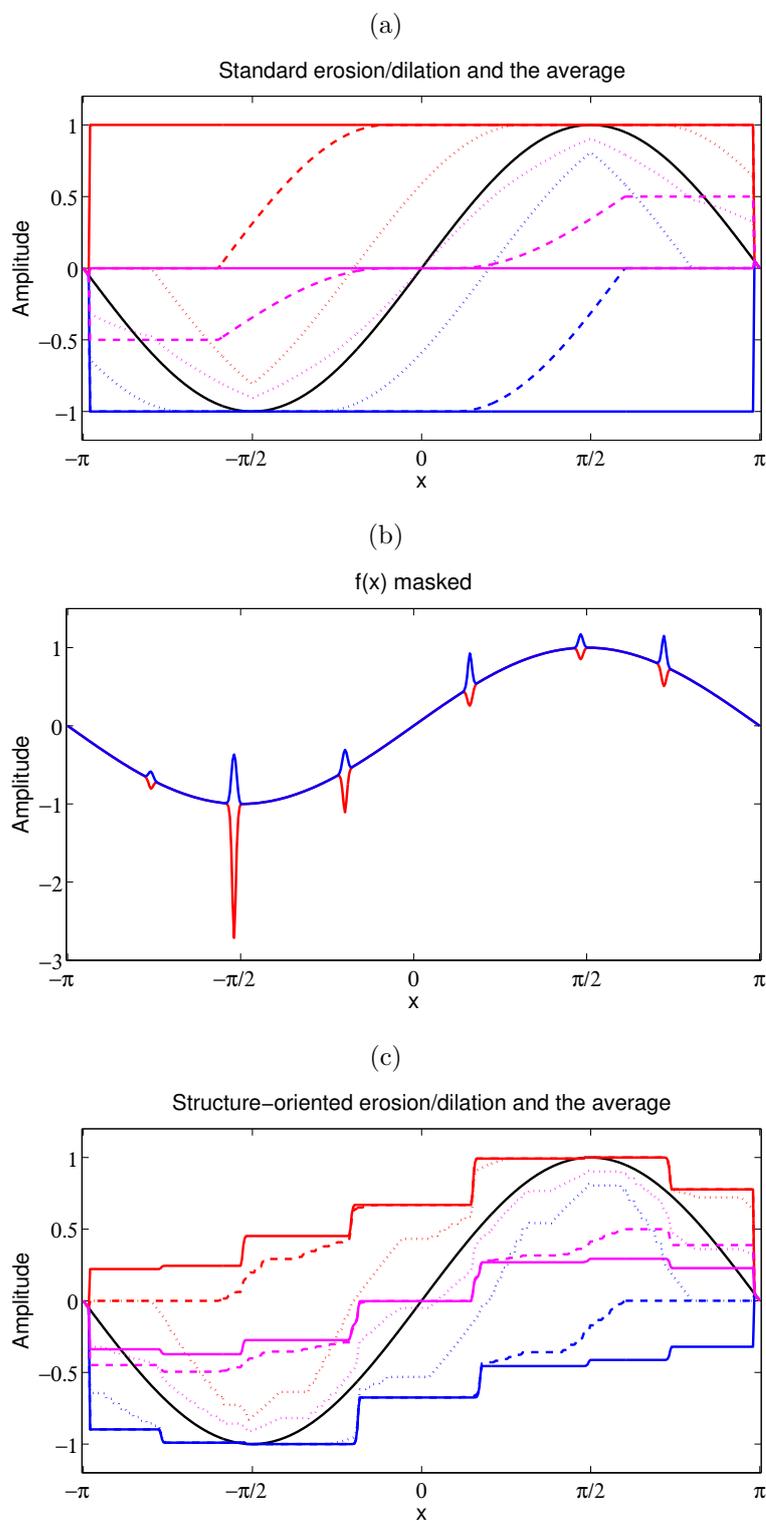
$$A^k(\mathbf{x}) = \frac{1}{2} \left[(f \ominus g)^k(\mathbf{x}) + (f \oplus g)^k(\mathbf{x}) \right]. \quad (4.8)$$

Let's exemplify this procedure by the 1D morphological operation in function $f(x) = \text{sen}(x)$ on the interval $[-\pi, \pi]$ using discrete function f_i with 500 samples. First let's see the normal erosion and dilation applied with probe size equal to 9. The Figure 4.6a shows the behavior for different values of sequential applications. As we can see the erosion (dilation) expand the smallest (highest) values until the minimum (maximum) bound -1.0 ($+1.0$) prior settled.

Taking a prior information $R_i = \sum_j^{11} r_{i+j-6} \exp[-0.1(i-5)^2]$ where $[r_{60}, r_{120}, r_{200}, r_{290}, r_{370}, r_{430}] = [-0.2, 0.5, -0.4, 0.4, 0.2, -0.32]$, putting it into a mask function, and applying $f(x)/H(x)$ and $f(x)M(x)$ will produce the gaps shown in Figure 4.6b. Repeating the previous experiment with these modified models we got the $\text{sen}(x)$ approximation by staircase function, as we can see in Figure 4.6c.

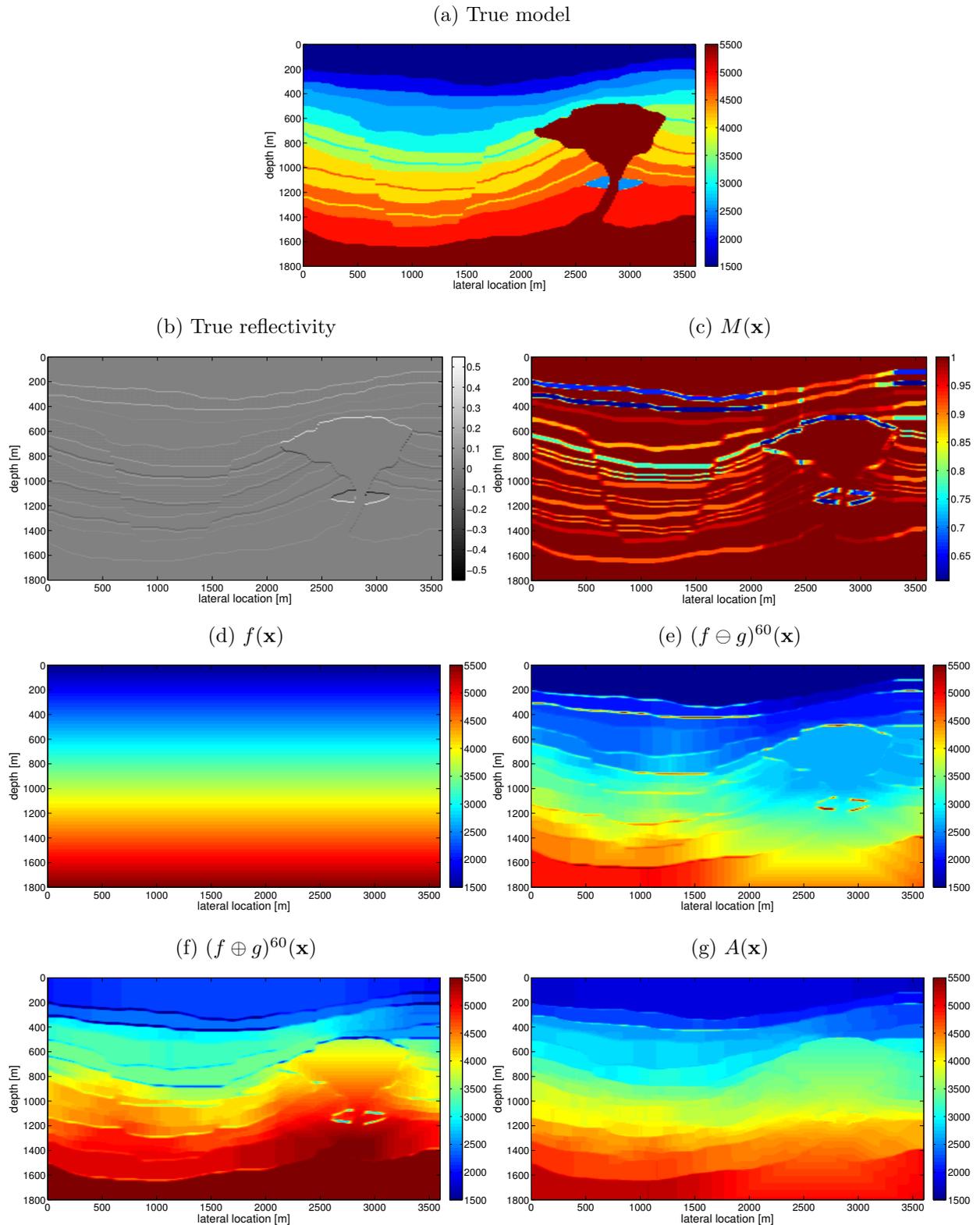
An example for a 2D seismic model is shown in Figure 4.7. Here we used the true reflectivity, Figure 4.7b, taken from the true velocity, 4.7a. We convolved the reflectivity map with a Gaussian filter to create a reflectivity signal gap of the same size of the probe, 3×3 . Then putting it into the mask function expression, normalized by the trace image, gives us the mask map shown in Figure 4.7c. We then applied the structure-oriented morphological operators, starting from a vertical gradient, Figure 4.7d. After 60 sequential applications we got the erosion map, Figure 4.7e, dilation map, Figure 4.7f and the final solution $A(x)$ shown in Figure 4.7g.

Figure 4.6 – 1D morphological operators applied to function $f(x) = \sin(x)$ (black curve): (a) Standard erosion (blue) and dilation (red) and the average function (magenta), for $k = 10$ (dotted line), $k = 30$ (dashed line) and $k = 100$ (continuous line). (b) Constrained curves $f(x)/H(x)$ (blue) and $f(x)H(x)$ (red). (c) Structure-oriented erosion and dilation with the same values of k as in (a).



Source: By the author

Figure 4.7 – Structure-oriented morphological operator example in 2D. (a) True model, (b) true reflectivity, (c) mask function $M(\mathbf{x})$, (d) initial model $f(\mathbf{x})$, (e) eroded model $(f \ominus g)^{60}(\mathbf{x})$, (f) dilated model $(f \oplus g)^{60}(\mathbf{x})$ and (g) the final solution $A(\mathbf{x})$. Color bars are in m/s except the reflectivity model which is dimensionless.



Source: By the author

4.4 Application of morphological operators in the JMI method

Joint Migration Inversion (JMI) simultaneously estimates a seismic velocity model and an image, using all reflection data including primaries, surface multiples and internal multiples. It employs Full Wavefield Modeling (BERKHOUT, 2012; BERKHOUT, 2014a) and a gradient descent scheme to minimize the residual data between modeled data and recorded data, in a least-squares sense.

The morphological operations are introduced in the JMI algorithm as velocity preconditioner, before the new reflectivity model is estimated. The algorithm is simple and doesn't represent a significant additional computational cost to JMI. Appendix D shows pseudo codes representing the implemented version.

The next two subsections show the potential of velocity preconditioning with examples using two synthetic data sets. In both cases the test data was generated with Full Wave Modeling.

4.4.1 2D simple model with strong internal multiples

We applied JMI on a simple 2D model with strong internal multiples. It consists of a lens embedded in a homogeneous background and three horizontal layers as target, see Figure 4.8a. The true reflectivity is calculated as

$$R(z, x) = \mu \frac{f(x_1, x_2) - f(x_1 - \Delta x_1, x_2)}{f(x_1, x_2) + f(x_1 + \Delta x_1, x_2)}, \quad (4.9)$$

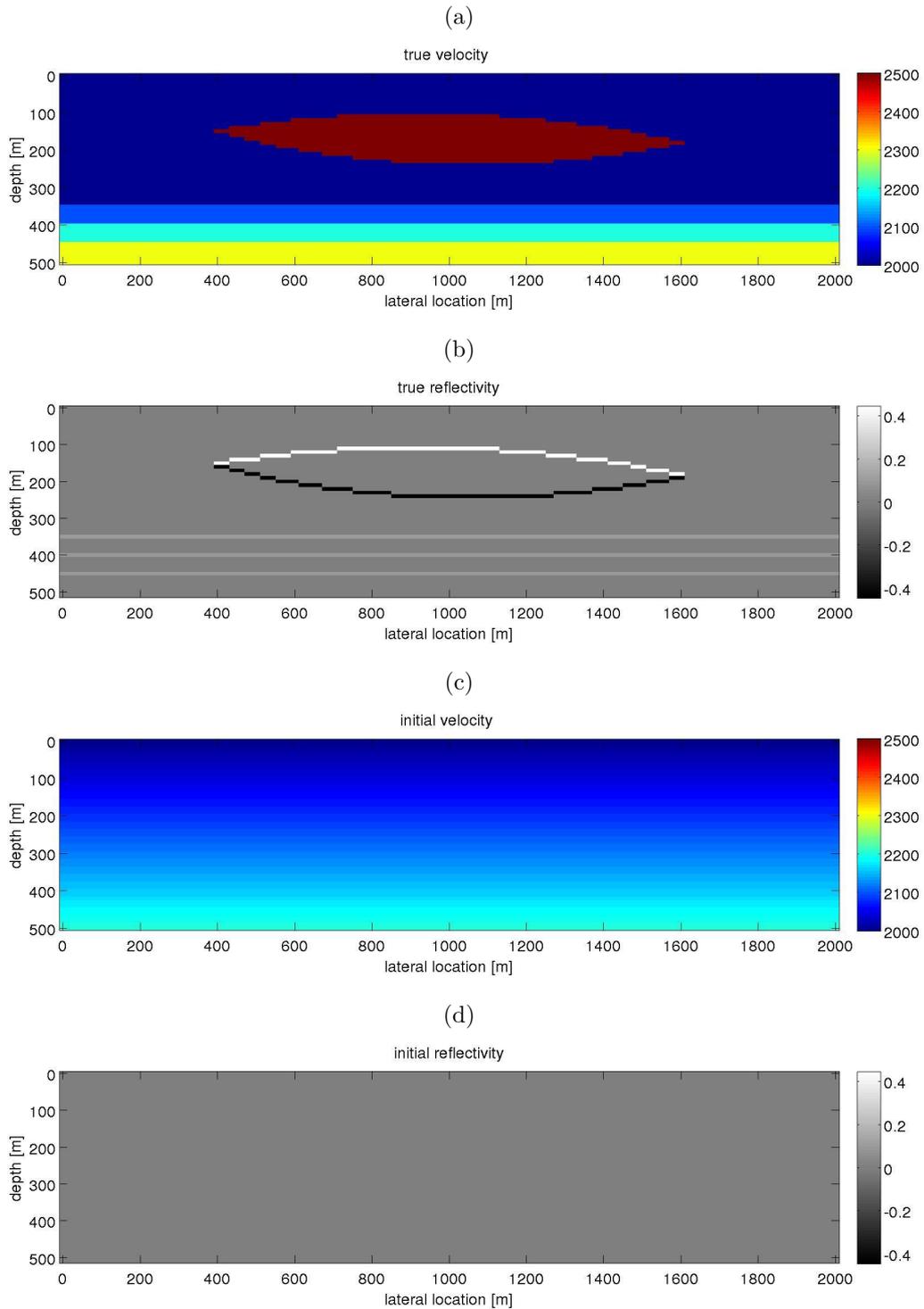
where $f(x_1, x_2)$ is the velocity model, shown in Figure 4.8b, and μ is a gain to improve the image amplitude contrast ($\mu = 4.0$). The JMI process starts with a vertical velocity gradient (Figure 4.8c) and a homogeneous zero reflectivity model Figure (4.8d), and stops after 45 iterations. It uses a Ricker wavelet with bandwidth between 5 Hz and 40 Hz.

Figure 4.9 shows results using standard JMI. We can see that the method is effective to estimate the reflectivity, shown in Figures 4.9a and 4.9c (right). Unfortunately, like other image inversion methods, there is ambiguity in the velocity estimation. In this case the velocity looks like a smoothed version of the true velocity, see Figures 4.9b and 4.9c (left). The convergence can be confirmed by means of a minimum data residual (Figure 4.9d).

The smoothness of the velocity model reflects the linear treatment given to the velocity perturbation in JMI. Including nonlinear filters like morphological operators in the JMI scheme. We obtained a very similar reflectivity, see Figure 4.10a and 4.10c (right), but with a better resemblance between optimized and true velocities, as visible in Figure 4.10b and 4.10c (left). The residual and objective function are shown in Figures 4.10d and

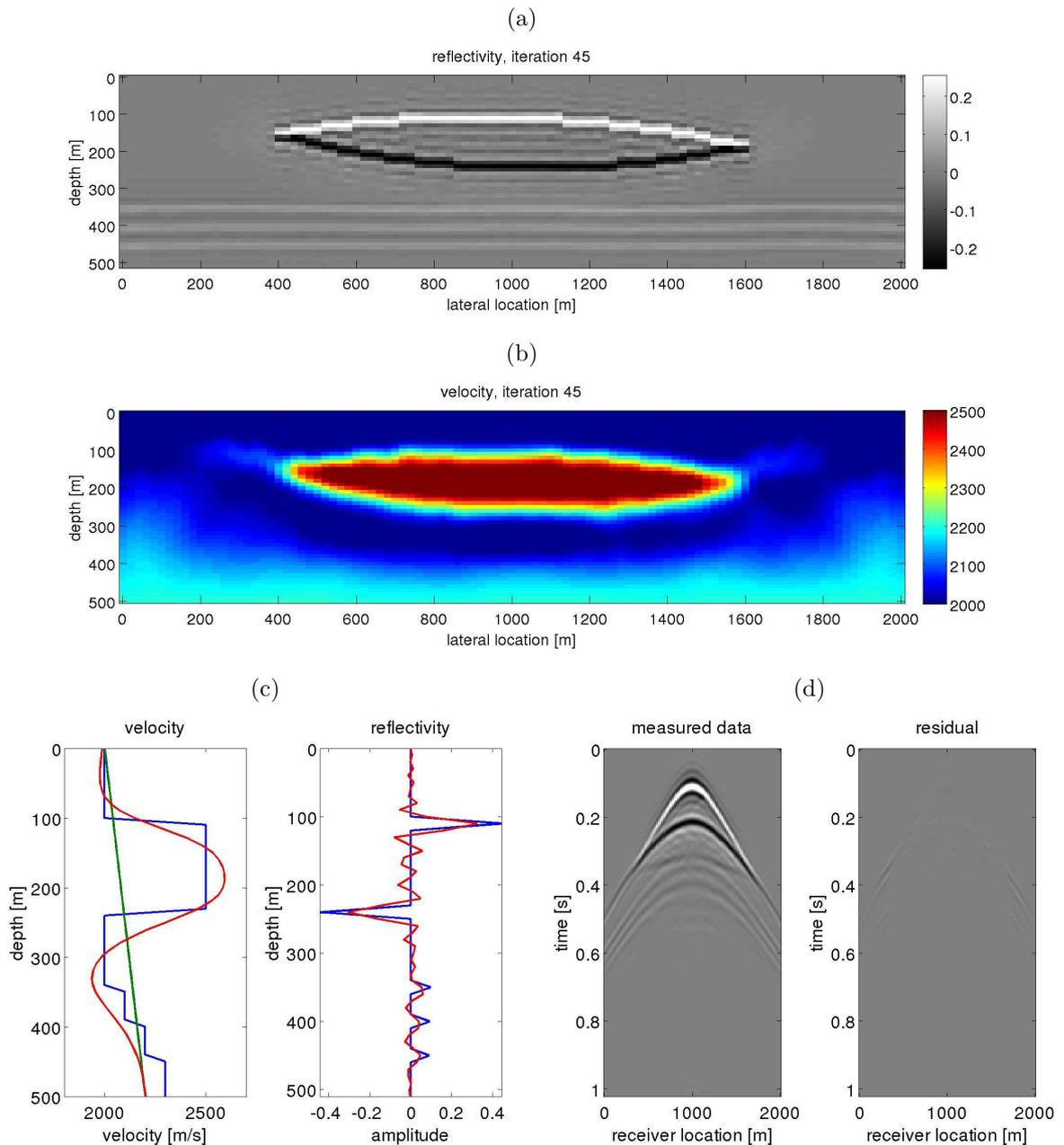
4.10e, respectively. To get these results, we used $g(\mathbf{x})$ of size 3×3 , $k = 100$, the minimum bound $f_{min} = 1700$ m/s and maximum bound $f_{max} = 2800$ m/s.

Figure 4.8 – Subsurface model for JMI test. (a) True velocity model, (b) true reflectivity, (c) initial velocity model and (d) initial reflectivity. Color bars are in m/s except the reflectivity model which is dimensionless.



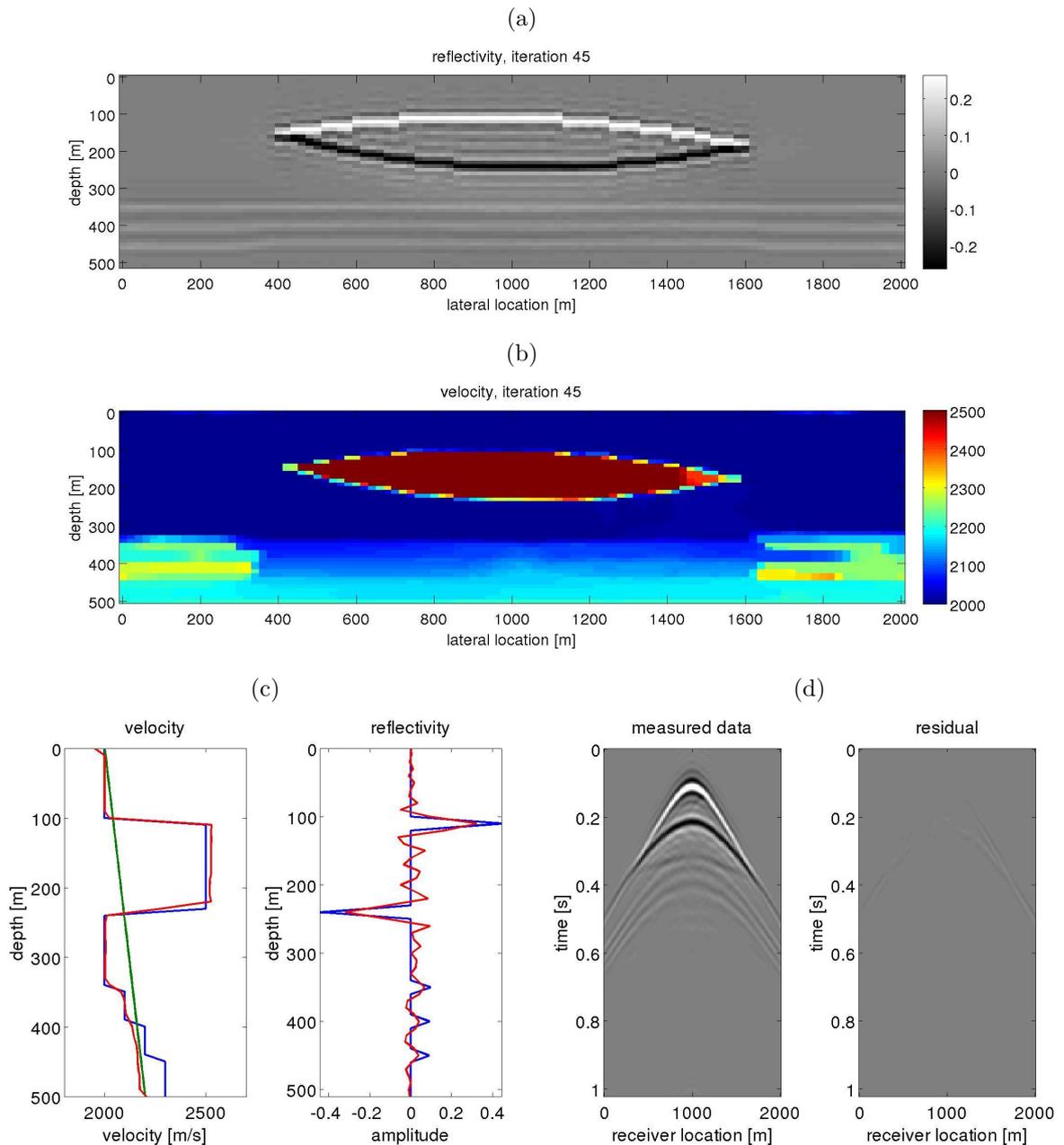
Source: By the author

Figure 4.9 – Standard JMI results. (a) Reflectivity; (b) velocity; (c) profiles at $x = 1000$ m: velocity (left), true velocity (blue), initial velocity (green) and final velocity (red); reflectivity (right), true reflectivity (blue) and final reflectivity (red); (d) measured data (left) and the residual data (right). Color bars are in m/s except the reflectivity model which is dimensionless.



Source: By the author

Figure 4.10 – JMI results with morphological constraints. (a) Reflectivity; (b) velocity; (c) profiles at $x = 1000$ m: velocity (left), true velocity (blue), initial velocity (green) and final velocity (red); reflectivity (right), true reflectivity (blue) and final reflectivity (red); (d) measured data (left), residual data (right). Color bars are in m/s except the reflectivity model which is dimensionless.



Source: By the author

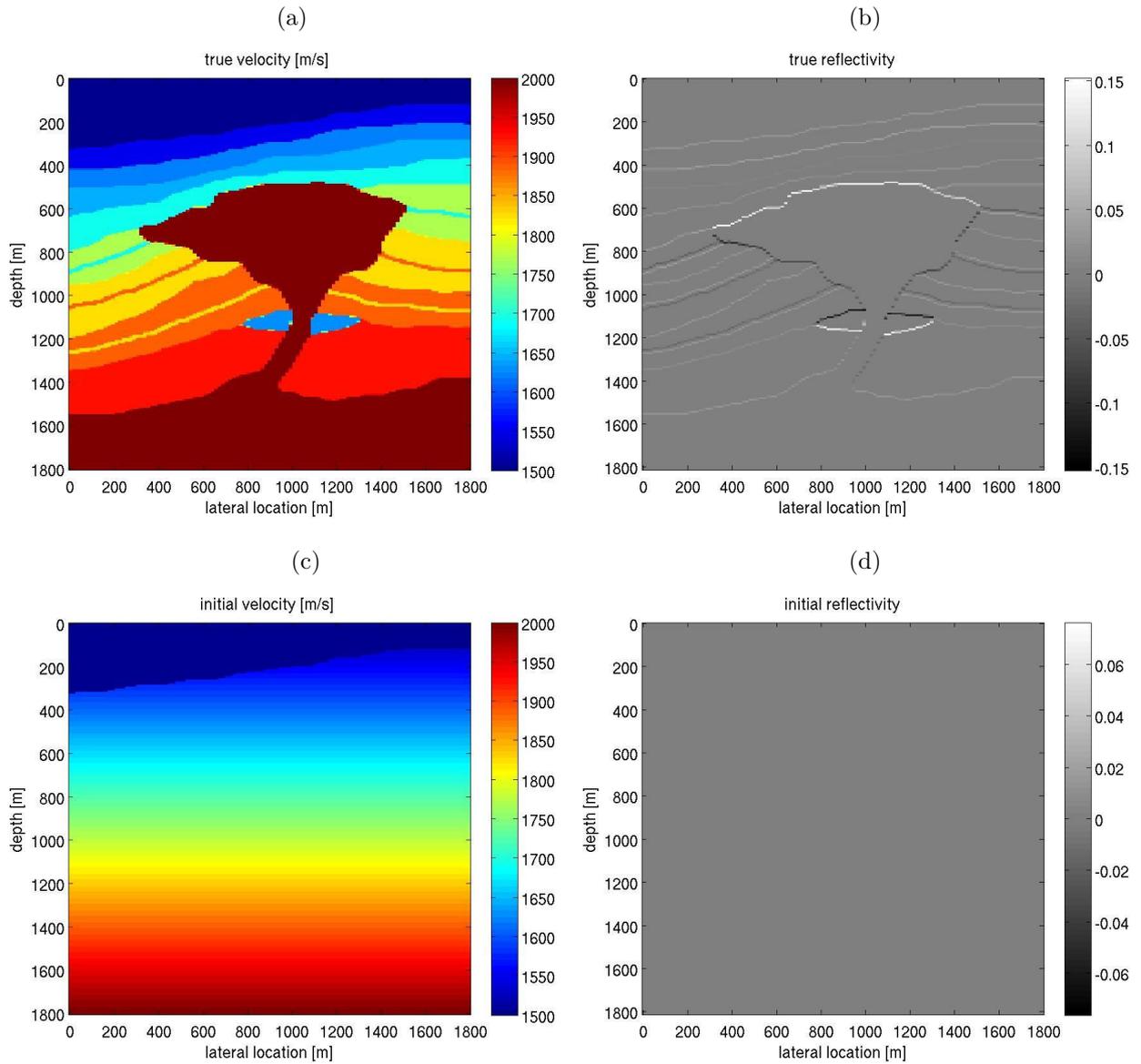
4.4.2 2D complex model with strong internal multiples

Next, we repeat the last experiment, but changing the velocity model for another one that is more realistic. The new velocity again has a high velocity region in the middle, with a complex heterogeneous background and low velocity target on the bottom, as displayed in Figure 4.11a. The true reflectivity, from equation 4.9, is shown in Figure 4.11b (using $\mu = 1.5$). The data was modeled with a Ricker wavelet containing frequencies between 5 Hz and 60 Hz. Like the previous experiment we started from a simple gradient velocity model, Figure 4.11c, a homogeneous zero reflectivity (Figure 4.11d) and the algorithm closes after 400 iterations, maximum iterations prior established by user.

The JMI results without constraints are shown in Figure 4.12. Like in the previous experiment, JMI was effective to estimate the reflectivity, see Figures 4.12a and 4.12c (right), but the final velocity model still represents a smoothed version of the true one, see Figures 4.12b and 4.12c (left). The ambiguity is understood when we see that a smoothed solution is good enough to reduce the residual data, Figure 4.12d.

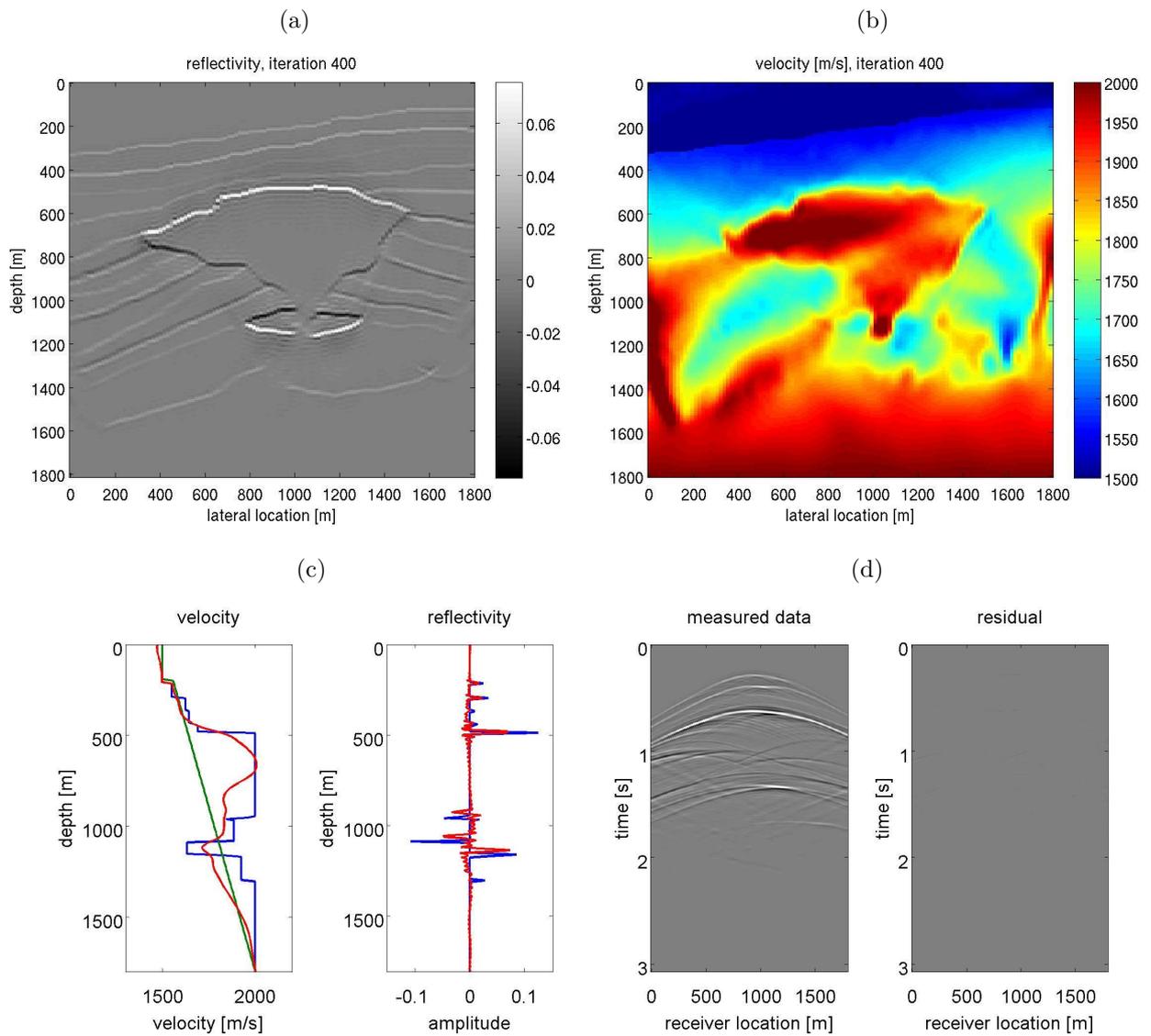
Besides enhancing the sharpness of the velocity contrasts, morphological operators also improve the JMI convergence power, making the method more robust with complex velocity models. This can be observed from a better reflectivity fit in Figures 4.13a and 4.13c (right), and a largely improved velocity fit in Figures 4.13b and 4.13c (left). The residual data decreased more than standard JMI, Figure 4.13d. These results were obtained with a structural function of size 3×3 , with $k = 8$, the minimum bound $f_{min} = 1500$ m/s and maximum bound $f_{max} = 2000$ m/s.

Figure 4.11 – Subsurface model for JMI test. (a) True velocity model, (b) True reflectivity, (c) initial velocity model and (d) initial reflectivity. Color bars are in m/s except the reflectivity model which is dimensionless.



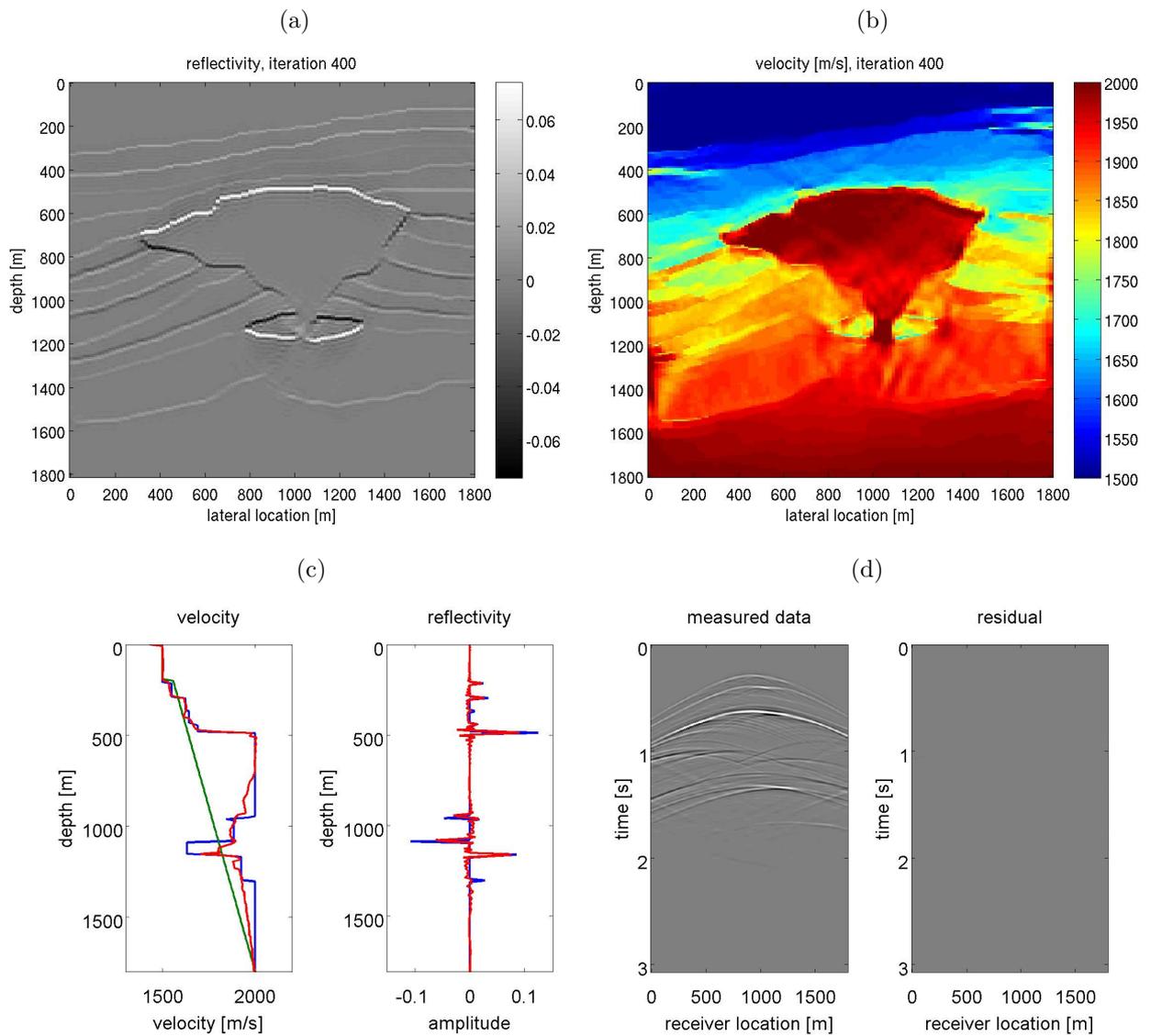
Source: By the author

Figure 4.12 – Standard JMI results. (a) Reflectivity; (b) velocity; (c) profiles at $x = 900$ m: velocity (left), true velocity (blue), initial velocity (green) and final velocity (red); reflectivity (right), true reflectivity (blue) and final reflectivity (red); (d) measured data (left), residual data (right). Color bars are in m/s except the reflectivity model which is dimensionless.



Source: By the author

Figure 4.13 – JMI results with morphological constraints. (a) Reflectivity; (b) velocity; (c) profiles at $x = 900$ m: velocity (left), true velocity (blue), initial velocity (green) and final velocity (red); reflectivity (right), true reflectivity (blue) and final reflectivity (red); (d) measured data (left), residual data (right). Color bars are in m/s except the reflectivity model which is dimensionless.



Source: By the author

4.5 Conclusion

To improve the geological meaning of JMI velocity solutions, we propose to precondition the inversion with nonlinear filters as used in morphological image processing theory. We combined the gray-scale erosion and dilation operators with the JMI reflectivity model, to produce a staircase function wherein each step matches with the reflectivity position. The method improves the layer definition in the velocity model, the JMI convergence with complex models and provides an improved image and smaller data residuals.

The results of our numerical experiments suggest the method can be applied to any problem that involves filling and homogenization of well-defined regions in a digital image. The success of homogenization and the definition of the contrast between the region boundaries clearly depends on the quality of edge detection. If the structures are not well defined in the migrated images the morphological operators might failed to enforce homogeneity and produce blurred velocity regions, although the migrated images are not always compromised. For future developments of this method a more detailed study is necessary about how to define the mask function given a migrated image. Another topic for further investigation is the selection of structure adapted probes which can improve the model resolution.

5 GENERAL CONCLUSIONS

This thesis discussed structural regularization techniques and its applications to velocity model building. Two new preconditioning techniques for wave-equation inversion were proposed and validated in WEMVA and JMI. In order to better condition WEMVA, the cross-gradient between the velocity wavefield and the image intensity was applied to enforce correlation between the estimated velocity models and the migrated image. To improve resolution in JMI a new type of regularization, based on morphological operators, was proposed to enforce homogenization and sharp contrasts across major velocity discontinuities.

We redefine the traditional objective functions of Stack Power and Differential in a compact function, which combines SP and DS functions with cross-gradient regularization, without adjustment of a regularization parameter. That strategy adds to WEMVA a new information of zero-offset, which enforces the velocity model contrasts obey the reflectivity distribution from migrated image. This new information increase the short wavenumber components of the velocity model, and is indicated to be applied when the long wavenumber components is already fitted well. To include the cross-gradient regularization from the initial of optimization process, we proposed to use the maximization of Partial Stack Power, an objective function that combines the Stack Power and Differential Semblance optimization, multiplying it with its cross-gradient version. As the inversion proceeds and the energy becomes more focused the cross-gradient factor is enforced, in order to avoid premature convergence to local minima. This analysis uses naturally the structural information in objective function gradient calculus, which avoids an extra filtering by anisotropic diffusion equation at objective function gradient model. The experiments on Marmousoft dataset indicate that the proposed method improved the convergence of the objective function, resulting in focused common image gathers and a velocity model correlated with the image. The RTM image computed with the estimated velocity model is very close to the image using the exact velocity model except in regions compromised by lack of illumination.

Joint migration inversion separates the short and long wavenumber components of subsurface model in terms of reflectivity and velocity models. Differently from WEMVA, JMI can handle with all kind of scattering information, which possibility no limitation with velocity contrasts as WEMVA do. To increase the velocity sharpening, we proposed to use the erosion and dilation operators modified. The proposal is divide/multiply the input model with a mask function, which is a Gaussian function of the reflectivity map. The application produces gaps or channels in the velocity model. The erosion and dilation detect this features preserving their geometry and at the same time homogenize the regions

inside of the layers. An upper and lower bound staircase functions are produced. An average between these two limits is taken in sense to re-approximate the values to the original model. The result is an approximation of original model by staircase function where the steps coincide with reflector position. The experiments in two synthetic data shown the increase of velocity resolution, defining well thin layers and the model targets. The erosion and dilation do not need any modeling of differential equation. They just use a simple algorithm of convolution.

The application of cross-gradient regularization and the morphological operators of erosion and dilation are not mutually excluding. But the user need to take care about the limitations of inversion method and the prior conditions that this regularizations impose. WEMVA image condition, for example, does not account the multiple scattering. Impose sharp contrast of velocity can bring it inaccurate velocity update. As well as the application of cross-gradient regularization proposed for WEMVA in JMI will not work. because JMI has a data driven objective function. In both cases of regularization is assumed that the layer need to be constant, then the problems that involve shallow gas, which there is not a reflector, the regularization will fail.

Bibliography

- BERKHOUT, A. Combining full wavefield migration and full waveform inversion, a glance into the future of seismic imaging. *Geophysics*, v. 77, n. 2, p. S43–S50, 2012.
- BERKHOUT, A. Review Paper: an outlook on the future of seismic imaging, Part I: forward and reverse modelling. *Geophysical Prospecting*, v. 62, n. 5, p. 911–930, sep 2014.
- BERKHOUT, A. Review Paper: an outlook on the future of seismic imaging, Part II: Full-Wavefield Migration. *Geophysical Prospecting*, v. 62, n. 5, p. 931–949, sep 2014.
- BERKHOUT, A. Review Paper: an outlook on the future of seismic imaging, Part III: Joint Migration Inversion. *Geophysical Prospecting*, v. 62, n. 5, p. 950–971, sep 2014.
- BILLETTE, F. et al. Practical aspects and applications of 2D stereotomography. *Geophysics*, v. 68, n. 3, p. 1008–1021, 2003.
- BIONDI, B.; SYMES, W. W. Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging. *Geophysics*, v. 69, n. 5, p. 1283–1298, 2004.
- BROYDEN, C. G. A new double-rank minimization algorithm. *Notices of the American Mathematical Society*, v. 16, p. 670, 1969.
- BYRD, R. et al. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, v. 16, n. 5, p. 1190–1208, 1995.
- CHAVENT, G.; JACEWITZ, C. A. Determination of background velocities by multiple migration fitting. *Geophysics*, v. 60, n. 2, p. 476–490, 1995.
- CLAERBOUT, J. F. Toward a unified theory of reflector mapping. *Geophysics*, v. 36, n. 3, p. 467–481, 1971.
- DAVIDON, W. Variable metric method for minimization. *SIAM Journal on Optimization*, v. 1, n. 1, p. 1–17, 1991.
- DOUGHERTY, E.; LOTUFO, R. *Hands-on morphological image processing*. [S.l.]: SPIE Press, 2003.
- FEHMERS, G. C.; HÖCKER, C. F. W. Fast structural interpretation with structure-oriented filtering. *Geophysics*, v. 68, n. 4, p. 1286–1293, 2003.
- FEI, W.; WILLIAMSON, P. On the gradient artifacts in migration velocity analysis based on differential semblance optimization. In: *SEG Denver 2010 Annual Meeting*. [S.l.: s.n.], 2010. p. 4071–4076.
- FLETCHER, R. A new approach to variable metric algorithms. *The Computer Journal*, Oxford University Press, v. 13, n. 3, p. 317–322, jan 1970.
- FLETCHER, R.; POWELL, M. J. D. A rapidly convergent descent method for minimization. *The Computer Journal*, v. 6, n. 2, p. 163–168, aug 1963.
- FREGOSO, E.; GALLARDO, L. A. Cross-gradients joint 3D inversion with applications to gravity and magnetic data. *Geophysics*, v. 74, n. 4, p. L31–L42, 2009.

- GOLDFARB, D. A family of variable-metric methods derived by variational means. *Mathematics of Computation*, American Mathematical Society, v. 24, n. 109, p. 23–26, jan 1970.
- HALE, D. *Structure-oriented smoothing and semblance*. Technical Report CWP-635, Center for Wave Phenomena, Colorado School of Mines, 2009.
- HALE, D. *Structure-oriented bilateral filtering*. Technical Report CWP-695, Center for Wave Phenomena, Colorado School of Mines, 2011.
- KAPIO, J. P. et al. Inverse problems with structural prior information. *Inverse Problems*, v. 15, n. 3, p. 713, 1999.
- KRAMER, H. P.; BRUCKNER, J. B. Iterations of a non-linear transformation for enhancement of digital images. *Pattern Recognition*, v. 7, n. 1–2, p. 53–58, 1975.
- LAMELOISE, C. A.; CHAURIS, H.; NOBLE, M. Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis. *Geophysical Prospecting*, v. 63, n. 2, p. 391–404, 2015.
- MACIEL, J. S.; COSTA, J. C.; SCHLEICHER, J. Coherence measures in automatic time-migration velocity analysis. *Journal of Geophysics and Engineering*, v. 9, n. 5, p. 569, 2012.
- MARAGOS, P. Morphological filtering. In: *The essential guide to image processing*. [S.l.]: Academic Press, 2009. cap. 13, p. 293–321.
- MATHERON, G. *Random sets and integral geometry*. [S.l.]: John Wiley & Sons Inc, 1975.
- MEYER, F.; SERRA, J. Contrasts and activity lattice. *Signal Processing*, v. 16, n. 4, p. 303–317, 1989.
- NOCEDAL, J.; WRIGHT, S. *Numerical optimization*. 2. ed. New York, NY: Springer, 2006. (Springer series in operations research and financial engineering).
- PLESSIX, R. E. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, v. 167, n. 2, p. 495–503, 2006.
- RICKETT, J. E.; SAVA, P. C. Offset and angle-domain common image-point gathers for shot-profile migration. *Geophysics*, v. 67, n. 3, p. 883–889, 2002.
- SAVA, P.; BIONDI, B. Wave-equation migration velocity analysis. I. Theory. *Geophysical Prospecting*, v. 52, n. 6, p. 593–606, 2004.
- SAVA, P.; FOMEL, S. Time-shift imaging condition in seismic migration. *Geophysics*, v. 71, n. 6, p. S209–S217, 2006.
- SAVA, P.; VASCONCELOS, I. Extended imaging conditions for wave-equation migration. *Geophysical Prospecting*, Blackwell Publishing Ltd, v. 59, n. 1, p. 35–55, 2011.
- SERRA, J. *Image analysis and mathematical morphology*. [S.l.]: Academic Press, 1983.

- SHANNO, D. F. Conditioning of quasi-Newton methods for function minimization. *Mathematics of computation*, v. 24, n. 111, p. 647–656, jul 1970.
- SHEN, P.; SYMES, W. W. Automatic velocity analysis via shot profile migration. *Geophysics*, v. 73, n. 5, p. VE49–VE59, 2008.
- SHEN, P.; SYMES, W. W. Horizontal contraction in image domain for velocity inversion. *Geophysics*, v. 80, n. 3, p. R95–R110, 2015.
- STAAL, X.; VERSCHUUR, D. Joint migration inversion: Implementation and examples. In: *DELPHI - The multiple removal and structural imaging project*. [S.l.]: QuantesArtoos, 2014. XXV, cap. 15, p. 265–285.
- SYMES, W. W.; CARAZZONE, J. J. Velocity inversion by differential semblance optimization. *Geophysics*, v. 56, n. 5, p. 654–663, 1991.
- TIKHONOV, A. N.; ARSENIN, V. Y. *Solutions of ill-posed problems*. [S.l.]: W.H. Winston, 1977.
- TRYGGVASON, A.; LINDE, N. Local earthquake (LE) tomography with joint inversion for P- and S-wave velocities using structural constraints. *Geophysical Research Letters*, v. 33, n. 7, p. L07303, 2006.
- VERSTEEG, R. J.; GRAU, G. The marmousi experience. *Proc. EAGE Workshop on Practical Aspects of Seismic Data Inversion. Copenhagen*, 1991.
- VIRIEUX, J.; OPERTO, S. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, v. 74, n. 6, p. WCC1–WCC26, 2009.
- WEIBULL, W. W.; ARNTSEN, B. Automatic velocity analysis with reverse-time migration. *Geophysics*, v. 78, n. 4, p. S179–S192, 2013.
- WILLIAMSON, P. et al. Regularization of wave-equation migration velocity analysis by structure-oriented smoothing. In: *SEG San Antonio 2011 Annual Meeting*. [S.l.: s.n.], 2011. p. 3877–3881.
- YANG, T.; SAVA, P. Moveout analysis of wave-equation extended images. *Geophysics*, v. 75, n. 4, p. S151–S161, 2010.
- ZHANG, Y.; SHAN, G. Wave-equation migration velocity analysis using partial stack-power maximization. In: *SEG Houston 2013 Annual Meeting*. [S.l.: s.n.], 2013. p. 4847–4852.
- ZHU, C. et al. *L-BFGS-B - Fortran subroutines for large-scale bound constrained Optimization*. [S.l.], 1994.

APPENDICES

APPENDIX A – OBJECTIVE FUNCTION GRADIENT

This appendix shown the calculation of objective function gradient by adjoint state method (PLESSIX, 2006), but first let's summarize the notation used here. We use a discrete notation for easy understanding of the algorithm implementation. A 2D wave field $P(\mathbf{x}, t)$ is identified as $P_{i,j}^k$, where i is the i th sample in x_1 direction, j is the j th sample in x_2 direction and k is the k th time modeling sample. For example, equation 2.1 can be approximated by the following finite difference scheme

$$P_{i,j}^{k+1} - 2P_{i,j}^k + P_{i,j}^{k-1} - \Delta t^2 c_{i,j}^2 \left[\frac{1}{\Delta x_1^2} \sum_{q=0}^{N-1} a_q^{(2)} (P_{i+q,j}^k + P_{i-q,j}^k) + \frac{1}{\Delta x_2^2} \sum_{q=0}^{N-1} a_q^{(2)} (P_{i,j+q}^k + P_{i,j-q}^k) \right] = F_{i,j}^k, \quad (\text{A.1})$$

where $a^{(2)}$ are the finite difference coefficients for the second-order derivative, N is the order of the finite difference operator, Δx_1 , Δx_2 are the space sample intervals and Δt is the time sample interval.

The adjoint state method applied in this thesis tries to minimize the objective function Φ by the minimization of the associated Lagrangian $\mathcal{L} = \mathcal{L}(D_{i,j}^{k,s}, U_{i,j}^{k,r}, \Lambda_{i,j}^{k,s}, \Lambda_{i,j}^{k,r}, c_{i,j})$, with the D and U fields subject to obey the acoustic wave equation with constant density

$$\begin{aligned} \text{Min } \mathcal{L} = & \Phi + \sum_{s=1}^{N_s} \sum_{j=1}^{N_{x_2}} \sum_{i=1}^{N_{x_1}} \sum_{k=1}^{N_t} \Lambda_{i,j}^{k,s} \left[D_{i,j}^{k+1,s} - 2D_{i,j}^{k,s} + D_{i,j}^{k-1,s} - c_{i,j}^2 \Delta t^2 \nabla^2 D_{i,j}^{k,s} - F_{i,j}^{k,s} \right] \\ & + \sum_{s=1}^{N_s} \sum_{r=1}^{N_r(s)} \sum_{j=1}^{N_{x_2}} \sum_{i=1}^{N_{x_1}} \sum_{k=1}^{N_t} \Lambda_{i,j}^{k,r} \left[U_{i,j}^{k+1,r} - 2U_{i,j}^{k,r} + U_{i,j}^{k-1,r} - c_{i,j}^2 \Delta t^2 \nabla^2 U_{i,j}^{k,r} - F_{i,j}^{k,r} \right], \end{aligned} \quad (\text{A.2})$$

where the index s and r refers to sth shot and rth receiver, the fields $\Lambda_{i,j}^{k,s}$ and $\Lambda_{i,j}^{k,r}$ are the adjoint states of $D_{i,j}^{k,s}$ and $U_{i,j}^{k,r}$, respectively.

Before finding out the expression of $\delta\mathcal{L}$, let's find out the generic expression of $\delta\Phi$. An ODCIG for a 2D model can be represent as a data cube, which saves 3 components, i, j, h , where h is the subsurface offset index. In this case we can write the objective function as

$$\Phi(I, c) = \sum_{i=1}^{N_{x_1}} \sum_{j=1}^{N_{x_2}} \sum_{h=-N_{h_2}}^{N_{h_2}} \mathcal{K}_{i,j,h}(I_{i,j,h}; c_{i,j}), \quad (\text{A.3})$$

where \mathcal{K} is the kernel of the function. A perturbation in Φ is described by

$$\delta\Phi(I, c) = \sum_{i=1}^{N_{x_1}} \sum_{j=1}^{N_{x_2}} \sum_{h=-N_{h_2}}^{N_{h_2}} \frac{\partial \mathcal{K}_{i,j,h}}{\partial I} \delta I_{i,j,h} + \frac{\partial \mathcal{K}_{i,j,h}}{\partial c} \delta c_{i,j}. \quad (\text{A.4})$$

Remembering the extended Claerbout image condition, equation 2.3 in [$\mathbf{h} = (0, h_2, 0)$; $\tau = 0$] case, the image perturbation is

$$\delta I_{i,j,h} = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r(s)} \sum_{k=1}^{N_t} \left[\delta D_{i,j-h}^{k,s} U_{i,j+h}^{k,r} + D_{i,j-h}^{k,s} \delta U_{i,j+h}^{k,r} \right]. \quad (\text{A.5})$$

Substituting the equation A.5 in A.4 we get

$$\delta \Phi = \sum_{srkijh} \frac{\partial \mathcal{K}_{i,j+h,h}}{\partial I} U_{i,j+2h}^{k,r} \delta D_{i,j}^{k,s} + \sum_{srkijh} \frac{\partial \mathcal{K}_{i,j-h,h}}{\partial I} D_{i,j-2h}^{k,s} \delta U_{i,j}^{k,r} + \sum_{ijh} \frac{\partial \mathcal{K}_{i,j,h}}{\partial c} \delta c_{i,j}. \quad (\text{A.6})$$

Now let's return to the $\delta \mathcal{L}$ calculus. From equation A.2 we have:

$$\begin{aligned} \delta \mathcal{L} &= \delta \Phi + \sum_{sij} \left[\sum_{k=1}^{N_t} \Lambda_{i,j}^{k,s} \delta D_{i,j}^{k+1,s} - 2 \sum_{k=1}^{N_t} \Lambda_{i,j}^{k,s} \delta D_{i,j}^{k,s} + \sum_{k=1}^{N_t} \Lambda_{i,j}^{k,s} \delta D_{i,j}^{k-1,s} \right] \\ &+ \sum_{srij} \left[\sum_{k=1}^{N_t} \Lambda_{i,j}^{k,r} \delta U_{i,j}^{k+1,r} - 2 \sum_{k=1}^{N_t} \Lambda_{i,j}^{k,r} \delta U_{i,j}^{k,r} + \sum_{k=1}^{N_t} \Lambda_{i,j}^{k,r} \delta U_{i,j}^{k-1,r} \right] \\ &- \Delta t^2 \sum_{sijk} c_{i,j}^2 \Lambda_{i,j}^{k,s} \nabla^2 \delta D_{i,j}^{k,s} - \Delta t^2 \sum_{srijk} c_{i,j}^2 \Lambda_{i,j}^{k,r} \nabla^2 \delta U_{i,j}^{k,r} \\ &- 2\Delta t^2 \sum_{ij} c_{i,j} \left[\sum_{sk} \Lambda_{i,j}^{k,s} \nabla^2 D_{i,j}^{k,s} \right] \delta c_{i,j} - 2\Delta t^2 \sum_{ij} c_{i,j} \left[\sum_{srk} \Lambda_{i,j}^{k,r} \nabla^2 U_{i,j}^{k,r} \right] \delta c_{i,j}. \quad (\text{A.7}) \end{aligned}$$

Remembering that the boundary conditions are equal to zero, it is not necessary to write the expressions for $\delta \Lambda_{i,j}^{k,s}$ and $\delta \Lambda_{i,j}^{k,r}$. Next we need to isolate the terms of $\delta D_{i,j}^{k,s}$ e $\delta U_{i,j}^{k,r}$, but there is a Laplacian operator acting on them. We solved this remembering that the Laplacian is self-adjoint operator. Exemplify for the case of $\delta D_{i,j}^{k,s}$ by finite difference approximation

$$\begin{aligned} \sum_{sijk} c_{i,j}^2 \Lambda_{i,j}^{k,s} \nabla^2 \delta D_{i,j}^{k,s} &= \sum_{sijk} c_{i,j}^2 \Lambda_{i,j}^{k,s} \left[\sum_{l=0}^{N-1} a_l^{(2)} (\delta D_{i+l,j}^{k,s} + \delta D_{i-l,j}^{k,s}) + \sum_{l=0}^{N-1} b_l^{(2)} (\delta D_{i,j+l}^{k,s} + \delta D_{i,j-l}^{k,s}) \right] \\ &= \sum_{sjk} \left[\sum_{l=0}^{N-1} a_l^{(2)} \left(\sum_{p=1}^{N_{x_1}} c_{p,j}^2 \Lambda_{p,j}^{k,s} \delta D_{p+l,j}^{k,s} + \sum_{p=1}^{N_{x_1}} c_{p,j}^2 \Lambda_{p,j}^{k,s} \delta D_{p-l,j}^{k,s} \right) \right] \\ &+ \sum_{sik} \left[\sum_{l=0}^{N-1} b_l^{(2)} \left(\sum_{q=1}^{N_{x_2}} c_{i,q}^2 \Lambda_{i,q}^{k,s} \delta D_{i,q+l}^{k,s} + \sum_{q=1}^{N_{x_2}} c_{i,q}^2 \Lambda_{i,q}^{k,s} \delta D_{i,q-l}^{k,s} \right) \right]. \quad (\text{A.8}) \end{aligned}$$

In the following expression, we reorganize the summations, changing the the index i to p and j to q , then setting separately $i = p \pm l$ e $j = q \pm l$, so that

$$\begin{aligned} \sum_{sijk} c_{i,j}^2 \Lambda_{i,j}^{k,s} \nabla^2 \delta D_{i,j}^{k,s} &= \sum_{sjk} \left[\sum_{l=0}^{N-1} a_l^{(2)} \left(\sum_{i=1+l}^{N_{x_1}+l} c_{i-l,j}^2 \Lambda_{i-l,j}^{k,s} + \sum_{i=1-l}^{N_{x_1}-l} c_{i+l,j}^2 \Lambda_{i+l,j}^{k,s} \right) \right] \delta D_{i,j}^{k,s} \\ &+ \sum_{sik} \left[\sum_{l=0}^{N-1} b_l^{(2)} \left(\sum_{j=1+l}^{N_{x_2}+l} c_{i,j-l}^2 \Lambda_{i,j-l}^{k,s} + \sum_{j=1-l}^{N_{x_2}-l} c_{i,j+l}^2 \Lambda_{i,j+l}^{k,s} \right) \right] \delta D_{i,j}^{k,s}. \quad (\text{A.9}) \end{aligned}$$

Out side of the model domain, the adjoint field should be zero. Regroup again the summations for i and j ,

$$\begin{aligned}
 \sum_{sijk} c_{i,j}^2 \Lambda_{i,j}^{k,s} \nabla^2 \delta D_{i,j}^{k,s} &= \sum_{sijk} \left[\sum_{l=0}^{N-1} a_l \left(c_{i-l,j}^2 \Lambda_{i-l,j}^{k,s} + c_{i+l,j}^2 \Lambda_{i+l,j}^{k,s} \right) \right] \delta D_{i,j}^{k,s} \\
 &+ \sum_{sijk} \left[\sum_{l=0}^{N-1} b_l \left(c_{i,j-l}^2 \Lambda_{i,j-l}^{k,s} + c_{i,j+l}^2 \Lambda_{i,j+l}^{k,s} \right) \right] \delta D_{i,j}^{k,s} \\
 &= \sum_{sijk} \nabla^2 \left[c_{i,j}^2 \Lambda_{i,j}^{k,s} \right] \delta D_{i,j}^{k,s}. \tag{A.10}
 \end{aligned}$$

The identity A.10 is analogous for the case $\delta U_{i,j}^{k,r}$, and doing similarly for index k , we have the expression

$$\begin{aligned}
 \delta \mathcal{L} &= \delta \Phi + \sum_{sijk} \left[\Lambda_{i,j}^{k-1,s} - 2\Lambda_{i,j}^{k,s} + \Lambda_{i,j}^{k+1,s} - \Delta t^2 \nabla^2 \left(c_{i,j}^2 \Lambda_{i,j}^{k,s} \right) \right] \delta D_{i,j}^{k,s} \\
 &+ \sum_{srijk} \left[\Lambda_{i,j}^{k-1,r} - 2\Lambda_{i,j}^{k,r} + \Lambda_{i,j}^{k+1,r} - \Delta t^2 \nabla^2 \left(c_{i,j}^2 \Lambda_{i,j}^{k,r} \right) \right] \delta U_{i,j}^{k,r} \\
 &- 2\Delta t^2 \sum_{ij} c_{i,j} \left[\sum_{sk} \Lambda_{i,j}^{k,s} \nabla^2 D_{i,j}^{k,s} \right] \delta c_{i,j} - 2\Delta t^2 \sum_{ij} c_{i,j} \left[\sum_{srk} \Lambda_{i,j}^{k,r} \nabla^2 U_{i,j}^{k,r} \right] \delta c_{i,j}. \tag{A.11}
 \end{aligned}$$

The final expression for $\delta \mathcal{L}$ is obtained from the substitution of equation A.6 in A.11, resulting in

$$\begin{aligned}
 \delta \mathcal{L} &= \sum_{srijk} \left[\Lambda_{i,j}^{k-1,s} - 2\Lambda_{i,j}^{k,s} + \Lambda_{i,j}^{k+1,s} - \Delta t^2 \nabla^2 \left(c_{i,j}^2 \Lambda_{i,j}^{k,s} \right) + \sum_{h=-N_{h_2}}^{N_{h_2}} \frac{\partial \mathcal{K}_{i,j+h,h}}{\partial I} U_{i,j+2h}^{k,r} \right] \delta D_{i,j}^{k,s} \\
 &+ \sum_{srijk} \left[\Lambda_{i,j}^{k-1,r} - 2\Lambda_{i,j}^{k,r} + \Lambda_{i,j}^{k+1,r} - \Delta t^2 \nabla^2 \left(c_{i,j}^2 \Lambda_{i,j}^{k,r} \right) + \sum_{h=-N_{h_2}}^{N_{h_2}} \frac{\partial \mathcal{K}_{i,j-h,h}}{\partial I} D_{i,j-2h}^{k,s} \right] \delta U_{i,j}^{k,r} \\
 &- 2\Delta t^2 \sum_{ij} c_{i,j} \left[\sum_{sk} \Lambda_{i,j}^{k,s} \nabla^2 D_{i,j}^{k,s} \right] \delta c_{i,j} - 2\Delta t^2 \sum_{ij} c_{i,j} \left[\sum_{srk} \Lambda_{i,j}^{k,r} \nabla^2 U_{i,j}^{k,r} \right] \delta c_{i,j} \\
 &+ \sum_{ijh} \frac{\partial \mathcal{K}_{i,j,h}}{\partial c} \delta c_{i,j}. \tag{A.12}
 \end{aligned}$$

We are interested in the variance δc , so the expressions inside the brackets must be equal zero. With that we recognize the adjoint wave equations for the source field

$$\Lambda_{i,j}^{k+1,s} = 2\Lambda_{i,j}^{k,s} - \Lambda_{i,j}^{k-1,s} + \Delta t^2 \nabla^2 \left(c_{i,j}^2 \Lambda_{i,j}^{k,s} \right) - \sum_{h=-N_{h_2}}^{N_{h_2}} \frac{\partial \mathcal{K}_{i,j+h,h}}{\partial I} U_{i,j+2h}^{k,r} \tag{A.13}$$

and the adjoint wave equation for the receiver field

$$\Lambda_{i,j}^{k+1,r} = 2\Lambda_{i,j}^{k,r} - \Lambda_{i,j}^{k-1,r} + \Delta t^2 \nabla^2 \left(c_{i,j}^2 \Lambda_{i,j}^{k,r} \right) - \sum_{h=-N_{h_2}}^{N_{h_2}} \frac{\partial \mathcal{K}_{i,j-h,h}}{\partial I} D_{i,j-2h}^{k,s}. \tag{A.14}$$

The gradient, for any objective function $\Phi(I, c)$, is written as:

$$\frac{\partial \mathcal{L}}{\partial c_{i,j}} = -2\Delta t^2 c_{i,j} \left[\sum_{s=1}^{N_s} \sum_{k=1}^{N_t} \Lambda_{i,j}^{k,s} \nabla^2 D_{i,j}^{k,s} \right]$$

$$- 2\Delta t^2 c_{i,j} \left[\sum_{s=1}^{N_s} \sum_{r=1}^{N_r(s)} \sum_{k=1}^{N_t} \Lambda_{i,j}^{k,r} \nabla^2 U_{i,j}^{k,r} \right] + \sum_{h=-N_{h_2}}^{N_{h_2}} \frac{\partial \mathcal{K}_{i,j,h}}{\partial c}, \quad (\text{A.15})$$

where $\Lambda_{i,j}^{k,s}$ and $\Lambda_{i,j}^{k,r}$ are given by equations A.13 and A.14 respectively. Specifying the gradient for our proposed objective function, equation 2.16, is necessary to identify the terms $\frac{\partial \mathcal{K}_{i,j,h}}{\partial I}$ and $\frac{\partial \mathcal{K}_{i,j,h}}{\partial c}$. Starting from

$$\Phi = \varphi \psi, \quad (\text{A.16})$$

where we can summarize it as

$$\varphi = \frac{1}{2} \sum_{j=1}^{N_{x_2}} \frac{\bar{\varphi}_{Gj}}{\bar{\varphi}_j}, \quad (\text{A.17})$$

with

$$\bar{\varphi}_{Gj} = \sum_{h=-N_{h_2}}^{N_{h_2}} \sum_{i=1}^{N_{x_1}} G_h^2 I_{i,j,h}^2, \quad (\text{A.18a})$$

$$\bar{\varphi}_j = \sum_{h=-N_{h_2}}^{N_{h_2}} \sum_{i=1}^{N_{x_1}} I_{i,j,h}^2. \quad (\text{A.18b})$$

In the same way, the second factor is given by

$$\psi = \frac{1}{2} \sum_{j=1}^{N_{x_2}} \frac{\bar{\psi}_{\tilde{G}j}}{\bar{\psi}_j} \quad (\text{A.19})$$

where

$$\bar{\psi}_{\tilde{G}j} = \sum_{h=-N_{h_2}}^{N_{h_2}} \sum_{i=1}^{N_{x_1}} \tilde{G}_h^2 [\nabla I_{i,j,h} \cdot \nabla c_{i,j}]^2, \quad (\text{A.20a})$$

$$\bar{\psi}_j = \sum_{h=-N_{h_2}}^{N_{h_2}} \sum_{i=1}^{N_{x_1}} [\nabla I_{i,j,h} \cdot \nabla c_{i,j}]^2. \quad (\text{A.20b})$$

The perturbation of equation A.16 is

$$\begin{aligned} \delta \Phi &= \frac{\psi}{2} \sum_{j=1}^{N_{x_2}} \left\{ \frac{\bar{\varphi}_j \delta \bar{\varphi}_{Gj} - \bar{\varphi}_{Gj} \delta \bar{\varphi}_j}{\bar{\varphi}_j^2} \right\} \\ &+ \frac{\varphi}{2} \sum_{j=1}^{N_{x_2}} \left\{ \frac{\bar{\psi}_j \delta \bar{\psi}_{\tilde{G}j} - \bar{\psi}_{\tilde{G}j} \delta \bar{\psi}_j}{\bar{\psi}_j^2} \right\}, \end{aligned} \quad (\text{A.21})$$

from which we find using equations A.18a, A.18b, A.20a and A.20b,

$$\begin{aligned} \delta \Phi &= \psi \sum_{ijh} \left\{ \frac{\bar{\varphi}_j G_h^2 - \bar{\varphi}_{Gj}}{\bar{\varphi}_j^2} \right\} I_{i,j,h} \delta I_{i,j,h} \\ &+ \varphi \sum_{ijh} \left\{ \frac{\bar{\psi}_j \tilde{G}_h^2 - \bar{\psi}_{\tilde{G}j}}{\bar{\psi}_j^2} \right\} [\nabla I_{i,j,h} \cdot \nabla c_{i,j}] [\nabla \delta I_{i,j,h} \cdot \nabla c_{i,j}] \end{aligned}$$

$$+ \varphi \sum_{ijh} \left\{ \frac{\bar{\psi}_j \tilde{G}_h^2 - \bar{\psi} \tilde{G}_j}{\bar{\psi}_j^2} \right\} [\nabla I_{i,j,h} \cdot \nabla c_{i,j}] [\nabla I_{i,j,h} \cdot \nabla \delta c_{i,j}]. \quad (\text{A.22})$$

Using the Leibniz rule to take out from derivate δI and δc and assuming on the edges of the model $\delta I = 0$ and $\delta c = 0$ we get

$$\begin{aligned} \delta \Phi &= \psi \sum_{ijh} \left\{ \frac{\bar{\varphi}_j G_h^2 - \bar{\varphi} G_j}{\bar{\varphi}_j^2} \right\} I_{i,j,h} \delta I_{i,j,h} \\ &- \varphi \sum_{ijh} \nabla \cdot \left\{ \left[\frac{\bar{\psi}_j \tilde{G}_h^2 - \bar{\psi} \tilde{G}_j}{\bar{\psi}_j^2} \right] [\nabla I_{i,j,h} \cdot \nabla c_{i,j}] \nabla c_{i,j} \right\} \delta I_{i,j} \\ &- \varphi \sum_{ijh} \nabla \cdot \left\{ \left[\frac{\bar{\psi}_j \tilde{G}_h^2 - \bar{\psi} \tilde{G}_j}{\bar{\psi}_j^2} \right] [\nabla I_{i,j,h} \cdot \nabla c_{i,j}] \nabla I_{i,j,h} \right\} \delta c_{i,j}. \end{aligned} \quad (\text{A.23})$$

Finally, identifying the gradient of the kernels, according the equation A.4, we have:

$$\begin{aligned} \frac{\partial \mathcal{K}_{i,j,h}}{\partial I} &= \psi \left\{ \frac{\bar{\varphi}_j G_h^2 - \bar{\varphi} G_j}{\bar{\varphi}_j^2} \right\} I_{i,j,h} - \varphi \nabla \cdot \left\{ \left[\frac{\bar{\psi}_j \tilde{G}_h^2 - \bar{\psi} \tilde{G}_j}{\bar{\psi}_j^2} \right] [\nabla I_{i,j,h} \cdot \nabla c_{i,j}] \nabla c_{i,j} \right\}, \\ \frac{\partial \mathcal{K}_{i,j,h}}{\partial c} &= -\varphi \nabla \cdot \left\{ \left[\frac{\bar{\psi}_j \tilde{G}_h^2 - \bar{\psi} \tilde{G}_j}{\bar{\psi}_j^2} \right] [\nabla I_{i,j,h} \cdot \nabla c_{i,j}] \nabla I_{i,j,h} \right\}. \end{aligned} \quad (\text{A.24a})$$

APPENDIX B – QUASI-NEWTON METHOD

The quasi-Newton optimization methods are algorithms based on Newton's method for searching stationary points. Both, Newton's method and quasi-Newton methods, presupposes that the nonlinear functions can be approximated locally by quadratic curves. This approximation can be represented by a Taylor series, around position x_m

$$f(x) = f(x_m) + f'(x_m)(x - x_m) + \frac{1}{2}(x - x_m)f''(x_m)(x - x_m). \quad (\text{B.1})$$

A stationary point is the solution of

$$f'(x) = f'(x_m) + f''(x_m)(x - x_m) = 0 \quad (\text{B.2})$$

so that $x = x_m - [f''(x_m)]^{-1}f'(x_m)$. If $f(x)$ is not a quadratic function, then the solution needs to be found iteratively, updating the solution according to

$$x_{m+1} = x_m - \delta_m \quad (\text{B.3})$$

with m representing the current iteration and

$$\delta_m = [f''(x_m)]^{-1}f'(x_m). \quad (\text{B.4})$$

Most of the problems have more than one variable, in this case x_m represents the vector $\mathbf{x}_m = (x_m^1, x_m^2, \dots, x_m^n)$ with n variables, $f'(x_m)$ the gradient vector

$$\mathbf{g}_m = \left(\frac{\partial f(\mathbf{x}_m)}{\partial x_m^1}, \frac{\partial f(\mathbf{x}_m)}{\partial x_m^2}, \dots, \frac{\partial f(\mathbf{x}_m)}{\partial x_m^n} \right) \quad (\text{B.5})$$

and $f''(x_m)$ the Hessian matrix

$$\mathbf{H}_m = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x}_m)}{\partial x_m^1 \partial x_m^1} & \frac{\partial^2 f(\mathbf{x}_m)}{\partial x_m^2 \partial x_m^1} & \dots & \frac{\partial^2 f(\mathbf{x}_m)}{\partial x_m^n \partial x_m^1} \\ \frac{\partial^2 f(\mathbf{x}_m)}{\partial x_m^1 \partial x_m^2} & \frac{\partial^2 f(\mathbf{x}_m)}{\partial x_m^2 \partial x_m^2} & \dots & \frac{\partial^2 f(\mathbf{x}_m)}{\partial x_m^n \partial x_m^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x}_m)}{\partial x_m^1 \partial x_m^n} & \frac{\partial^2 f(\mathbf{x}_m)}{\partial x_m^2 \partial x_m^n} & \dots & \frac{\partial^2 f(\mathbf{x}_m)}{\partial x_m^n \partial x_m^n} \end{bmatrix}, \quad (\text{B.6})$$

which is symmetric positive definite.

Often the objective function doesn't have good behavior close to \mathbf{x}_m . In the sense that the surface in that is poorly approximate by a quadratic function. An additional step is included to correct that, with redefinition of B.3:

$$\begin{aligned} \mathbf{x}_{m+1} &= \mathbf{x}_m - \alpha_m \boldsymbol{\delta}_m \\ &= \mathbf{x}_m - \alpha_m \mathbf{H}_m^{-1} \mathbf{g}_m. \end{aligned} \quad (\text{B.7})$$

The Hessian calculus is computationally expensive. However, efforts have been made to reduce these problems. The main idea which makes the Newton method different from quasi-Newton methods was introduced by Broyden (1969). The Hessian can be evaluated indirectly by the analysis of successive gradient vectors, in other words, the idea is to use the information of current and previous iterations to calculate an approximation of the Hessian. Defining the quantities

$$\mathbf{s}_m = \mathbf{x}_{m+1} - \mathbf{x}_m \quad (\text{B.8})$$

and

$$\boldsymbol{\eta}_m = \mathbf{g}_{m+1} - \mathbf{g}_m, \quad (\text{B.9})$$

the \mathbf{H}_{m+1} is a solution of the linear system

$$\mathbf{H}_{m+1}\mathbf{s}_m = \boldsymbol{\eta}_m. \quad (\text{B.10})$$

The equation B.10 is known as quasi-Newton condition (or secant condition). There are a lot of solutions for this set of equations. The most convenient to guarantee a symmetric positive definite Hessian is the update formula of category two

$$\mathbf{H}_{m+1} = \mathbf{H}_m + a\mathbf{u}\mathbf{u}^t + b\mathbf{v}\mathbf{v}^t, \quad a, b \in \mathbb{R}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^n. \quad (\text{B.11})$$

Replacing the equation B.11 in B.10 we find

$$\begin{aligned} \mathbf{H}_m\mathbf{s}_m + a\mathbf{u}\mathbf{u}^t\mathbf{s}_m + b\mathbf{v}\mathbf{v}^t\mathbf{s}_m &= \boldsymbol{\eta}_m \\ \mathbf{u}(a\mathbf{u}^t\mathbf{s}_m) + \mathbf{v}(b\mathbf{v}^t\mathbf{s}_m) &= \boldsymbol{\eta}_m - \mathbf{H}_m\mathbf{s}_m. \end{aligned} \quad (\text{B.12})$$

The vectors \mathbf{u} and \mathbf{v} are not determined in a unique way, but equation B.12 suggests that a good choice is

$$\mathbf{u} = \boldsymbol{\eta}_m \quad \text{and} \quad \mathbf{v} = \mathbf{H}_m\mathbf{s}_m, \quad (\text{B.13})$$

in this case the terms inside the brackets must be $(a\mathbf{u}^t\mathbf{s}_m) = 1$ and $(b\mathbf{v}^t\mathbf{s}_m) = -1$, which implies the coefficients

$$a = \frac{1}{\boldsymbol{\eta}_m^t\mathbf{s}_m} \quad \text{and} \quad b = \frac{-1}{\mathbf{s}_m^t\mathbf{H}_m\mathbf{s}_m}. \quad (\text{B.14})$$

Substitution of B.13 and B.14 in B.11 produces the update rule

$$\mathbf{H}_{m+1} = \mathbf{H}_m + \frac{\boldsymbol{\eta}_m\boldsymbol{\eta}_m^t}{\boldsymbol{\eta}_m^t\mathbf{s}_m} - \frac{\mathbf{H}_m\mathbf{s}_m(\mathbf{H}_m\mathbf{s}_m)^t}{\mathbf{s}_m^t\mathbf{H}_m\mathbf{s}_m} \quad (\text{B.15})$$

known as update BFGS formula (BROYDEN, 1969; FLETCHER, 1970; GOLDFARB, 1970; SHANNO, 1970).

It is worth mentioning that there are other update formulas, for example the DFP formula (DAVIDON, 1991; FLETCHER; POWELL, 1963), which is considered the dual representation of BFGS formula.

The BFGS update formula is generally evaluated as the most efficient updating equation. However the BFGS algorithm has the disadvantage of storing in memory all values of the Hessian. An algorithm that circumvents this problem is known as limited memory BFGS or L-BFGS. This algorithm stores only a small amount of recent values of the equations B.8 and B.9, and evaluates recursively $\delta_m = \alpha_m \mathbf{H}_m^{-1} \mathbf{g}_m$ using

$$\mathbf{H}_m^{-1} = \left(\mathbf{I} - \frac{\mathbf{s}_m \boldsymbol{\eta}_m^t}{\boldsymbol{\eta}_m^t \mathbf{s}_m} \right) \mathbf{H}_{m-1}^{-1} \left(\mathbf{I} - \frac{\boldsymbol{\eta}_m \mathbf{s}_m^t}{\boldsymbol{\eta}_m^t \mathbf{s}_m} \right) + \frac{\mathbf{s}_m \mathbf{s}_m^t}{\boldsymbol{\eta}_m^t \mathbf{s}_m}. \quad (\text{B.16})$$

A recent modification for L-BFGS algorithm, is known as L-BFGS-B (NOCEDAL; WRIGHT, 2006; BYRD et al., 1995; ZHU et al., 1994), which extend the technique to bounded constraint optimization. This algorithm will be used in this work.

APPENDIX C – B-SPLINES INTERPOLATION

Interpolation reduces the number of parameters, accelerating the optimization process, as well as intrinsically producing smooth models, eliminating small wavelength artifacts. Good results initially obtained with sparse grids, can be refined with the gradual increase of the grid density. Interpolation in B-Splines is widely used in computing, due to its ease of construction, smoothness, evaluation accuracy and best fit into complex shapes. The 2D transformation between two grids can be made as

$$M(x_1^i, x_2^j) = \sum_{p=0}^{N_{x_1}} \sum_{q=0}^{N_{x_2}} m_{pq} B_p^n(x_1^i) B_q^n(x_2^j), \quad (\text{C.1})$$

where m is the input model, N_{x_1} e N_{x_2} are their number of samples, $B_p^n(x_1^i)$ is the B-Spline function of order n in position x_1^i , and i correspond to i th node of new grid. The B-Spline function is defined as

$$B_i^n(x) = \left(\frac{x - x_i}{x_{i+n} - x_i} \right) B_i^{n-1}(x) + \left(\frac{x_{i+n+1} - x}{x_{i+n+1} - x_{i+1}} \right) B_{i+1}^{n-1}(x), \quad (\text{C.2})$$

which $B_i^n(x)$ is obtained recursively, starting from the box function

$$B_i^0(x) = \begin{cases} 1, & \text{se } x_i < x < x_{i+1}, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{C.3})$$

This work used cubic B-Splines ($n = 3$), so that the grid transformation for the velocity model is

$$c(x_1^i, x_2^j) = \sum_{p=0}^{N_{x_1}} \sum_{q=0}^{N_{x_2}} c_{pq} B_p^3(x_1^i) B_q^3(x_2^j), \quad (\text{C.4})$$

and for the objective function gradient model

$$\frac{\partial \mathcal{L}}{\partial c_{i,j}}(x_1^i, x_2^j) = \sum_{p=0}^{N_{x_1}} \sum_{q=0}^{N_{x_2}} \frac{\partial \mathcal{L}}{\partial c_{p,q}} B_p^3(x_1^i) B_q^3(x_2^j). \quad (\text{C.5})$$

APPENDIX D – EROSION AND DILATION PSEUDO CODE

In this appendix we show the 2D pseudo codes for erosion and dilation operators applied for structural constraints velocity models. The algorithm's main parameters are show in Table D.1 and the codes on Algorithms D.1 (erosion) and D.2 (dilation).

Table D.1 – The algorithm's main parameters.

Parameter	Description
k	kth sequential application
n	Related to structural function size, $(2n + 1) \times (2n + 1)$
$fmin, fmax$	Minimum and maximum bounds of the velocity model
n_1, n_2	Vertical and horizontal sample numbers
$M(n_1, n_2)$	Mask function
$f(n_1, n_2)$	Velocity model

Figure D.1 – 2D Erosion algorithm.

```

input :  $k, n, fmin, fmax, n_1, n_2, M(n_1, n_2), f(n_1, n_2)$ 
output:  $E(n_1, n_2)$ 
 $f0(:, :) \leftarrow f(:, :)$ ;
for  $ik \leftarrow 1$  to  $k$  do
  for  $i_2 \leftarrow 1 + n$  to  $n_2 - n$  do
    for  $i_1 \leftarrow 1 + n$  to  $n_1 - n$  do
       $val \leftarrow fmax$ ;
      for  $j \leftarrow -n$  to  $n$  do
        for  $i \leftarrow -n$  to  $n$  do
           $val \leftarrow \max(\min(f(i_1 + i, i_2 + j)/M(i_1 + i, i_2 + j), val), fmin)$ 
        end
      end
       $f0(i_1, i_2) \leftarrow val$ ;
    end
  end
   $f(:, :) \leftarrow f0(:, :)$ ;
end
 $E(:, :) \leftarrow f(:, :)$ 

```

Figure D.2 – 2D Dilation algorithm.

```

input :  $k, n, fmin, fmax, n_1, n_2, M(n_1, n_2), f(n_1, n_2)$ 
output:  $D(n_1, n_2)$ 
 $f0(:, :) \leftarrow f(:, :)$ ;
for  $ik \leftarrow 1$  to  $k$  do
  for  $i_2 \leftarrow 1 + n$  to  $n_2 - n$  do
    for  $i_1 \leftarrow 1 + n$  to  $n_1 - n$  do
       $val \leftarrow fmin$ ;
      for  $j \leftarrow -n$  to  $n$  do
        for  $i \leftarrow -n$  to  $n$  do
           $val \leftarrow \min(\max(f(i_1 + i, i_2 + j)M(i_1 + i, i_2 + j), val), fmax)$ 
        end
      end
       $f0(i_1, i_2) \leftarrow val$ ;
    end
  end
   $f(:, :) \leftarrow f0(:, :)$ ;
end
 $D(:, :) \leftarrow f(:, :)$ 

```