

UNIVERSIDADE FEDERAL FLUMINENSE
INSTITUTO DE GEOCIÊNCIAS
DEPARTAMENTO DE GEOLOGIA E GEOFÍSICA



FELIPE TIMOTEO DA COSTA

**COMPARAÇÃO DOS MÉTODOS DE OTIMIZAÇÃO E ANÁLISE DOS
NÚMEROS DE ONDA RECUPERADOS PELA INVERSÃO DO CAMPO
DE ONDA COMPLETO**

TESE DE DOUTORADO

PROGRAMA DE PÓS-GRADUAÇÃO
DINÂMICA DOS OCEANOS E DA TERRA (DOT)

Niterói, RJ
2021

Ficha catalográfica automática - SDC/BIG
Gerada com informações fornecidas pelo autor

C837c Costa, Felipe Timóteo da
Comparação dos métodos de otimização e análise dos
números de onda recuperados pela inversão do campo de onda
completo / Felipe Timóteo da Costa ; Marco Cetale,
orientador. Niterói, 2019.
238 p. : il.

Tese (doutorado)-Universidade Federal Fluminense, Niterói,
2019.

DOI: <http://dx.doi.org/10.22409/PPGDOT.2019.d.05368212755>

1. Geofísica aplicada. 2. Sismologia. 3. Inversão
sísmica. 4. Prospeção sísmica. 5. Produção intelectual.
I. Cetale, Marco, orientador. II. Universidade Federal
Fluminense. Instituto de Geociências. III. Título.

CDD -

FELIPE TIMOTEO DA COSTA

**COMPARAÇÃO DOS MÉTODOS DE OTIMIZAÇÃO E ANÁLISE DOS NÚMEROS
DE ONDA RECUPERADOS PELA INVERSÃO DO CAMPO DE ONDA COMPLETO**

Tese apresentada ao Programa de Pós-Graduação em Dinâmica dos Oceanos e da Terra, da Universidade Federal Fluminense, como requisito parcial para a obtenção do título de Doutor. Área de concentração: Geofísica Aplicada.

Área de concentração: Geologia e Geofísica.

Aprovada em 23 de outubro de 2019 pela banca examinadora abaixo:

Prof. Marco Cetale, Dr. (Orientador)
UFF / DOT / GISIS

Prof. Luiz Alberto Santos, Dr.
UFF / PETROBRÁS / GISIS

Prof. Roger Matsumoto Moreira, Dr.
UFF / GISIS

Prof. Leandro Di Bartolo, Dr.
Observatório Nacional

André Bulcão, Dr.
CENPES / PETROBRÁS

Bruno Pereira Dias, Dr.
CENPES / PETROBRÁS

Niterói, RJ
2021

*Este trabalho é dedicado à,
Maria Helena Timóteo da Costa,
eternamente em minhas lembranças.*

Agradecimentos

À Universidade Federal Fluminense e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por viabilizar e financiar a minha pesquisa e formação; e a Agência Nacional de Petróleo (ANP/BDEP) por disponibilizar os dados utilizados nesse estudo.

Ao meu orientador Marco Cetale, por propor o desafio, pelo incentivo e todo conhecimento compartilhado durante nossa parceria. E ao Rodrigo Stern, por todo apoio técnico na área de tecnologia da informação.

Ao meu amor, Gabriela Zuquim pelo apoio incondicional durante minha jornada acadêmica.

À toda minha família, em especial aos meus pais Antônio e Maria da Penha, à meus irmãos e sobrinhos Fabrício, Fábio, Rafael, Anthony e Miguel. E também minha querida avó, Maria José.

Aos meus queridos amigos Felipe Souza, Thiago Barros, Thiago Capone e Daniel Franco por sempre acreditarem em mim.

À todos os professores que me ajudaram ao longo da minha caminhada, em especial aos professores Gildo, Alex e Adalberto.

À todos os amigos feitos durante a passagem pelo Grupo de Imageamento e Inversão Sísmica (GISIS). E aos colegas Denis, Victor Martins, Victor Xavier, Felipe Capuzzo, Felipe Cruz e Roberto Miyamoto pelas conversas enriquecedoras.

Muito obrigado.

Resumo

A aplicação da inversão do campo de onda completo (FWI - *Full Waveform Inversion*) para a construção de modelos de velocidade de alta resolução se tornou amplamente utilizada, entretanto, ainda permanece desafiadora. Nesse trabalho, foram comparados 3 métodos de otimização não-linear tipicamente usados na FWI, gradiente descendente, gradiente conjugado e L-BFGS. Para cada caso foram avaliadas a utilização da norma L1 e também a aplicação da regularização por *total variation*. Nos testes realizados, a otimização pelo método gradiente descendente combinado com a norma L1 forneceram os melhores resultados. Nestes testes, foram avaliados os modelos de velocidade Marmousi e um modelo baseado nas informações sísmicas do campo de Búzios na Bacia de Santos onde, em ambos os casos, a inversão foi capaz de aumentar a resolução do modelo. Além disso, ao fluxo FWI foi incluída a estimativa da *wavelet* fonte a partir dos dados pré-empilhamento, tornando os experimentos numéricos mais realistas. Por fim, foi proposta a análise dos números de onda recuperados pela FWI para o diagnóstico mais preciso dos resultados de inversão. Nos testes realizados foi mostrado que uma *wavelet* fonte com espectro de frequência amplo é capaz de recuperar mais números de onda, melhorando a resolução dos modelos estimados.

Palavras-chave: FWI, Modelagem Sísmica, Métodos de Otimização, Norma L1, Bacia de Santos.

Abstract

The full-waveform inversion (FWI) has been widely used for constructing high-resolution velocity models even though remains challenging. In this work, the application of three optimization methods was compared: steepest descent, conjugate gradient, and L-BFGS. For each case, tests were carried out with L1 norm and with the total variation regularization. In these tests, the steepest descent optimization method together with the L1 norm achieved the best results. The benchmark Marmousi model and a model based on the Buzios Field on Santos Basin, built with previous seismic data information, were used for the validation of the developed algorithms. In both cases, the FWI enhanced the model resolution. Besides that, in the FWI flow was added a source wavelet estimation step, became the numeric experiment more realistic. Finally, the analysis of the recovered wavenumber by FWI was proposed to a better diagnostic of the results. The tests carried out showed that a wavelet with a broadband source wavelet can recover more wavenumbers, consequently improving the resolution of the estimated models.

Keywords: Seismic Inversion, Seismic Modeling, Optimization Methods.

Lista de ilustrações

Figura 1 – Representação de uma aquisição sísmica.	9
Figura 2 – a) Geometria de aquisição lanço simétrico (<i>split-spread</i>). b) Geometria de aquisição lanço lateral (<i>end-on</i>). Os raios representam as reflexões primárias de duas camadas quaisquer.	11
Figura 3 – Principais eventos presentes em um sismograma .a) Lanço simétrico em aquisição terrestre. b) Lanço lateral em aquisição marinha. . . .	12
Figura 4 – Exemplo de dois sismogramas de uma aquisição terrestre. À esquerda, os dados brutos de campo. À direita, os mesmos sismogramas após a etapa de pré-condicionamento dos dados sísmicos.	13
Figura 5 – Representação da construção do agrupamento CMP e aplicação da correção NMO para a construção de um traço empilhado. As seções sísmicas são construídas com um conjunto de traços empilhados. .	14
Figura 6 – Esquema representando as etapas para a solução do problema direto.	18
Figura 7 – Esquema representando as etapas para a solução do problema inverso	20
Figura 8 – Representação do método de Newton. A função quadrática aproximada pelo método de Newton tangencia a função objetivo por várias iterações até que o mínimo da função objetivo seja alcançado.	29
Figura 9 – Representação dos diferentes passos do método Gradiente Descendente.	34
Figura 10 – Funções objetivo utilizadas nos testes de performance: a) Função objetivo Rosenbrock. b) Função objetivo Ackley.	39
Figura 11 – Funções objetivo utilizadas nos testes de performance do método gradiente descendente. a) Atualização do modelo utilizando a função objetivo Rosenbrock. b) Evolução em valores relativos da função objetivo Rosenbrock. c) Atualização do modelo utilizando a função objetivo Ackley d) Evolução em valores relativos da função objetivo Ackley.	41
Figura 12 – Funções objetivo utilizadas nos testes de performance do método gradiente conjugado. a) Atualização do modelo utilizando a função objetivo Rosenbrock. b) Evolução em valores relativos da função objetivo Rosenbrock. c) Atualização do modelo utilizando a função objetivo Ackley. d) Evolução em valores relativos da função objetivo Ackley.	42

Figura 13 – Funções objetivo utilizadas nos testes de performance do método quasi-Newton LBFGS. a) Atualização do modelo utilizando a função objetivo Rosenbrock b) Evolução em valores relativos da função objetivo Rosenbrock. c) Atualização do modelo utilizando a função objetivo Ackley d) Evolução em valores relativos da função objetivo Ackley.	43
Figura 14 – Representação do método adjunto.	48
Figura 15 – Representação da função objetivo e seu comportamento oscilatório para escalas diferentes. (a)-(d) ilustram a mesma função objetivo com escalas crescentes. A função objetivo possui menos mínimos locais quanto maior for a escala.	60
Figura 16 – Esquema que representa o problema de salto de ciclo. Quando existe uma diferença de fase maior que meio período do comprimento de onda, os ciclos do sinal são ajustados de forma errada (indicação em vermelho) pelo algoritmo de otimização. Isso provoca uma diminuição da função objetivo porém estima o modelo utilizando uma informação equivocada.	61
Figura 17 – Representação do comportamento da função objetivo para as diferentes escalas.	62
Figura 18 – Modelo de referência Marmousi. Número de amostras $N_x = 851$ e $N_z = 351$ com espaçamento entre os pontos da malha de 10 m. A linha pontilhada indica a posição do perfil apresentado na lateral. As setas indicam anomalias de baixa velocidade.	71
Figura 19 – Modelo inicial criado a partir da suavização do modelo de referência Marmousi (Figura 18).	72
Figura 20 – Modelo de velocidade conceitual baseado nos dados pós empilhamento do campo de Búzios. Número de amostras $N_x = 865$ e $N_z = 651$ com espaçamento entre os pontos da malha de 10 m. A linha pontilhada indica a posição do perfil lateral. As setas indicam a base do sal e as rochas reservatórios. O retângulo indica uma falha simulada artificialmente.	73
Figura 21 – Seção sísmica e os horizontes interpretados das principais formações utilizada para criação do modelo geológico conceitual do campo de Búzios. O horizonte azul é o fundo marinho (Fm. Marambaia), em verde é o topo do sal (Fm. Ariri), o horizonte roxo é a Discordância Pré-Alagoas e o horizonte vermelho é a Fm. Camboriú.	74
Figura 22 – Modelo de velocidade inicial criado a partir da suavização do modelo de referência Búzios (Figura 20)	75
Figura 23 – Interface gráfica para controle de qualidade	79

Figura 24 – <i>Snapshots</i> . Distribuição das amplitudes do campo de pressão para diferentes instantes de tempo utilizando o modelo Marmousi.	83
Figura 25 – Sismogramas na posição central do modelo a) Dado observado utilizando o modelo de referência Marmousi. b) Dado calculado utilizando o modelo inicial. A linha pontilha indica a posição do perfil lateral.	83
Figura 26 – a) Sismograma residual, foram utilizados o modelo de referência e o modelo inicial. b) Sinal do sismograma residual usado como fonte adjunta para o caso da norma L1.	84
Figura 27 – <i>Snapshots</i> . Distribuição das amplitudes do campo de pressão para diferentes instantes de tempo utilizando o modelo Búzios.	85
Figura 28 – a) Dado gerado utilizando o modelo de referência Búzios.b) Sismograma gerado utilizando o modelo inicial.	86
Figura 29 – a) Simograma residual para o modelo inicial de Búzios. b) Sinal do resíduo usado como fonte adjunta no caso da norma L1.	86
Figura 30 – Gradiente da função objetivo para um único par fonte receptor, caminho da onda. Para <i>offsets</i> a) 3000 m e b) 1000 m.	87
Figura 31 – Gradiente da função objetivo para um único par fonte receptor, caminho da onda. a) Modelo Marmousi b) Modelo Búzios.	88
Figura 32 – Gradiente utilizando o modelo de velocidade inicial e parametrizado na velocidade.	88
Figura 33 – Gradiente utilizando o modelo de velocidade inicial e parametrizado no quadrado da vagarosidade.	89
Figura 34 – Gradiente utilizando o modelo de velocidade inicial e parametrizado no quadrado da vagarosidade. A fonte adjunta utilizada foi baseada na norma L1.	89
Figura 35 – a) Gradiente incluindo o termo de regularização por <i>total variation</i> . O modelo de velocidade utilizado foi atualizado por uma iteração do algoritmo de inversão. b) Termo de regularização <i>total variation</i> após a primeira iteração.	90
Figura 36 – Gradiente utilizando o modelo de velocidade inicial. Nesse caso o gradiente foi parametrizado na velocidade (Equação 2.65).	91
Figura 37 – Gradiente da função objetivo parametrizado na vagarosidade utilizando o modelo de velocidade inicial.	91
Figura 38 – Gradiente parametrizado na vagarosidade com a norma L1 utilizando o modelo de velocidade inicial.	92
Figura 39 – a) Gradiente da função objetivo incluindo o termo de regularização por <i>total variation</i> . b) Termo de regularização <i>total variation</i> após a primeira atualização do modelo de velocidade.	93

Figura 40 – a) Modelo estimado utilizando o método gradiente descendente. b) Função objetivo.	96
Figura 41 – a) Modelo estimado utilizando o método gradiente descendente incluindo a regularização por <i>total variation</i> . b) Função objetivo. . . .	97
Figura 42 – a) Modelo estimado utilizando o método gradiente descendente utilizando a norma L1. b) Função Objetivo	97
Figura 43 – a) Modelo estimado utilizando o método gradiente descendente incluindo a norma L1 e a regularização por <i>total variation</i> . b) Função objetivo.	98
Figura 44 – a) Modelo estimado utilizando o método Gradiente Conjugado. b) Função objetivo.	98
Figura 45 – Modelo estimado utilizando o método Gradiente Conjugado incluindo a regularização por <i>total variation</i> . b) Função objetivo.	99
Figura 46 – a) Modelo estimado utilizando o método Gradiente Conjugado incluindo a norma L1. b) Função objetivo.	100
Figura 47 – a) Modelo estimado utilizando o método Gradiente Conjugado incluindo a norma L1 e a regularização por <i>total variation</i> . b) Função objetivo.	100
Figura 48 – a) Modelo estimado utilizando o método LBFGS. b) Função objetivo.	101
Figura 49 – Modelo estimado utilizando o método LBFGS incluindo a regularização por <i>total variation</i> . b) Função objetivo.	101
Figura 50 – Modelo estimado utilizando o método LBFGS incluindo a norma L1. b) Função objetivo.	102
Figura 51 – Modelo estimado utilizando o método LBFGS incluindo a norma L1 e a regularização por <i>total variation</i> . b) Função objetivo.	103
Figura 52 – a) Modelo estimado utilizando o método gradiente descendente. b) Função objetivo.	105
Figura 53 – Modelo de velocidade estimado pelo método gradiente descendente incluindo a regularização por <i>total variation</i> . b) Função objetivo. . . .	105
Figura 54 – a) Modelo de velocidade estimado pelo método gradiente descendente utilizando a norma L1. b) Função objetivo.	106
Figura 55 – Modelo de velocidade estimado pelo método Gradiente Descendente incluindo a norma L1 e a regularização por <i>total variation</i> . b) Função objetivo.	107
Figura 56 – a) Modelo de velocidade estimado pelo método gradiente conjugado. b) Função objetivo.	107
Figura 57 – a) Modelo de velocidade estimado pelo método gradiente conjugado incluindo a regularização por <i>total variation</i> . b) Função objetivo. . . .	108

Figura 58 – a) Modelo de velocidade estimado pelo método gradiente conjugado incluindo a norma L1. b) Função objetivo.	108
Figura 59 – a) Modelo de velocidade estimado pelo método gradiente conjugado incluindo a norma L1 e a regularização por <i>total variation</i> . b) Função objetivo.	109
Figura 60 – a) Modelo de velocidade estimado pelo método LBFGS. b) Função objetivo.	110
Figura 61 – a) Modelo de velocidade estimado pelo método LBFGS incluindo a regularização por <i>total variation</i> . b) Função objetivo.	110
Figura 62 – a) Modelo de velocidade estimado pelo método LBFGS incluindo a norma L1. b) Função objetivo.	111
Figura 63 – a) Modelo de velocidade estimado pelo método LBFGS incluindo a norma L1 e a regularização por <i>total variation</i> . b) Função objetivo.	111
Figura 64 – a) <i>Wavelet</i> wide Band B-spline (CAO; HAN, 2011), usada para gerar os dados observados. b) Sismograma sintético gerado usando a <i>wavelet</i> BSpline.	113
Figura 65 – Filtro passa baixa aplicado ao dado observado com frequência de corte de a) 12 Hz e b) 21 Hz. A linha pontilhada indica a posição do traço apresentado a esquerda.	114
Figura 66 – a) Comparação entre a <i>wavelet</i> inicial, <i>wavelet</i> fonte e <i>wavelet</i> estimada. b) Decomposição da <i>wavelet</i> estimada nas faixas de frequências utilizadas na inversão.	115
Figura 67 – a) Modelo de velocidade estimado pelo método gradiente descendente utilizando a <i>wavelet</i> estimada. A norma L1 foi utilizada. b) Função objetivo.	116
Figura 68 – a) <i>Wavelet</i> BSpline utilizada para a geração dos dados observados. b) Exemplo de sismograma sintético usando a <i>wavelet</i> BSpline.	117
Figura 69 – Exemplos de dois sismogramas observados filtrados com frequência de corte de a) 12 Hz e b) 21 Hz.	118
Figura 70 – a) Comparação entre a <i>wavelet</i> inicial, <i>wavelet</i> fonte e <i>wavelet</i> estimada. b) Decomposição da <i>wavelet</i> estimada nas faixas de frequências utilizadas na inversão.	119
Figura 71 – a) Modelo de velocidade estimado pelo método gradiente descendente incluindo a norma L1 e utilizando a <i>wavelet</i> estimada. b) Função objetivo.	119

Figura 72 – Confiabilidade da informação obtida a partir de medidas sísmicas na superfície. O imageamento sísmico pode ser visto através de duas bandas de frequências (ou números de ondas) distintas, o modelo de velocidade nas baixas frequências e a imagem migrada (refletividade) nas altas frequências. O método sísmico possui uma lacuna de informação nos comprimentos de ondas intermediários. .	120
Figura 73 – a) Números de onda de referência. b) Números de onda recuperados pelo método de otimização gradiente descendente combinado com a norma L1 e utilizando a <i>wavelet</i> estimada.	122
Figura 74 – Números de onda recuperados pela FWI utilizando o método gradiente descendente a) d) g) e j); método gradiente conjugado b),e), h) e k); e método L-BFGS c),f), i) e l). A aplicação da norma L2 é apresentada na primeira linha a),b) e c); a aplicação da regularização por <i>total variation</i> na segunda linha d), e) e f); Na terceira linha está a aplicação da norma L1 g),h) e i); e a combinação da norma L1 com a regularização está na última linha j), k) e i) . . .	123
Figura 75 – a) Números de onda de referência. b)Números de onda recuperados utilizando uma <i>wavelet</i> fonte estimada e o método de otimização gradiente descendente e a norma L1.	124
Figura 76 – Números de onda recuperados pela FWI utilizando o método gradiente descendente a) d) g) e j); método gradiente conjugado b),e), h) e k); e método L-BFGS c),f), i) e l). A aplicação da norma L2 é apresentada na primeira linha a),b) e c); a aplicação da regularização por <i>total variation</i> , na segunda linha d), e) e f); Na terceira linha está a aplicação da norma L1 g),h) e i); e a combinação da norma L1 com a regularização está na última linha j), k) e i) . . .	125
Figura 77 – Seção migrada utilizando o modelo de velocidade inicial.	127
Figura 78 – Seção migrada utilizando o modelo de velocidade estimado pelo método gradiente descendente incluindo a norma L1 no caso em que a <i>wavelet</i> foi estimada.	127
Figura 79 – Seção migrada utilizando o modelo de velocidade inicial.. . . .	128
Figura 80 – Seção migrada utilizando o modelo de velocidade estimado pelo método gradiente descendente incluindo a norma L1 no caso em que a <i>wavelet</i> foi estimada.	128
Figura 81 – Perturbação no meio unidimensional provocada pela chegada da frente de onda.	147
Figura 82 – A diferença de pressão no meio produz uma força resultante.	148

Figura 83 – Corpo D qualquer, onde são analisadas as forças de superfície e as forças de corpo. Note que o ponto P está na superfície do corpo D e o ponto Q está no interior do corpo D	150
Figura 84 – Representação do deslocamento de partícula provocado pela deformação do corpo D entre os instantes de tempo t e t'	152
Figura 85 – a) Representação do vetor tração. b) Representação do tensor de tensões.	154
Figura 86 – a) Exemplos com diversas <i>wavelets</i> Ricker com frequências de corte de 4 até 32 Hz. Os tons de vermelho indicam as frequências mais altas. b) Espectro de amplitude da <i>wavelet</i> Ricker com frequência de corte de 32 Hz	165
Figura 87 – a) <i>Wavelet Wide-Band B-spline</i> . b) Espectro de amplitude da <i>wavelet Wide-Band B-spline</i> . Foram utilizados os seguintes parâmetros: $p = 2$ Hz, $q = 27$ Hz, $m = 40$ e $f_b = 25$ Hz.	166
Figura 88 – a) <i>Wavelet Wide-Band B-spline</i> . b) Espectro de frequência da <i>wavelet Wide-Band B-spline</i> . Foram utilizados os parâmetros: $p = 4$ Hz, $q = 6$ Hz, $m = 1$ e $f_b = 1$ Hz.	167
Figura 89 – Representação geométrica das diferenças finitas. As retas DP, DC e DR são, respectivamente, as diferenças progressivas, centrais e regressivas.	171
Figura 90 – Representação da discretização de um modelo geológico (Marmousi). Em cada nó do <i>grid</i> são armazenados os valores do modelo de propriedade e os campos do fenômeno ondulatório.	175
Figura 91 – Representação do <i>stencil</i> para a aproximação de 4ª ordem da 2ª derivada espacial.	176
Figura 92 – Representação de uma malha espacial e temporal no domínio computacional que discretiza os campos contínuos. Nesse exemplo o campo contínuo é discretizado usando duas dimensões espaciais e uma dimensão temporal.	177
Figura 93 – Representação da malha de modelagem indicando em vermelho as posições da malha onde a equação da onda será modificada para simular um meio infinito (bordas não reflexivas).	180
Figura 94 – Exemplos de sismogramas sintéticos. a) Sem aplicação de bordas absorcivas. b) Aplicação de bordas absorcivas. c) Aplicação de camada absorciva do tipo Cerjan com 30 pontos. d) Aplicação de camada absorciva do tipo Cerjan com 100 pontos.	182

Figura 95 – a) Representação do domínio computacional onde será resolvida a equação da onda. A região em vermelho indica onde a equação da onda será modificada para atenuar os efeitos das bordas. Essa camada é adicionada externamente ao domínio de interesse. Função de amortecimento utilizando uma camada de amortecimento de a) 30 pontos e b) 100 pontos.	183
Figura 96 – Visão geral do algoritmo de FWI utilizando a abordagem multiescala no domínio da frequência.	186
Figura 97 – Visão geral para o algoritmo implementado que atualiza o modelo de velocidade utilizando um método de otimização baseado no gradiente da função objetivo. O pacote disponibilizado pelo consórcio Seiscope foi utilizado para aplicação do método de otimização.	187
Figura 98 – Visão geral para o algoritmo implementado para a extrapolação do campo de onda para o caso acústico utilizando paralelização no domínio dos tiros.	188
Figura 99 – Visão geral do algoritmo implementado de extrapolação do campo da onda	188
Figura 100–Visão geral do algoritmo que calcula o resíduo utilizando a paralelização no domínio do tiro.	189
Figura 101–Visão geral do algoritmo que calcula o gradiente da função objetivo através do método adjunto no domínio do tempo e utilizando a paralelização no domínio do tiro.	189
Figura 102–Visão geral do algoritmo implementado de extrapolação da fonte adjunta (modelagem reversa no tempo).	190
Figura 103–Visão geral do algoritmo que atualiza o modelo através do método LBFGS. Pacote disponibilizado pelo consórcio Seiscope.	190
Figura 104–Visão geral do algoritmo que realiza a busca linear. Algoritmo disponibilizado pelo consórcio Seiscope.	191

Lista de tabelas

Tabela 1 – Parâmetros de aquisição utilizados nos exemplos dessa seção. . . .	82
Tabela 2 – Parâmetros de aquisição utilizados nos exemplos dessa seção . . .	85
Tabela 3 – Parâmetros de aquisição utilizados nos exemplos dessa seção . . .	94
Tabela 4 – Parâmetros de otimização utilizados na FWI	95
Tabela 5 – Parâmetros de aquisição utilizados nos exemplos dessa seção . . .	104
Tabela 6 – Parâmetros de aquisição utilizados nos exemplos dessa seção . . .	112
Tabela 7 – Parâmetros de aquisição utilizados nos exemplos dessa seção . . .	117
Tabela 8 – Tabela com resumo dos resultados das inversões do modelo Marmousi	130
Tabela 9 – Tabela com resumo dos resultados das inversões do modelo Buzios	131
Tabela 10 – Coeficientes dos operadores de diferença central. Modificado de Bartolo (2013)	174

Lista de abreviaturas e siglas

FWI	<i>Full Waveform Inversion</i> - Inversão do campo da onda completo
CDP	<i>Commom depth point</i> - Ponto comum em profundidade
CMP	<i>Commom midpoint</i> - Ponto médio comum
RTM	<i>Reverse Time Migration</i> - Migração reversa no tempo
RMS	<i>Root mean square</i> - Média quadrática
OBN	<i>Ocean bottom nodes</i> - sismômetro telemétrico autônomo
OBC	<i>Ocean bottom cables</i> - Cabos de fundo oceânico
MPI	<i>Message passing interface</i>
HPC	<i>High performance computing</i> - Computação de alto desempenho
CSG	<i>Commom shot gather</i> - Família de tiro comum
NMO	<i>Normal moveout</i>

Lista de símbolos

t	Tempo
x, y, z	Coordenadas espaciais
$\mathbf{x}$	Vetor Posição
m	Massa
ρ	Densidade
κ	Incompressibilidade
v_p	Velocidade de propagação da onda P
v_s	Velocidade de propagação da onda S
v_{RMS}	Velocidade RMS
λ	Parâmetro de Lamé
μ	Módulo de cisalhamento
$\mathbf{m}$	Parâmetros do modelo
s	Quadrado da vagarosidade
λ_R	Parâmetro de regularização
R	Termo de regularização
$\mathbf{u}$	Deslocamento de partícula
$\mathbf{v}$	Velocidade de partícula
p	Campo de pressão
$p^\dagger$	Campo de adjunto
ϕ	Tempo de trânsito
λ_L	Multiplicador de Lagrange
w	Termo fonte (<i>wavelet</i>)
σ_{ij}	Tensor de tensões

ε_{ij}	Tensor de deformações
c_{ijkl}	Tensor Elástico de 4ª ordem
δ_{ij}	Delta de Kronecker
$\delta(\mathbf{x})$	Delta de Dirac
$\mathbf{n}$	Vetor normal
$\mathbf{d}$	Conjunto de dados dados
η	Ruído
χ	Função objetivo
∇	Operador Nabla
∇^2	Operador Laplaciano
$\mathbf{I}$	Matriz Identidade
$\mathbf{H}$	Matriz Hessiana
$\mathbf{B}$	Aproximação da matriz Hessiana
$\mathcal{L}$	Operador problema direto
L	Lagrangeana
$\mathcal{H}$	Transformada de Hilbert

Sumário

	Introdução	3
I	REFERENCIAIS TEÓRICOS	7
1	O MÉTODO SÍSMICO	9
1.1	A aquisição sísmica	10
1.2	Construção do modelo de velocidade inicial	13
2	MÉTODOS DE OTIMIZAÇÃO	17
2.1	O problema inverso	18
2.2	Função objetivo	21
2.2.1	Norma L2	21
2.2.2	Norma L1	22
2.2.3	Outras normas	22
2.3	Regularização do problema inverso	25
2.3.1	Regularização <i>Total Variation</i>	26
2.4	Métodos de otimização não-linear	28
2.4.1	Método de Newton	28
2.4.2	Método do máximo declive (gradiente descendente)	32
2.4.3	Método do gradiente conjugado	33
2.4.4	Método quasi-Newton (LBFGS)	36
2.4.5	Busca linear	37
2.4.6	Exemplos numéricos	39
2.5	Cálculo do gradiente da função objetivo e a derivada de Fréchet	44
2.6	O método do estado adjunto e os multiplicadores de Lagrange	45
2.6.1	Derivada da Lagrangeana com relação aos parâmetros	47
2.6.2	Derivada da Lagrangeana com relação ao campo direto	49
2.6.3	A fonte adjunta	52
2.6.4	Campo adjunto	54
2.6.5	Derivada da Lagrangeana com relação ao campo adjunto	55
2.6.6	Reparametrização do problema direto e o cálculo de $\frac{\partial L}{\partial \mathbf{m}}$	56
2.6.7	A Lagrangeana regularizada por <i>Total Variation</i>	58
2.7	O problema de salto de ciclo e a abordagem multiescala	60
2.8	Estimativa da <i>wavelet</i>	63
2.8.1	Método adjunto para estimativa da wavelet	64

II	METODOLOGIA, RESULTADOS E DISCUSSÕES	67
3	METODOLOGIA	69
3.1	Preparação do modelo inicial	70
3.2	Modelo Marmousi	70
3.3	Modelo baseado no Campo de Búzios	72
3.3.1	Contexto geológico	73
3.3.2	Construção do modelo de velocidades conceitual	74
3.4	Os dados calculados	76
3.5	Linguagem de programação	76
3.6	Paralelização	77
3.7	A biblioteca de otimização SEISCOPE	78
3.8	Interface para controle de qualidade	78
4	RESULTADOS	81
4.1	Modelagem sísmica	81
4.1.1	Modelo Marmousi	82
4.1.2	Modelo Búzios	84
4.2	Gradiente da função Objetivo	87
4.2.1	Modelo Marmousi	88
4.2.2	Modelo Búzios	91
4.3	Inversão modelo Marmousi	94
4.3.1	Método gradiente descendente	96
4.3.1.1	Norma L1 + Regularização por <i>total variation</i>	98
4.3.2	Gradiente conjugado	98
4.3.2.1	Norma L1 + Regularização por <i>total variation</i>	99
4.3.3	Método LBFGS	100
4.4	Inversão modelo Búzios	104
4.4.1	Gradiente Descendente	104
4.4.2	Método gradiente conjugado	107
4.4.3	Método LBFGS	110
4.5	Estimativa da wavelet no contexto da FWI	112
4.5.1	Modelo Marmousi	113
4.5.2	Modelo Búzios	117
4.6	Os números de onda recuperados pela FWI	120
4.6.1	Modelo Marmousi	122
4.6.2	Modelo Búzios	124
4.7	Migração reversa no tempo	126
4.7.1	Modelo Marmousi	126

4.7.2	Modelo Búzios	127
5	DISCUSSÃO	129
6	CONCLUSÃO	135
	REFERÊNCIAS	137
	 APÊNDICES	 143
	APÊNDICE A – A TEORIA ONDULATÓRIA	145
A.1	A equação da onda 1D	145
A.2	A equação da onda em meios elásticos	149
A.2.1	Forças de superfície	149
A.2.2	Forças de corpo	151
A.2.3	Equação do movimento	151
A.2.4	O tensor de tensões	153
A.2.5	O tensor de deformação	155
A.2.6	O tensor elástico e a relação tensão-deformação	156
A.2.7	A equação da onda para meios elásticos	158
A.2.8	A equação da onda elástica em meios homogêneos e isotrópicos	159
A.3	A equação da onda em meios acústicos	162
A.4	O termo fonte	164
A.4.1	A <i>wavelet</i> Ricker	165
A.4.2	A <i>wavelet</i> B-spline	166
	 APÊNDICE B – IMPLEMENTAÇÃO COMPUTACIONAL	 169
B.1	O método das diferenças finitas	169
B.1.1	Aproximação de derivadas por diferenças finitas	170
B.1.2	Aproximação da 2ª derivada por diferenças finitas	172
B.2	A discretização da equação da onda 2D	174
B.2.1	A derivada temporal	174
B.2.2	A derivada espacial	175
B.2.3	O esquema de marcha no tempo	177
B.3	Crítérios de dispersão e estabilidade	178
B.4	Simulando um meio infinito	179
B.4.1	Bordas não-reflexivas	179
B.4.2	Camada de absorção Cerjan	181

APÊNDICE C – FLUXOGRAMAS	185
------------------------------------	-----

ANEXOS	193
--------	-----

ANEXO A – WAVENUMBERS ILLUMINATED BY TIME-DOMAIN ACOUSTIC FWI USING THE L1 AND L2 NORMS . . .	195
--	-----

Introdução

Este trabalho visa desenvolver e implementar a estratégia de inversão sísmica popularmente conhecida como Inversão do Campo de Onda Completo (*Full-Wave Form Inversion* - FWI). Em termos gerais, o objetivo principal desse método é obter as propriedades do meio em subsuperfície, através de um experimento sísmico e a utilização da equação da onda. Nas últimas décadas, essa técnica vem se tornando amplamente conhecida devido ao seu grande potencial em recuperar as propriedades do meio em subsuperfície, com alta resolução e riqueza de detalhes quando comparada a técnicas similares como, por exemplo, a tomografia sísmica.

A FWI é um problema de otimização não-linear que permite estimar as propriedades do meio, a partir dos dados sísmicos e um modelo de propriedades inicial. Na aproximação para meios acústicos, o modelo de propriedades é o campo de velocidade no qual as ondas se propagam. O modelo de velocidade inicial, necessário para a aplicação do método, pode ser construído nas etapas convencionais de processamento sísmico, ou obtido por uma tomografia sísmica.

A estimativa do modelo de propriedade de alta resolução pela FWI é alcançada através da solução de um problema inverso, assim como na tomografia. Entretanto, a solução da equação da onda utiliza o campo completo (*full waveform*), ao contrário da tomografia e as outras etapas do processamento sísmico, que utilizam uma aproximação assintótica desta solução (CARCIONE; HERMAN; KROODE, 2002).

Por se tratar de um problema não-linear e com grande quantidade de parâmetros a serem estimados, os métodos de otimização utilizados para a solução do problema inverso são métodos locais e iterativos baseados no gradiente (NOCEDAL; WRIGHT, 2006). A partir do modelo inicial, o método de otimização gera uma sequência de estimativas, que a cada iteração se aproxima da solução desejada. Um critério de parada é estabelecido a partir de uma comparação entre os dados observados e dados calculados (função objetivo, função erro ou função custo). A forma como essas estimativas são feitas a cada iteração determina o método de otimização.

A primeira etapa para a solução do problema inverso é a solução do problema

direto. Neste caso, é a extrapolação dos campos de onda no meio físico a ser estudado, reproduzindo o experimento de aquisição sísmica, no qual o sismograma calculado é o produto final. Após a etapa da modelagem direta, os dados calculados são comparados aos observados no experimento sísmico real e, então, é medida sua similaridade através de uma função objetivo.

A atualização do modelo utiliza a primeira derivada da função objetivo com relação a cada parâmetro do modelo. a depender do método de otimização, também pode ser utilizada a segunda derivada. Estas derivadas são conhecidas como derivadas de Fréchet e para problemas com grande quantidade de parâmetros seu cálculo é proibitivo (PRATT, 1999). Porém, Lailly (1983) e Tarantola (1984) apresentaram uma forma de obtenção do gradiente da função objetivo sem a necessidade do cálculo das derivadas de Fréchet, conhecido como método do estado adjunto (PLESSIX, 2006), que viabilizou a resolução do problema inverso para o caso sísmico, que contém um grande número de parâmetros.

Apesar da forma eficiente de calcular a direção de atualização no espaço dos modelos, a aplicação da FWI de forma bem-sucedida requer que alguns desafios inerentes ao método sejam superados. O principal deles é a solução do problema de salto de ciclo (*cycle skipping problem*), que faz com que a solução encontrada pelo método de otimização fique presa em um mínimo local. Devido a natureza oscilatória dos dados sísmicos, o problema de salto de ciclo faz com que o dado calculado se ajuste a um ciclo errado do dado observado, levando o problema a uma solução espúria, sem significado geológico.

Existem diversas formas de mitigar o problema de convergência provocado pelo salto de ciclo, uma delas é a abordagem multiescala (BUNKS et al., 1995). Nesta abordagem o dado é decomposto em diferentes escalas, e o problema inverso é resolvido para cada escala, partindo das maiores para as menores escalas. Essa estratégia reduz os números de mínimos locais na função objetivo nas maiores escalas permitindo que o método de otimização convirja gradualmente para uma solução mais próxima do mínimo global.

Entretanto, apenas a abordagem multiescala não é suficiente para a construção de modelos de velocidade com boa qualidade e resolução. O pre-condicionamento do gradiente de forma adequada é fundamental para a convergência do método de

otimização. O silenciamento da atualização do modelo na coluna de água, a reparametrização do problema direto (CARNEIRO, 2017; CARNEIRO et al., 2018) e a escolha da norma da função objetivo (PYUN; SON; SHIN, 2009; JIMÉNEZ TEJERO et al., 2015) são exemplos de pre-condicionamento do problema inverso aplicados neste estudo.

Além disso, a regularização do problema inverso também pode evitar o modelo convergir para soluções espúrias, porém, em geral piora a taxa de convergência. Em regiões onde a presença de fortes contrastes de velocidade existem, regularizações que permitem a preservação das variações acentuadas no modelo de velocidades podem evitar soluções espúrias. A regularização por *Total Variation* é conhecida por possuir essas propriedades (ANAGAW; SACCHI, 2011).

Tendo em vista as diversas dificuldades na aplicação da FWI, esta tese apresenta a implementação da FWI acústica no domínio no tempo aplicada ao modelo de velocidade de testes de inversão Marmousi (VERSTEEG, 1994; MARTIN; WILEY; MARFURT, 2006) e a um modelo de testes construído utilizando os dados sísmicos do campo de Búzios, na Bacia de Santos. Foram comparados os métodos de otimização gradiente descendente, gradiente conjugado e quasi-Newton LBFGS. Além disso, foi utilizada a parametrização do problema direto utilizando o quadrado da vagarosidade, que forneceu melhores propriedades de convergência ao gradiente da função objetivo. Para cada método de otimização também foram avaliados se a regularização por *Total Variation* ou a utilização da norma L1 fornecem melhores propriedades de convergência.

Neste trabalho foram realizados testes com uma fonte sísmica com uma banda de frequência larga para gerar os dados observados nos testes de inversão, simulando os levantamentos sísmicos mais recentes que são capazes de adquirir um dado com mais qualidade. O espectro amplo da fonte permite que as baixas frequências garantam a estabilidade da inversão e, as altas frequências ajudem a construir modelos com mais resolução. Sabendo que as características dessa fonte melhoram as propriedades de convergência da inversão, foi proposto neste trabalho aplicar um método para estimativa da fonte sísmica a partir dos dados sísmicos pré-empilhamento. A introdução dessa abordagem ao fluxo de aplicação da FWI torna o experimento numérico mais próximo ao problema aplicado em dados reais.

Por fim, é apresentada uma análise dos números de onda recuperados pela FWI

utilizando a transformada de Fourier 1D, que permite identificar os limites de resolução alcançados por cada método. Os melhores modelos de velocidade obtidos foram utilizados na migração reversa no tempo para a construção das imagens sísmicas. Essas imagens foram comparadas àquelas construídas com os respectivos modelos de velocidades iniciais (antes da aplicação da FWI), e foi confirmada a sua melhora.

Estrutura da Tese

Este trabalho está dividido em duas partes, na primeira parte da tese são apresentados os referenciais teóricos e na segunda parte a metodologia, os resultados e as discussões. O capítulo 1 aborda uma breve revisão sobre o método sísmico. O capítulo 2 apresenta os fundamentos dos métodos de otimização, principal base para o entendimento dos algoritmos de FWI. Neste capítulo o problema inverso é definido e são apresentadas algumas funções objetivo que podem ser utilizadas. A regularização por *Total Variation* é formulada também neste capítulo. Os métodos de Newton, gradiente descendente, gradiente conjugado e quasi-Newton são abordados, assim como, a busca linear utilizada em todos eles. O método adjunto, a forma de se obter a direção de atualização do problema não-linear (gradiente da função objetivo), é derivado utilizando o método dos multiplicadores de Lagrange. No fim do capítulo é explicado como mitigar os problemas de salto de ciclo com a abordagem multiescala e como realizar a estimativa da *wavelet* fonte a partir dos dados pre empilhamento.

Na segunda parte desse trabalho, o capítulo 3 apresenta a metodologia aplicada aos testes de inversão, os modelos de velocidades utilizados e como os dados foram gerados. No capítulo 4, os resultados da modelagem sísmica e das diversas etapas da inversão aplicadas aos modelos Marmousi e Búzios são apresentados, incluindo os números de onda recuperados pelas inversões e exemplos de seções migradas pelo método de migração reversa no tempo (RTM). No capítulo 5 é feita uma discussão sobre os resultados obtidos indicando os melhores de acordo com as métricas estabelecidas. E por fim, no capítulo 6 são apresentadas as conclusões. No apêndice podem ser encontrados mais detalhes sobre a formulação da teoria ondulatória e a implementação computacional. E em anexo, o artigo que trata sobre a análise dos números de onda recuperados pela FWI pelas normas L1 e L2.

Parte I

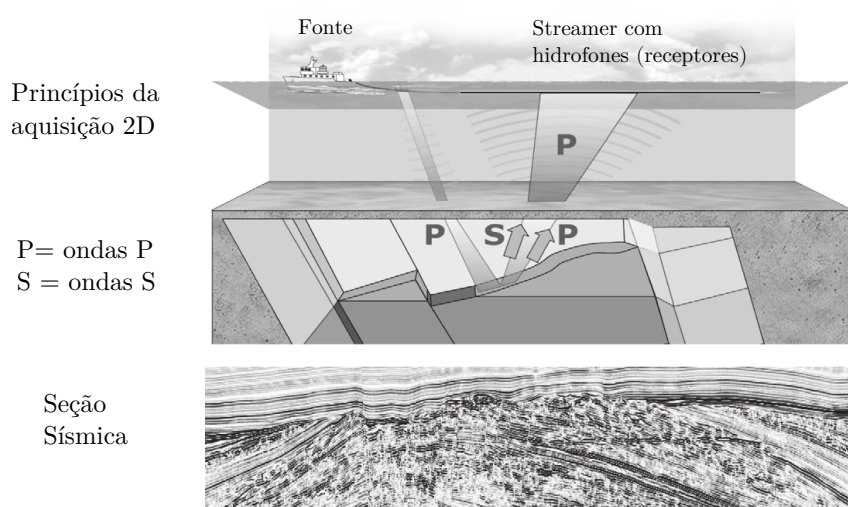
Referenciais teóricos

1 O método sísmico

O principal objetivo de um levantamento sísmico é mapear as reservas de recursos minerais em sub superfície, como por exemplo aquíferos ou reservatórios de hidrocarbonetos (SCHUSTER, 2017). Levantamentos sísmicos podem utilizar fontes sísmicas artificiais para a geração das ondas elásticas, que se propagam no meio geológico em subsuperfície. Os levantamentos também podem ocorrer utilizando apenas fontes sísmicas naturais, ou sísmica passiva. Conforme as frentes de ondas encontram as camadas com diferentes propriedades físicas, a energia sísmica é refletida, transmitida e refratada, e por fim registrada pelos receptores posicionados na superfície (Figura 1).

De acordo com Yilmaz (2000), tipicamente o método sísmico é decomposto em três etapas: aquisição, processamento e interpretação. Na primeira etapa, o registro dos dados sísmicos é realizado na região de estudo. O processamento sísmico consiste em utilizar diversas técnicas para atenuar os ruídos contidos nos dados registrados e construir as imagens de subsuperfície. Por último, a interpretação sísmica identifica as principais feições geológicas e os possíveis recursos minerais a partir das seções sísmicas construídas pelo processamento.

Figura 1 – Representação de uma aquisição sísmica.



Fonte: Modificado de Ikelle e Amundsen (2013).

As ondas sísmicas refletidas e refratadas registradas na superfície contêm as

informações das camadas litológicas da subsuperfície. O processamento dos tempos de trânsito registrados permite construir imagens sísmicas que representam a geometria das camadas litológicas. Vale ressaltar que as ondas refratadas são fundamentais para a recuperação dos baixos números de onda do modelo durante a aplicação da FWI, ou seja, esses eventos são responsáveis pela determinação das macro estruturas dos modelo.

Por fim, as seções sísmicas (Figura 1) são construídas de acordo com o modelo matemático que descreve o fenômeno ondulatório. As aproximações da solução da equação da onda podem levar em consideração meios acústicos, elásticos, viscoelásticos, entre outros.

1.1 A aquisição sísmica

A aquisição sísmica consiste em gerar um pulso sísmico que se propaga no interior da Terra. As ondas sísmicas refletidas pelas estruturas em subsuperfície são registradas por receptores posicionados na superfície, esse registro é conhecido como sismograma. O sismograma contém as informações dos tempos de trânsito das reflexões, provocadas pelas mudanças de propriedades físicas das rochas em subsuperfície que foram gerados pelo pulso sísmico.

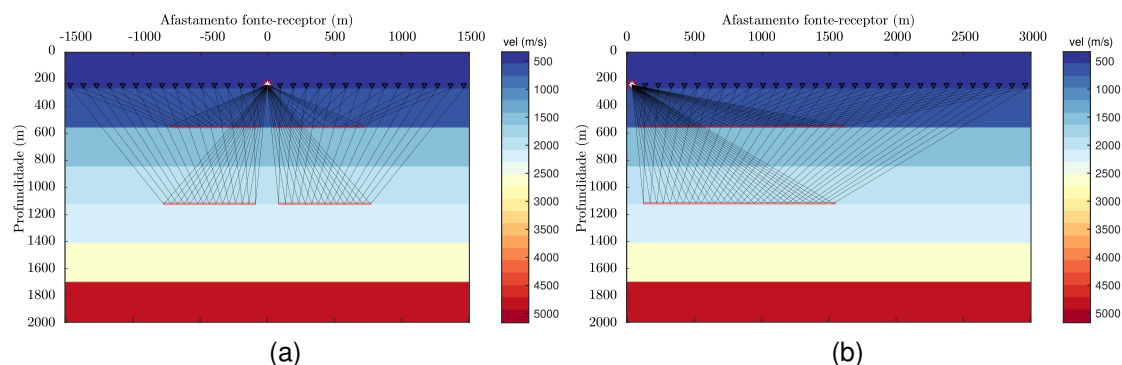
Os diferentes tempos de trânsito, devido às variações do afastamento fonte-receptor, fornecem um método útil para distinguir os diferentes tipos de eventos sísmicos utilizados para imagear a subsuperfície. Como por exemplo, as reflexões, as refrações e as múltiplas, que podem ser identificadas por sua coerência (SHERIFF; GELDART, 1995).

As posições relativas entre uma fonte e os receptores durante o registro dos tempos de trânsito são conhecidas como lanço (*spread*). As geometrias de aquisição mais comuns são o lanço simétrico (*split-spread*) em aquisições terrestres (Figura 3a), e o lanço lateral (*end-on*) em aquisições marinhas (Figura 3b).

Recentemente, outras geometrias de aquisição também são utilizadas, principalmente em levantamentos marinhos utilizando os OBN (*ocean bottom nodes*) ou os cabos de fundo oceânico (OBC - *Ocean bottom cable*), tipicamente usados em levantamentos sísmicos para monitoramento de reservatórios de petróleo. Nessas

aquisições, os receptores são fixados no assoalho oceânico e registram 4 componentes (pressão, v_x , v_y , v_z), e as fontes sísmicas são detonadas na superfície gerando lanços assimétricos. Porém, permitem o registro dos longos afastamentos e possuem uma alta razão sinal-ruído, fundamentais para a FWI.

Figura 2 – a) Geometria de aquisição lanço simétrico (*split-spread*). b) Geometria de aquisição lanço lateral (*end-on*). Os raios representam as reflexões primárias de duas camadas quaisquer.



As Figuras 4a e 4b apresentam os sismogramas para as geometrias de lanço simétrico e lanço lateral em uma situação hipotética, em que a geologia em subsuperfície é plano paralela. O sismograma consiste em visualizar todos os registros dos receptores, lado a lado, para uma mesma fonte. Esse arranjo de visualização é chamado de agrupamento de tiro comum (CSG - *common shot gather*).

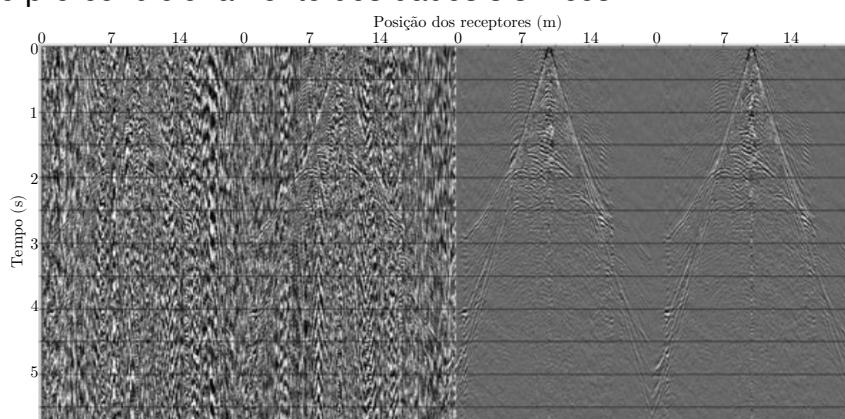
Em dados terrestres, existe uma grande quantidade de eventos que precisam ser atenuados pois não são previstos pela equação da onda acústica, como por exemplo, as ondas convertidas S para P, o *ground roll*, entre outros. Além disso, os dados terrestres, em geral, possuem uma baixa relação sinal-ruído. Em levantamentos sísmicos marinhos, a quantidade de eventos indesejados no sismograma é menor. Tipicamente, os dados marinhos possuem uma relação sinal-ruído superior quando comparada a levantamentos terrestres.

A aplicação da FWI é muito sensível aos dados de entrada, em particular, a relação sinal-ruído, os máximos afastamentos fonte-receptor registrados e o conteúdo de baixas frequências (BIZZI et al., 2016), por essa razão as principais aplicações da FWI são em dados marinhos.

Em geral, os dados sísmicos convencionais possuem um conteúdo de frequência entre 10 e 80 Hz (LATIMER; DAVIDSON; RIEL, 2002), porém o avanço tecnológico dos

A Figura 4 apresenta um exemplo de aplicação de diversos filtros em uma sísmica terrestre (BAETEN et al., 2013) para o pré condicionamento dos dados para a FWI. É importante ressaltar que o registro das baixas frequências e dos longos *offsets* permite que os números de onda intermediários, do modelo estimado, sejam recuperados corretamente (SIRGUE, 2003), desse modo permitindo construir um modelo de propriedades que esteja de acordo com as estruturas geológicas em subsuperfície.

Figura 4 – Exemplo de dois sismogramas de uma aquisição terrestre. À esquerda, os dados brutos de campo. À direita, os mesmos sismogramas após a etapa de pré-condicionamento dos dados sísmicos.



Fonte: Modificado de Baeten et al. (2013)

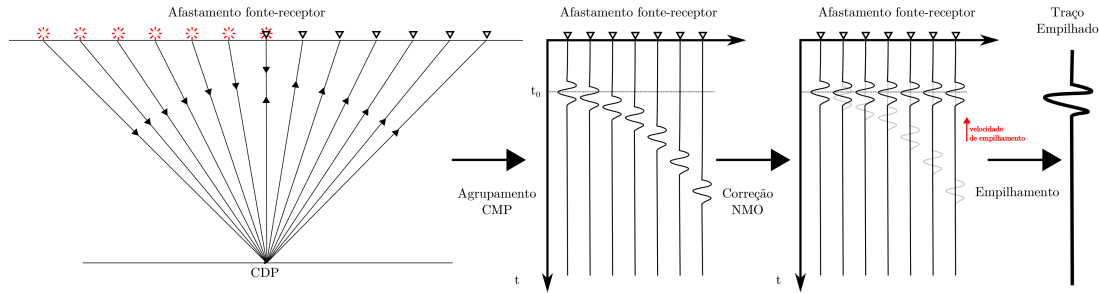
1.2 Construção do modelo de velocidade inicial

Para aplicação da FWI, é necessário um modelo de velocidade inicial que contenha os baixos números de onda. Esse modelo pode ser obtido pelo processamento na etapa de análise de velocidades, ou ser o mesmo modelo de velocidade utilizado na migração sísmica. Esses modelos de velocidades podem ser alcançados através de técnicas consolidadas como por exemplo análise de velocidade de migração (*Migration Velocity Analysis* - MVA), a tomografia de refração ou tomografia de reflexão (JONES, 2014).

Os modelos de velocidades obtidos na etapa do processamento são construídos com base no método CDP (*common depth-point*), na etapa de correção NMO (*Normal moveout*) (Figura 5).

Segundo Rosa (2010), assumindo um modelo de camadas plano-paralelas, o método CDP consiste em amostrar um mesmo ponto em subsuperfície diversas vezes,

Figura 5 – Representação da construção do agrupamento CMP e aplicação da correção NMO para a construção de um traço empilhado. As seções sísmicas são construídas com um conjunto de traços empilhados.



cada uma delas com um afastamento fonte-receptor distinto. Assim, a visualização dos registros agrupados representa a subsuperfície abaixo do ponto médio entre a fonte e o receptor, sendo denominado agrupamento CMP (*common midpoint*).

A visualização dos dados sísmicos agrupados por CMP permite aplicar a equação que estima os tempos de trânsito t entre a fonte e o receptor, que descrevem uma hipérbole no agrupamento CMP, dada por:

$$t^2 = t_0^2 + \frac{x_{CMP}^2}{v_{emp}^2}, \quad (1.1)$$

onde t_0 é o tempo de ida e volta até o refletor quando o afastamento é nulo, x_{CMP} são as coordenadas dos pontos médios e v_{emp} é a velocidade de empilhamento. A velocidade de empilhamento é encontrada através do método de análise de velocidades que testa diferentes velocidades que alinham a hipérbole de reflexão do agrupamento CMP. Devido às aproximações utilizadas no método CDP, a velocidade de empilhamento utilizada durante a correção NMO é aproximada pela velocidade RMS (*root mean square*):

$$v_{rms_N} = \sqrt{\frac{\sum_{i=1}^N v_i^2 t_i}{\sum_{i=1}^N t_i}}, \quad (1.2)$$

onde N é a última camada do modelo considerado para o cálculo de v_{rms} . A velocidade RMS pode ser relacionada com as velocidade de cada camada, a velocidade intervalar

v_i , recursivamente através da fórmula de Dix (AMERY, 1993):

$$v_p = v_i = \left[\frac{v_{rms_i}^2 t_i - v_{rms_{i-1}}^2 t_{i-1}}{t_i - t_{i-1}} \right]^{1/2}, \quad (1.3)$$

onde t_i é o tempo de trânsito até a i -ésima camada do modelo e v_{rms_i} é a velocidade RMS considerando até a i -ésima camada.

Apesar do modelo de velocidade construído durante a etapa de correção NMO ser um possível ponto de partida para a FWI, esse modelo possui limitações relacionadas as camadas com mergulho e as variações laterais de velocidade.

Um modelo de velocidade inicial mais apropriado para a FWI pode ser aquele obtido através da tomografia de refração, pois as aproximações do fenômeno ondulatório, utilizadas para a construção do modelo de velocidade, são menos restritas quando comparadas ao método CDP.

A tomografia de refração é um processo de inversão assim como a FWI, pois ajusta os dados observados aos dados calculados, através da minimização por mínimos quadrados. Na maior parte das aplicações são utilizadas apenas as primeiras chegadas do dado observado. Os dados calculados são obtidos através da equação eikonal, uma aproximação para altas frequências da equação da onda acústica, expressa por:

$$(\nabla\phi)^2 = \frac{1}{v_p^2}, \quad (1.4)$$

onde ϕ são os tempos de trânsito das primeiras chegadas da frente de onda (CARCIONE; HERMAN; KROODE, 2002). A solução da Equação 1.4 permite a estimativa dos dados calculados utilizados no processo de inversão tomográfica.

Os modelos de velocidade construídos por métodos tomográficos são os modelos iniciais mais consistentes para a aplicação da FWI (VIRIEUX; OPERTO, 2009), pois contém os baixos números de onda que podem ser caracterizados pelas feições suaves e ausência de feições abruptas, fundamentais para o sucesso da FWI. Esses modelos também são conhecidos como modelos de fundo (*background models*) e não sofrem tantas limitações quanto àqueles obtidos pelo método CDP, que pressupõe camadas plano-paralelas e ausência de variação lateral.

2 Métodos de otimização

Nesse capítulo serão apresentados os principais conceitos que envolvem a solução do problema inverso. Dentro do contexto da FWI acústica, a solução do problema inverso consiste em construir um modelo de velocidade, a partir de um modelo de velocidade inicial, comparando os dados gerados sinteticamente com os dados observados, através de uma função que mede a similaridade entre os dados calculados e observados, a função objetivo.

Os métodos de otimização não-linear buscam no espaço dos parâmetros do modelo as possíveis soluções que minimizem a função objetivo. Essa busca pode ser feita com o auxílio do gradiente da função objetivo, que fornece a direção de busca dentro do espaço dos parâmetros do modelo. O método dos multiplicadores de Lagrange foi utilizado para determinar a direção de atualização em cada iteração, dessa forma, garantindo que o método de otimização convirja para um mínimo local ou global da função objetivo.

A aplicação de um método de otimização não garante a convergência do modelo de velocidades para um modelo que represente a geologia da região estudada e, conseqüentemente, não permitindo a construção de boas imagens sísmicas. A abordagem multiescala é necessária para tentar evitar os problemas de salto de ciclo e garantir que o modelo de velocidade convirja para uma solução satisfatória.

Portanto, nesse capítulo será formulado o problema inverso e apresentadas as funções objetivos mais comuns, utilizadas no contexto da FWI. Será introduzido brevemente o problema inverso regularizado, que permite a introdução de informação a priori nas soluções, e os principais métodos de otimização baseados no gradiente da função objetivo. A dedução de como é obtido o gradiente da função objetivo pelo método do estado adjunto é apresentada, assim como a abordagem multiescala para evitar o problema de saltos de ciclo. Por fim, o método de estimativa para *wavelet* fonte é apresentada.

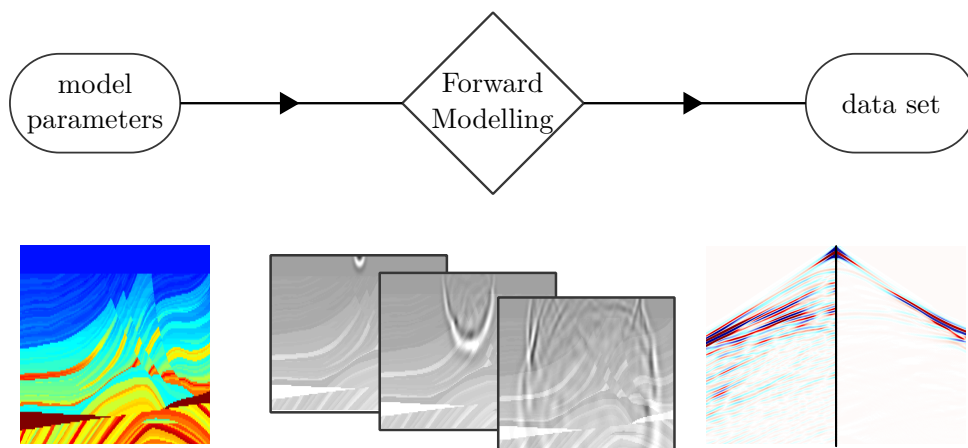
2.1 O problema inverso

De uma forma geral, os métodos de otimização são técnicas utilizadas na solução do problema inverso. Este, por sua vez, permite a estimativa de um conjunto de parâmetros desconhecidos, a partir dos dados observados em um experimento físico pré-estabelecido.

Grande parte dos problemas estudados pela geofísica podem ser abordados dessa forma, de modo que, a partir de um conjunto de dados geofísicos adquiridos, é possível inferir os parâmetros desejados, que normalmente são inacessíveis diretamente, utilizando a lei física que rege o fenômeno estudado (Apendice A). No caso sísmico, os dados observados são os sismogramas adquiridos, o parâmetro de interesse é o modelo de propriedade em subsuperfície e as leis que regem o fenômeno é a teoria ondulatória.

A solução do problema inverso começa com a definição de uma função objetivo (função custo, função erro). Essa função relaciona o dado observado e o dado que será calculado a partir do conhecimento das leis que regem o fenômeno físico e do modelo de propriedades. Essa função tipicamente é escolhida como uma medida da similaridade entre os dados observados e calculados. Portanto, o objetivo do problema inverso é minimizar a função objetivo. Ou seja, ajustar os dados calculados de forma que se tornem mais parecidos com os dados reais.

Figura 6 – Esquema representando as etapas para a solução do problema direto.
Forward Problem



Segundo Nocedal e Wright (2006), os métodos de otimização podem ser

classificados como sendo contínuos ou discretos, sem restrições ou com restrições, globais ou locais, e por fim, estocásticos ou determinísticos. A estratégia FWI pode ser classificada como um método de otimização contínuo, sem restrição, local e determinístico. Local, pois as direções de busca utilizadas garantem apenas a convergência para um mínimo local. Sem restrição, pois o método que calcula a direção de atualização faz uso do método dos multiplicadores de Lagrange sem restrição. Contínuo, pois toda a derivação do método ocorre no espaço contínuo. E, por fim, determinístico, pois não utiliza a teoria das probabilidades em sua formulação.

Como o termo já sugere, o problema inverso é contraposto ao problema direto. No problema direto, uma lei física é aplicada a um conjunto de parâmetros para obter os dados calculados, a modelagem direta (*forward modeling*). Usando novamente o exemplo do caso sísmico, a lei física é a teoria ondulatória, o conjunto de parâmetros é o modelo de propriedade em subsuperfície, e os dados calculados são os sismogramas obtidos com a modelagem direta.

Segundo Tarantola (2005), o problema inverso pode ser abordado em três etapas:

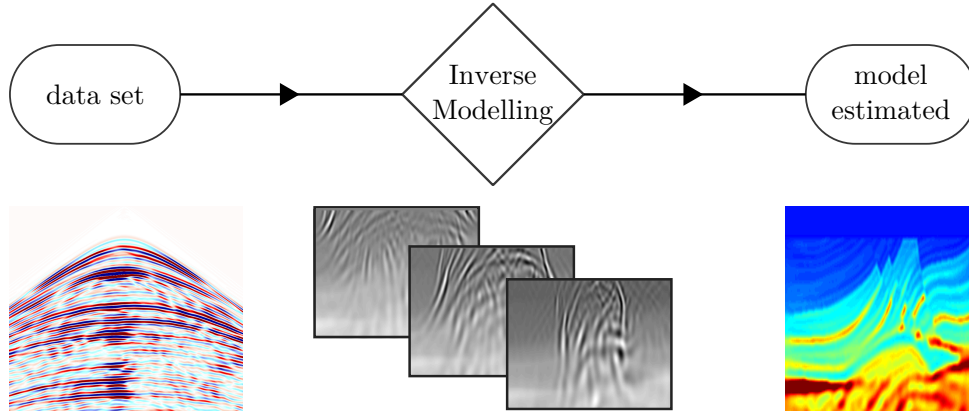
1. Parametrização do problema; determinação de um conjunto mínimo dos parâmetros do modelo de forma que caracterize completamente o problema.
2. Modelagem direta: determinação da lei física que permite, para um conjunto de parâmetros do modelo, fazer previsões sobre os dados observados.
3. Modelagem inversa: utilização dos dados observados para inferir os valores dos parâmetros do modelo.

Matematicamente, o problema direto pode ser escrito como:

$$\mathcal{L}(\mathbf{m}) = \mathbf{d}, \quad (2.1)$$

onde $\mathbf{m} \in \mathbb{R}^N$ representa os parâmetros do modelo e $\mathbf{d} \in \mathbb{R}^M$ representa os dados, e $\mathcal{L} : \mathbb{R}^N \mapsto \mathbb{R}^M$ é o operador de modelagem direta, a lei física que relaciona os parâmetros aos dados.

Figura 7 – Esquema representando as etapas para a solução do problema inverso
Inverse Problem



Assim, o problema inverso pode ser definido como:

$$\mathbf{m} = \mathcal{L}^{-1}(\mathbf{d}), \quad (2.2)$$

portanto, nessa formulação é possível estimar os parâmetros do modelo a partir dos dados observados. No problema sísmico acústico, o parâmetro a ser estimado pode ser o modelo de velocidade, vagarosidade ou quadrado da vagarosidade. No caso elástico, os parâmetros estimados podem ser: a velocidade de propagação da onda P, a velocidade de propagação da onda S e a densidade; ou os parâmetros de Lamé (λ e μ) e a densidade. A escolha dos parâmetros para inversão influenciam as propriedades de convergência do problema inverso

Aster, Borchers e Thurber (2013) afirmam que uma abordagem razoável para a solução do problema inverso, que estimará os parâmetros do modelo $\mathbf{m}^*$, é encontrar a melhor solução aproximada que minimiza alguma medida da similaridade χ entre os dados observados ($\mathbf{d}_{obs}$) e os dados calculados ($\mathbf{d}_{cal}$). Essa medida é conhecida como função objetivo. Dessa forma, a solução do problema inverso pode ser obtida através de:

$$\mathbf{m}^* = \min [\chi(\mathbf{d}_{cal}(\mathbf{m}) - \mathbf{d}_{obs})], \quad (2.3)$$

ou seja, os parâmetros estimados $\mathbf{m}^*$ são aqueles que minimizam a função objetivo. Assim, o algoritmo de otimização modifica o conjunto de parâmetros de modo que os

dados calculados se tornem mais parecidos com os dados observados.

2.2 Função objetivo

A solução dos problemas inversos (Equação 2.3) é baseada em uma medida do resíduo entre os dados preditos pelo modelo estimado (ou inicial) e os dados observados, assim como no ajuste de uma curva pelo método dos mínimos quadrados, que utiliza a distância euclidiana, ou norma L2, como métrica (MENKE, 2012). O método dos mínimos quadrados estima a solução do problema inverso encontrando o conjunto de parâmetros que minimiza a função objetivo (medida do resíduo).

2.2.1 Norma L2

A norma L2 é amplamente usada na solução dos problemas inversos devido a sua simplicidade, mesmo quando a hipótese de distribuição gaussiana dos dados não é satisfeita (TARANTOLA, 2005). A função objetivo que será minimizada pode ser definida pela norma L2, que possui a seguinte expressão:

$$\chi_2(\mathbf{m}) = \frac{1}{2} \int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}(\mathbf{x}, t)|^2 d\mathbf{x} dt \quad (2.4)$$

onde $\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t)$ são os dados calculados e $\mathbf{d}_{obs}$ são os dados observados. O problema direto (Equação 2.1) é solucionado sobre todo o volume V e durante o período de tempo T para a obtenção dos dados calculados. No experimento sísmico, os dados observados são os sismogramas registrados durante a etapa de aquisição de dados, onde os campos de pressão gerados pela fonte sísmica são registrados na posição dos receptores a cada tiro. Os dados calculados são obtidos através da solução da equação da onda acústica (Equação A.56) e registrados os campos de onda (pressão ou deslocamento de partícula) para as mesmas posições de receptores dos dados observados a cada tiro.

Assim, a função objetivo para o experimento sísmico utilizando a norma L2 fica:

$$\chi_2(\mathbf{m}) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{t=1}^{N_t} \sum_{r=1}^{N_r} |\mathbf{d}_{cal}^s(\mathbf{m}; \mathbf{x}^r, t) - \mathbf{d}_{obs}^s(\mathbf{x}, t)|^2, \quad (2.5)$$

onde N_r é o número de receptores por tiro, N_t é o número de amostras de tempo registradas e N_s é o número total de tiros realizados.

2.2.2 Norma L1

Segundo Tarantola (2005), a norma L1 é conhecida por ser pouco sensível a valores espúrios (*outliers*) da distribuição dos dados, ou seja, um método mais robusto à presença de ruído. A norma L1 é definida pela diferença absoluta entre os dados calculados e dados observados:

$$\chi_1(\mathbf{m}) = \int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}(\mathbf{x}, t)| d\mathbf{x} dt. \quad (2.6)$$

Similarmente a norma L2, a expressão para a função objetivo utilizando norma L1 para o caso do experimento sísmico é dada pela soma do sismograma residual $\Delta \mathbf{d} = \mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}$:

$$\chi_1(\mathbf{m}) = \sum_{s=1}^{N_s} \sum_{t=1}^{N_t} \sum_{r=1}^{N_r} |\mathbf{d}_{cal}^s(\mathbf{m}; \mathbf{x}^r, t) - \mathbf{d}_{obs}^s(\mathbf{x}, t)| \quad (2.7)$$

2.2.3 Outras normas

Vários autores apresentaram diferentes normas para tentar atenuar a sensibilidade da função objetivo à não-linearidade do problema sísmico, e evitar que a solução alcance um mínimo local. Diferentes normas tentam tratar alguma peculiaridade dos dados sísmicos para melhorar a convergência do método de otimização (FICHTNER, 2011; BULCÃO et al., 2014; JIMÉNEZ TEJERO et al., 2015; SCHUSTER, 2017). Por exemplo, a norma

dada pela correlação cruzada ponderada entre os dados observados e calculados é dada por:

$$\chi_{CC}(\mathbf{m}) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{\tau=1}^{N_\tau} \left[\sum_{t=1}^{N_t} W_\tau \mathbf{d}_{cal}^s(\mathbf{m}; \mathbf{x}^r, t + \tau) \mathbf{d}_{obs}^s(\mathbf{x}^r, t) \right]^2. \quad (2.8)$$

onde $W = e^{-\alpha\tau^2}$ é um peso gaussiano utilizado para penalizar suavemente a correlação entre as amplitudes, onde α é um parâmetro que define a largura da função Gaussiana.

Segundo Schuster (2017), a função objetivo por correlação cruzada ponderada é um funcional que atua na fase do sinal e possui um comportamento mais robusto em relação à função objetivo convencional (norma L2) na presença de ruídos. Entretanto, essa função objetivo alcança apenas baixas resoluções, podendo ser utilizada para fornecer um modelo inicial para a inversão convencional.

A função objetivo que utiliza o envelope e a fase instantânea do sinal também possui melhores propriedades de convergência na presença de ruído e modelos iniciais ruins (TEJERO et al., 2014). Chi, Dong e Liu (2014) afirmam também que a função objetivo baseada no envelope é indicada para dados com ausência das baixas frequências. A amplitude instantânea (envelope) e a fase instantânea dos dados pode ser obtida através do conceito de função analítica, que utiliza a transformada de Hilbert (IKELLE; AMUNDSEN, 2013).

A função analítica $s(t)$ de um sinal qualquer $u(t)$ é definida como (WU; LUO; WU, 2014):

$$s(t) = u(t) + iv(t), \quad (2.9)$$

onde $v(t) = \mathcal{H}[u(t)]$, é a transformada de Hilbert do sinal de entrada $u(t)$. Como o sinal analítico é um número complexo, pode ser reescrito em sua forma exponencial:

$$s(t) = a(t)e^{i\theta(t)}, \quad (2.10)$$

onde

$$a(t) = \sqrt{u(t)^2 + v(t)^2} \quad (2.11)$$

e

$$\theta(t) = \arctan\left(\frac{u(t)}{v(t)}\right). \quad (2.12)$$

As funções $a(t)$ e $\theta(t)$ são a amplitude instantânea e a fase instantânea do sinal $u(t)$, respectivamente.

Assim, a função objetivo baseada no envelope, ou amplitude instantânea, pode ser escrita como (JIMÉNEZ TEJERO et al., 2015):

$$\chi_E(\mathbf{m}) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{t=1}^{N_t} |a(\mathbf{d}_{cal}^s(\mathbf{m}; \mathbf{x}^r, t)) - a(\mathbf{d}_{obs}^s(\mathbf{x}^r, t))|^2. \quad (2.13)$$

e a função objetivo baseada na fase instantânea do sinal pode ser escrita como:

$$\chi_\theta(\mathbf{m}) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{t=1}^{N_t} |\theta(\mathbf{d}_{cal}^s(\mathbf{m}; \mathbf{x}^r, t)) - \theta(\mathbf{d}_{obs}^s(\mathbf{x}^r, t))|^2. \quad (2.14)$$

A aplicação de uma função de amortecimento em conjunto com a norma L2 pode melhorar as propriedades de convergência da função objetivo, pois é possível ajustar o fator de amortecimento σ para controlar a profundidade da inversão. Assim, inicialmente, pode-se estimar as partes mais rasas do modelo com um fator de amortecimento alto e, com o avanço das iterações, ajustar o parâmetro σ para estimar as propriedades em maiores profundidades (CHEN; CHEN; WU, 2015). A função objetivo com amortecimento pode ser escrita conforme a expressão:

$$\chi_\sigma(\mathbf{m}) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{t=1}^{N_t} e^{-\sigma t} |(\mathbf{d}_{cal}^s(\mathbf{m}; \mathbf{x}^r, t)) - (\mathbf{d}_{obs}^s(\mathbf{x}^r, t))|^2. \quad (2.15)$$

2.3 Regularização do problema inverso

A regularização envolve a introdução de informação extra com o objetivo de melhorar a convergência na solução de um problema inverso mal-posto. Normalmente, essa informação vem sob a forma de alguma restrição sobre os parâmetros do modelo $\mathbf{m}$ (KORMANN et al., 2013). A função objetivo utilizando a norma L2 (Equação 2.4) é uma forma de penalização baseada nos dados, pois utiliza os dados observados e calculados para resolver o problema inverso.

Tipicamente, as regularizações utilizam um termo de penalização baseado em uma norma dos parâmetros do modelo. Essa abordagem auxilia no condicionamento do problema inverso, introduzindo informações a priori que são impostas à função objetivo baseada nos dados. Assim, de acordo com Scales, Smith e Treitel (2001), a regularização de uma função objetivo pode ser escrita como a soma entre a função objetivo baseada nos dados e um termo de regularização baseado nos parâmetros:

$$\chi(\mathbf{m}) = \chi_2(\mathbf{m}) + \lambda_R R(\mathbf{m}). \quad (2.16)$$

onde $\chi_2(\mathbf{m})$ é a norma L2 e λ_R é o parâmetro de regularização que controla o peso da aplicação da regularização. Se $\lambda_R \rightarrow 0$, a função objetivo se torna baseada somente nos dados.

A escolha do operador R determina o método de regularização. A regularização de Tikhonov é amplamente utilizada para a regularização de problemas mal-postos. Essa regularização, também conhecida como regularização de ordem zero, é dada por:

$$R_0(\mathbf{m}) = \int_V |\mathbf{m}(\mathbf{x}) - \mathbf{m}_0(\mathbf{x})|^2 d\mathbf{x}. \quad (2.17)$$

Essa regularização garante que o modelo estimado seja suave, isso quer dizer que a distância Euclidiana do modelo estimado em relação ao modelo inicial $\mathbf{m}_0$ é minimizada, penalizando os grandes valores de $\mathbf{m}$ durante a inversão.

A regularização de Tikhonov de primeira ordem, definida como:

$$R_1(\mathbf{m}) = \int_V |\nabla_{\mathbf{x}} [\mathbf{m}(\mathbf{x}) - \mathbf{m}_0(\mathbf{x})]|^2 d\mathbf{x}, \quad (2.18)$$

assume que o modelo estimado seja suave, aplicando a derivada espacial de primeira ordem $\nabla_{\mathbf{x}}$ no modelo $\mathbf{m}$, enquanto a regularização de Tikhonov de segunda ordem, definida como:

$$R_2(\mathbf{m}) = \int_V |\nabla_{\mathbf{x}}^2 [\mathbf{m}(\mathbf{x}) - \mathbf{m}_0(\mathbf{x})]|^2 d\mathbf{x}, \quad (2.19)$$

aplica a derivada espacial de segunda ordem aos parâmetros do modelo, e tende a penalizar as rugosidades existentes (MENKE, 2012; BOYD; VANDENBERGHE, 2004; ASTER; BORCHERS; THURBER, 2013).

2.3.1 Regularização *Total Variation*

A regularização *Total Variation* (TV) é conhecida por preservar as variações acentuadas (*sharp transition*) no modelo estimado, mesmo em dados contaminados por ruídos (BOYD; VANDENBERGHE, 2004). Diversos autores citam a utilização da regularização TV no algoritmo de FWI para recuperar as interfaces do modelo de velocidade, mantendo a suavidade do modelo (ANAGAW; SACCHI, 2011; ESSER et al., 2014; QU; VERSCHUUR; CHEN, 2017; YONG et al., 2018; SOUZA, 2018).

O termo de regularização TV é definido como:

$$R_{TV}(\mathbf{m}) = \int_V \sqrt{(\nabla_{\mathbf{x}} [\mathbf{m}(\mathbf{x}) - \mathbf{m}_0(\mathbf{x})])^2} d\mathbf{x}. \quad (2.20)$$

Para a aplicação da regularização é necessário discretizar o problema. Assim, o operador nãla $\nabla_{\mathbf{x}}$ assume a forma de um operador de diferença espacial. Na Equação 2.20 é utilizado o operador de diferença de 1ª ordem. Assim, o termo de regularização

por TV discretizado para o caso acústico 2D fica:

$$R_{TV}(\mathbf{m}) = \sum_{i=1}^{Nx} \sum_{j=1}^{Nz} \sqrt{\left(\frac{\Delta v_{pi,j} - \Delta v_{pi-1,j}}{\Delta x} \right)^2 + \left(\frac{\Delta v_{pi,j} - \Delta v_{pi,j-1}}{\Delta z} \right)^2}, \quad (2.21)$$

onde Δv_p é a diferença entre o modelo inicial e modelo atual.

Entretanto, a experiência mostra que a aplicação da regularização TV apenas na direção vertical apresenta melhores resultados (ESSER et al., 2016). Portanto, neste trabalho será usado apenas a direção vertical da regularização TV.

2.4 Métodos de otimização não-linear

Devido ao caráter não-linear da equação da onda e da função objetivo, a solução do problema inverso fará uso de métodos de otimização não-linear. Estes métodos buscam o “melhor” conjunto de parâmetros que explicam o conjunto de dados observados. O “melhor” conjunto de parâmetros será aquele que minimiza a função objetivo utilizada (CHONG; ZAK, 2008).

Nocedal e Wright (2006) explicam que os algoritmos de otimização não-linear são algoritmos iterativos, e necessitam de um modelo inicial $\mathbf{m}_0$ para estimar os parâmetros $\mathbf{m}^*$. A cada iteração do processo de otimização, o modelo é modificado de forma que a função objetivo $\chi(\mathbf{m}_k)$ avaliada com o modelo de parâmetros da iteração atual seja menor ou igual à função objetivo da iteração anterior $\chi(\mathbf{m}_{k-1})$. Portanto:

$$\chi(\mathbf{m}_k) \leq \chi(\mathbf{m}_{k-1}), \quad (2.22)$$

garante uma convergência para um mínimo local. O processo iterativo continua até que algum critério pré-estabelecido seja alcançado, como por exemplo, a definição de um número máximo de iterações ou a definição de uma variação mínima da função objetivo como critério de parada, expresso por:

$$\frac{\chi(\mathbf{m}_k)}{\chi(\mathbf{m}_0)} \leq \epsilon, \quad (2.23)$$

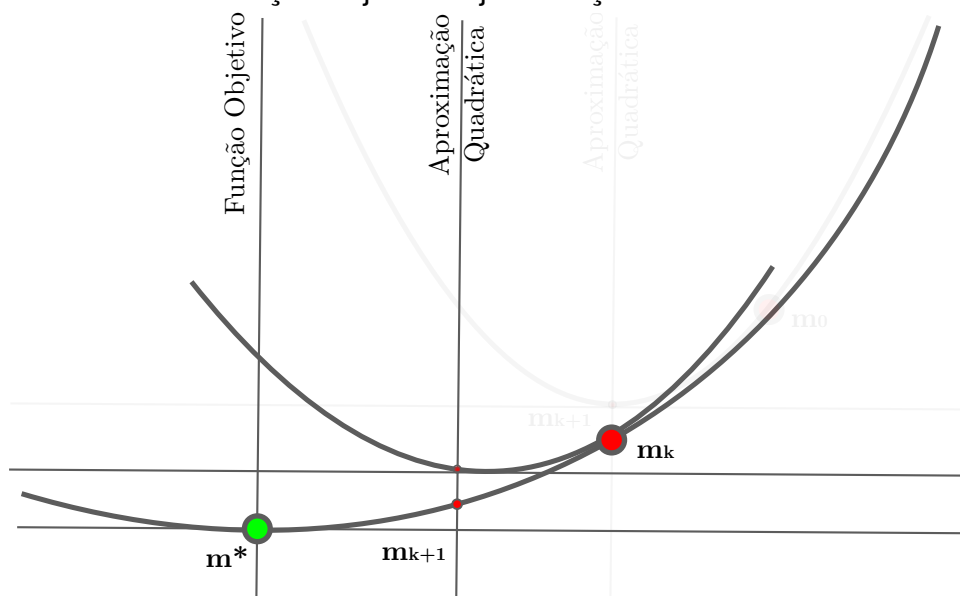
onde $\epsilon \geq 0$ é a tolerância.

2.4.1 Método de Newton

A forma mais natural de se obter a equação que estima os parâmetros do modelo iterativamente é através do estudo do método de Newton. Chong e Zak (2008) afirmam que a ideia central deste é minimizar a aproximação quadrática da função objetivo, ao invés da função objetivo original (Figura 8), e usar a solução que minimiza a função objetivo como entrada para próxima iteração. Esse processo continua iterativamente até que o mínimo da função objetivo original seja alcançado.

Em oposição aos métodos de otimização não-linear, em que a função objetivo tipicamente possui uma forma indefinida, na otimização linear, a função objetivo possui uma forma quadrática, permitindo o método de Newton convergir para o mínimo da função objetivo em apenas uma iteração. A complexidade da função objetivo nos problemas não-lineares exige que o modelo inicial esteja próximo o suficiente de um mínimo global, pois os métodos de otimização baseados no gradiente garantem apenas a convergência para um mínimo local.

Figura 8 – Representação do método de Newton. A função quadrática aproximada pelo método de Newton tangencia a função objetivo por várias iterações até que o mínimo da função objetivo seja alcançado.



Fonte: Modificado de Tarantola (2005).

Inicialmente, é necessário assumir que as perturbações ao modelo são desprezíveis, ou seja, $\mathbf{m} = \mathbf{m}_0 + \Delta\mathbf{m}$ com $\Delta\mathbf{m} \approx 0$. Dessa forma é possível expandir a função objetivo $\chi(\mathbf{m})$ em série de Taylor, em torno do modelo inicial:

$$\chi(\mathbf{m}) = \chi(\mathbf{m}_0 + \Delta\mathbf{m}) = \chi(\mathbf{m}_0) + (\mathbf{m} - \mathbf{m}_0) \nabla \chi(\mathbf{m}) \Big|_{\mathbf{m}=\mathbf{m}_0} + \frac{1}{2!} (\mathbf{m} - \mathbf{m}_0)^2 \nabla^2 \chi(\mathbf{m}) \Big|_{\mathbf{m}=\mathbf{m}_0} + \dots \quad (2.24)$$

O método de Newton utiliza uma aproximação parabólica da função objetivo pois são desprezados os termos maiores do que 2ª ordem na expansão da função objetivo

por série de Taylor (PRATT; SHIN, 1998). Portanto, a função objetivo é aproximada por:

$$\chi(\mathbf{m}) \approx \chi(\mathbf{m}_0) + (\mathbf{m}^* - \mathbf{m}_0) \nabla \chi(\mathbf{m}) \Big|_{\mathbf{m}=\mathbf{m}_0} + \frac{1}{2!} (\mathbf{m}^* - \mathbf{m}_0)^2 \nabla^2 \chi(\mathbf{m}) \Big|_{\mathbf{m}=\mathbf{m}_0} \quad (2.25)$$

Na equação anterior, $\mathbf{m}_0 \in \mathbb{R}^N$ são os parâmetros iniciais e $\mathbf{m}^* \in \mathbb{R}^N$ são os parâmetros estimados. O termo $\nabla \chi \in \mathbb{R}^N$ é chamado de gradiente da função objetivo, que é a 1ª derivada da função objetivo com relação aos parâmetros do modelo. O termo $\nabla^2 \chi \in \mathbb{R}^{N \times N}$ é a 2ª derivada da função objetivo com relação aos parâmetros, também conhecida como matriz Hessiana.

As condições necessárias para que $\mathbf{m}^*$ seja um mínimo local são descritas por Nocedal e Wright (2006) como condições de 1ª e 2ª ordem.

A condição de 1ª ordem garante que $\mathbf{m}^*$ é um mínimo local se a função objetivo χ é continuamente diferenciável na vizinhança aberta de $\mathbf{m}^*$, e além disso:

$$\nabla \chi(\mathbf{m}^*) = 0. \quad (2.26)$$

A condição de 2ª ordem inclui a condição anterior e define que $\mathbf{m}^*$ é um mínimo local se $\nabla^2 \chi$ existe, e é continuamente diferenciável na vizinhança aberta de $\mathbf{m}^*$, e além disso, positiva definida, ou seja:

$$\nabla^2 \chi(\mathbf{m}^*) \geq 0. \quad (2.27)$$

Utilizando a condição de 2ª ordem, é obtida a expressão que estima os parâmetros do modelo e que, simultaneamente, minimiza a função objetivo por mínimos quadrados a partir de um modelo inicial. Então, calculando o gradiente da função objetivo com relação aos parâmetros do modelo e igualando a zero, é encontrado:

$$\nabla_{\mathbf{m}} \chi(\mathbf{m}) \approx \nabla_{\mathbf{m}^*} \left[\chi(\mathbf{m}_0) + (\mathbf{m}^* - \mathbf{m}_0) \nabla \chi(\mathbf{m}_0) + \frac{1}{2!} (\mathbf{m}^* - \mathbf{m}_0)^2 \nabla^2 \chi(\mathbf{m}_0) \right] = 0.$$

Note que os termos $\chi(\mathbf{m}_0)$, $\nabla \chi(\mathbf{m}_0)$ e $\nabla^2 \chi(\mathbf{m}_0)$ não dependem de $\mathbf{m}^*$. Portanto:

$$\begin{aligned}
\nabla\chi(\mathbf{m}_0) + \frac{2}{2!}(\mathbf{m}^* - \mathbf{m}_0)\nabla^2\chi(\mathbf{m}_0) &= 0 \\
(\mathbf{m}^* - \mathbf{m}_0)\nabla^2\chi(\mathbf{m}_0) &= -\nabla\chi(\mathbf{m}_0).
\end{aligned}
\tag{2.28}$$

Desse modo, assumindo que a inversa de $\nabla^2\chi(\mathbf{m}_0)$ existe, é obtida a equação de atualização através do método de Newton:

$$\mathbf{m}^* = \mathbf{m}_0 - [\nabla^2\chi(\mathbf{m}_0)]^{-1}\nabla\chi(\mathbf{m}_0). \tag{2.29}$$

Devido ao caráter altamente não-linear da função objetivo e do problema direto, a estimativa dos parâmetros do modelo deve ser feita iterativamente. Portanto, a Equação 2.29 representa o cálculo da primeira iteração da resolução do problema inverso por mínimos quadrados.

Logo, a equação que estima o modelo iterativamente pelo método de Newton fica:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - [\nabla^2\chi(\mathbf{m}_k)]^{-1}\nabla\chi(\mathbf{m}_k) \tag{2.30}$$

onde o parâmetro estimado na iteração atual é $\mathbf{m}_{k+1}$ e o parâmetro da iteração anterior é $\mathbf{m}_k$.

Tipicamente, em problemas não-lineares com grande número de parâmetros, o cálculo da inversa da matriz Hessiana é computacionalmente proibitivo devido ao seu enorme número de componentes. Para ilustrar a dificuldade de se obter a matriz Hessiana, suponha um caso bidimensional, onde o problema é discretizado em uma malha regular e que o número de elementos na direção x seja N_x e na direção z seja N_z . Portanto, o número total de parâmetros é $N = N_x \times N_z$, logo os parâmetros podem ser escritos como:

$$\mathbf{m} = \begin{pmatrix} m_1 & m_2 & \dots & m_N \end{pmatrix}^T, \tag{2.31}$$

o gradiente da função objetivo também possui N componentes:

$$\nabla\chi(\mathbf{m})\Big|_{\mathbf{m}=\mathbf{m}_0} = \left(\frac{\partial\chi(\mathbf{m}_0)}{\partial m_1} \quad \frac{\partial\chi(\mathbf{m}_0)}{\partial m_2} \quad \cdots \quad \frac{\partial\chi(\mathbf{m}_0)}{\partial m_N} \right)^T, \quad (2.32)$$

e a matriz Hessiana pode ser escrita em sua forma matricial:

$$\mathbf{H} \equiv \nabla^2\chi(\mathbf{m})\Big|_{\mathbf{m}=\mathbf{m}_0} = \begin{pmatrix} \frac{\partial}{\partial m_1} \frac{\partial\chi(\mathbf{m}_0)}{\partial m_1} & \frac{\partial}{\partial m_1} \frac{\partial\chi(\mathbf{m}_0)}{\partial m_2} & \cdots & \frac{\partial}{\partial m_1} \frac{\partial\chi(\mathbf{m}_0)}{\partial m_N} \\ \frac{\partial}{\partial m_2} \frac{\partial\chi(\mathbf{m}_0)}{\partial m_1} & \frac{\partial}{\partial m_2} \frac{\partial\chi(\mathbf{m}_0)}{\partial m_2} & \cdots & \frac{\partial}{\partial m_2} \frac{\partial\chi(\mathbf{m}_0)}{\partial m_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial m_N} \frac{\partial\chi(\mathbf{m}_0)}{\partial m_1} & \frac{\partial}{\partial m_N} \frac{\partial\chi(\mathbf{m}_0)}{\partial m_2} & \cdots & \frac{\partial}{\partial m_N} \frac{\partial\chi(\mathbf{m}_0)}{\partial m_N} \end{pmatrix}. \quad (2.33)$$

Portanto, a matriz Hessiana $\mathbf{H}$ possui N^2 componentes. Além disso, a Equação 2.30 necessita do cálculo da inversa da matriz Hessiana, $\mathbf{H}^{-1}$, que torna a maioria dos casos com grande número de parâmetros impossíveis de resolver na prática.

De acordo com Pratt e Shin (1998), a inversa da Hessiana atua como um filtro que focaliza o gradiente da função objetivo. Além disso, a $\mathbf{H}^{-1}$ permite a atenuação dos artefatos no gradiente. É importante ressaltar que teoricamente um resultado semelhante pode ser alcançado usando um método baseado apenas no gradiente realizando iterações suficientes para compensar a aplicação da Hessiana.

Em problemas inversos não-lineares que são proibitivos os cálculos da inversa da matriz Hessiana, utilizam-se os métodos que fazem uso da aproximação da inversa da Hessiana (métodos quasi-Newton) ou métodos que evitam o cálculo da Hessiana (método do máximo declive ou método do gradiente conjugado). Tais técnicas, após obtida a direção de atualização através do cálculo do gradiente da função objetivo, resolvem outro problema inverso unidimensional que determina o tamanho do passo ótimo para a atualização do modelo.

2.4.2 Método do máximo declive (gradiente descendente)

De acordo com Nocedal e Wright (2006), entre os métodos de otimização baseados no gradiente, o método do máximo declive (do inglês, *steepest descent*) é o mais simples. A direção escolhida para a atualização é simplesmente a direção do gradiente

da função objetivo. Como o gradiente de uma função aponta para a direção de máximo crescimento dessa mesma função, então a escolha do negativo dessa direção ($-\nabla\chi(\mathbf{m})$) é a mais óbvia.

Uma vez determinada a direção de atualização do modelo, é necessário determinar um tamanho de passo que garanta a diminuição da função objetivo. Essa condição pode ser expressa matematicamente por:

$$\chi(\mathbf{m}_k - \alpha_k \nabla\chi(\mathbf{m}_k)) < \chi(\mathbf{m}_k), \quad (2.34)$$

onde α_k é conhecido como tamanho do passo (*step size*). Essa inequação define que para uma mesma direção de atualização $\nabla\chi(\mathbf{m}_k)$, o passo utilizado deve garantir a diminuição da função objetivo do modelo estimado. A solução do problema inverso unidimensional para a determinação do tamanho do passo que garante a diminuição da função objetivo é normalmente chamada de busca linear (*line search*).

Assim, o algoritmo que implementa essa ideia supõe que a partir de um conjunto de parâmetros iniciais $\mathbf{m}_k$, é possível estimar o próximo conjunto de parâmetros $\mathbf{m}_{k+1}$, utilizando a direção de atualização $-\alpha_k \nabla\chi(\mathbf{m}_k)$, onde α_k é o tamanho do passo, um escalar positivo. Portanto, a equação que atualiza o modelo de parâmetros iterativamente é:

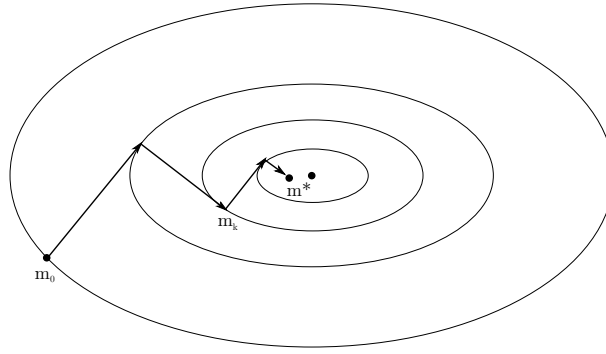
$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \nabla\chi(\mathbf{m}_k). \quad (2.35)$$

A Equação 2.35 é um algoritmo de otimização, que utiliza a estratégia de busca linear, uma vez definida a direção de atualização $\nabla\chi(\mathbf{m}_k)$. Essa classe de algoritmo busca um valor de α_k que garanta o decaimento da função objetivo (inequação 2.34). Após encontrar o mínimo, uma nova direção de atualização é calculada, e o processo se reinicia (Figura 9).

2.4.3 Método do gradiente conjugado

O método do gradiente conjugado pode ser visto como uma classe intermediária entre o método de Newton e o método do gradiente descendente (CHONG; ZAK, 2008). Esse

Figura 9 – Representação dos diferentes passos do método Gradiente Descendente.



Fonte: Nocedal e Wright (2006)

método evita caminhar por direções repetidas, utilizando a informação das direções de atualização dos passos anteriores. Essa abordagem tipicamente possui melhores propriedades de convergência que o gradiente descendente. Existem diversas formas de calcular a direção de atualização para o método gradiente conjugado, porém todas elas partem da equação de atualização:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \mathbf{h}_k, \quad (2.36)$$

onde $\mathbf{h}_k$ é a direção, que utiliza as informações dos gradientes anteriores, definida por:

$$\mathbf{h}_{k+1} = \nabla \chi(\mathbf{m}_{k+1}) + \beta^k \mathbf{h}_k, \quad (2.37)$$

onde β_k é um escalar que controla o cálculo da direção de atualização. No caso da primeira iteração, a direção de atualização é o próprio gradiente da função objetivo.

$$\mathbf{h}_0 = \nabla \chi(\mathbf{m}_0). \quad (2.38)$$

As direções subsequentes são obtidas utilizando a propriedade de conjugação (*conjugacy*), garantindo que é possível minimizar a função objetivo quadrática em N passos, sendo N o número de parâmetros do modelo (NOCEDAL; WRIGHT, 2006; CHONG; ZAK, 2008). As direções de atualização $\mathbf{h}_k$ são calculadas por meio de uma combinação linear das direções anteriores e da direção atual, de modo que todas as direções sejam mutuamente conjugadas. É recomendado que após um determinado

número de passos (dependendo do problema) a direção de atualização seja reiniciada. As diferentes formas para o cálculo da direção de atualização são obtidas a partir da definição de β_k na Equação 2.37.

De acordo com Chong e Zak (2008), as principais formas de cálculo das direções de atualização para problemas não-lineares, utilizando o método gradiente conjugado, são baseadas na manipulação algébrica de como β_k é obtido. Existem 3 fórmulas bem conhecidas, Hestenes-Stiefel:

$$\beta_k^{HS} = \frac{\nabla\chi(\mathbf{m}_{k+1}) \cdot [\nabla\chi(\mathbf{m}_{k+1}) - \nabla\chi(\mathbf{m}_k)]}{\mathbf{h}_k \cdot [\nabla\chi(\mathbf{m}_{k+1}) - \nabla\chi(\mathbf{m}_k)]}, \quad (2.39)$$

Polak-Ribière:

$$\beta_k^{PR} = \frac{\nabla\chi(\mathbf{m}_{k+1}) \cdot [\nabla\chi(\mathbf{m}_{k+1}) - \nabla\chi(\mathbf{m}_k)]}{\nabla\chi(\mathbf{m}_k) \cdot \nabla\chi(\mathbf{m}_k)}, \quad (2.40)$$

e, Fletcher-Reeves:

$$\beta_k^{FR} = \frac{\nabla\chi(\mathbf{m}_{k+1}) \cdot \nabla\chi(\mathbf{m}_{k+1})}{\nabla\chi(\mathbf{m}_k) \cdot \nabla\chi(\mathbf{m}_k)}. \quad (2.41)$$

Não existe unanimidade em afirmar qual a melhor fórmula, porém sabe-se que a busca linear exerce papel crucial na performance do algoritmo de gradiente conjugado. Nos casos em que a busca linear é imprecisa, recomenda-se a aplicação da fórmula Hestenes-Stiefel. A fórmula de Polak-Ribière é conhecida por possuir melhor performance, porém isso não é uma regra. Entretanto, a análise da convergência global sugere que a fórmula de Fletcher-Reeves é a melhor (CHONG; ZAK, 2008).

Uma quarta fórmula foi proposta por Dai e Yuan (1999), que tenta determinar a direção de atualização que garanta que as condições de Wolfe para a busca linear sejam respeitadas:

$$\beta_k^{DY} = \frac{\nabla\chi(\mathbf{m}_{k+1}) \cdot \nabla\chi(\mathbf{m}_{k+1})}{\mathbf{h}_k \cdot [\nabla\chi(\mathbf{m}_{k+1}) - \nabla\chi(\mathbf{m}_k)]}. \quad (2.42)$$

A biblioteca de algoritmos de otimização SEISCOPE (MÉTIVIER, 2014), utilizada nesse trabalho, usa a fórmula de Dai-Yuan para o cálculo das direções do gradiente conjugado. Essa equação permite que o mesmo algoritmo de busca linear usado no método gradiente descendente seja usado no gradiente conjugado.

2.4.4 Método quasi-Newton (LBFGS)

O método quasi-Newton também é baseado exclusivamente no gradiente da função objetivo, entretanto, ele possui uma taxa de convergência superior ao método gradiente conjugado e, em alguns casos, possui uma convergência melhor que o método de Newton (NOCEDAL; WRIGHT, 2006).

O método quasi-Newton surgiu a partir da dificuldade de estimar a inversa da matriz Hessiana $[\nabla^2 \chi(\mathbf{m}_k)]^{-1}$, para problemas não-lineares com grande quantidade de parâmetros. De acordo com Chong e Zak (2008), para evitar o cálculo de $[\nabla^2 \chi(\mathbf{m}_k)]^{-1}$, a matriz Hessiana é aproximada:

$$\mathbf{B}_k \approx \nabla^2 \chi(\mathbf{m}_k), \quad (2.43)$$

ao mesmo tempo que a Equação de atualização 2.30 é modificada para garantir que o algoritmo possua propriedades descendentes:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \mathbf{B}_k^{-1} \nabla \chi(\mathbf{m}_k), \quad (2.44)$$

onde α_k é escolhido de forma a garantir as propriedades de convergência $\chi(\mathbf{m}_{k+1}) < \chi(\mathbf{m}_k)$.

Segundo Nocedal e Wright (2006), os pesquisadores Broyden, Fletcher, Goldfarb e Shanno desenvolveram de forma independente o algoritmo mais utilizado para o método quasi-Newton, que é atualmente o mais efetivo entre todas as fórmulas de atualização.

O algoritmo BFGS possui um custo por iteração inferior ao método de Newton, pois evita o cálculo das segundas derivadas do gradiente da função objetivo. Por isso,

na maior parte dos problemas práticos os métodos quasi-Newton possuem melhor performance. A aproximação da Hessiana é calculada através do método BFGS por:

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{B}_k}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k} + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}, \quad (2.45)$$

onde:

$$\begin{aligned} \mathbf{s}_k &= \mathbf{m}_k - \mathbf{m}_{k-1} \\ e \\ \mathbf{y}_k &= \nabla \chi(\mathbf{m}_k) - \nabla \chi(\mathbf{m}_{k-1}). \end{aligned} \quad (2.46)$$

A inversa de $\mathbf{B}_{k+1}$ é obtida através de:

$$\mathbf{B}_{k+1}^{-1} = \left(\mathbf{I} - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) \mathbf{B}_k^{-1} \left(\mathbf{I} - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}, \quad (2.47)$$

onde $\mathbf{I}$ é a matriz identidade.

Byrd et al. (1994) apresentou um algoritmo eficiente para a estimativa da aproximação da inversa da matriz Hessiana, utilizando as matrizes de memória limitada do método BFGS (método LBFGS). O algoritmo armazena em cada iteração um pequeno número de pares m de correção $[\mathbf{s}_i, \mathbf{y}_i]$ obtidos pela Equação 2.46 com $i = k-1, \dots, k-m$. Os m pares de correção contêm informação sobre a curvatura da função objetivo, e em conjunto com a fórmula BFGS (Equação 2.45), definem uma versão da matriz $\mathbf{B}_k$ de memória limitada. Assim, o produto $\mathbf{B}_k^{-1} \nabla \chi(\mathbf{m}_k)$ pode ser obtido através de uma sequência de somas e produtos internos entre $\nabla \chi(\mathbf{m}_k)$ e os m pares mais recentes $[\mathbf{s}_i, \mathbf{y}_i]$.

2.4.5 Busca linear

Os métodos de otimização, baseados no gradiente, que evitam o cálculo da matriz Hessiana garantem a sua convergência através da escolha de um passo α_k , que obtenha uma função objetivo menor que a iteração anterior (Equação 2.34). Tais

métodos de otimização podem ser sintetizados em uma única equação de atualização:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \mathbf{h}_k, \quad (2.48)$$

onde a escolha do cálculo da direção de atualização de $\mathbf{h}_k$ define o método de otimização utilizado.

Os algoritmos que determinam os valores de α_k , tipicamente, testam uma sequência de possíveis valores, parando quando a condição de convergência é satisfeita. Devido ao alto custo para solução do problema direto e para o cálculo do gradiente da função objetivo, a busca linear atua em duas fases: a primeira seleciona um intervalo de possíveis valores para o passo, a segunda fase calcula um bom valor de passo dentro do intervalo escolhido (NOCEDAL; WRIGHT, 2006).

Para garantir uma diminuição consistente da função objetivo e a convergência para um mínimo local, são aplicadas as condições de Wolfe (MÉTIVIER; BROSSIER, 2016). A primeira é conhecida como condição de redução suficiente, ou condição de Armijo:

$$\chi(\mathbf{m}_k - \alpha_k \mathbf{h}_k) \leq \chi(\mathbf{m}_k) + c_1 \alpha \nabla_k \chi(\mathbf{m}_k) \mathbf{h}_k, \quad (2.49)$$

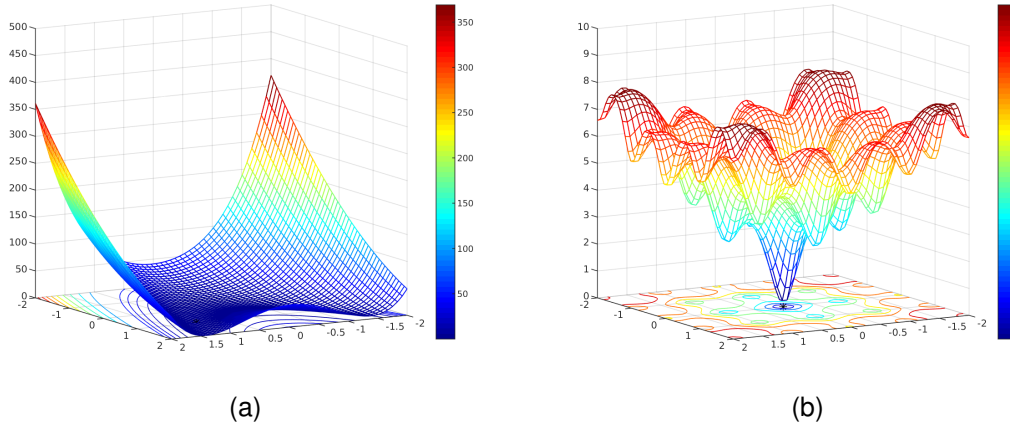
e a segunda é conhecida como condição de curvatura:

$$\nabla \chi(\mathbf{m}_k + \alpha_k \mathbf{h}_k) \mathbf{h}_k \geq c_2 \nabla \chi(\mathbf{m}_k) \mathbf{h}_k, \quad (2.50)$$

com $c_1 = 10^{-4}$ e $c_2 = 0,9$.

A primeira condição impõe que o passo α deve ser calculado de modo que garanta uma atualização na direção $\mathbf{h}_k$, que diminua suficientemente a função objetivo. A segunda condição é usada para desconsiderar valores de α pequenos que levem a uma atualização muito pequena da estimativa de $\mathbf{m}_k$. Vale ressaltar que a cada teste das condições de Wolfe é necessário realizar uma modelagem direta e o cálculo do gradiente da função objetivo. A escolha dos parâmetros c_1 e c_2 é empírica, restrita aos seguintes intervalos $c_1 \in (0, 1)$ e $c_2 \in (c_1, 1)$ (MÉTIVIER, 2014).

Figura 10 – Funções objetivo utilizadas nos testes de performance: a) Função objetivo Rosenbrock. b) Função objetivo Ackley.



2.4.6 Exemplos numéricos

A análise da performance entre os algoritmos de otimização não-linear pode ser realizada com as funções de teste Rosenbrock e Ackley como funções objetivo. A função Rosenbrock (Figura 11a) possui a seguinte expressão:

$$\chi(x_1, x_2) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2; \quad (2.51)$$

e uma forma suave em torno de seu mínimo global. A forma peculiar da função Rosenbrock permite avaliar o comportamento do método de otimização não linear, em condições onde a função objetivo possui apenas um mínimo local em uma região com valores muito próximos. De certa forma, essa função avalia a precisão do método de otimização. O mínimo global da função objetivo Rosenbrock pode ser calculado analiticamente, e está em $x_1 = 1$ e $x_2 = 1$.

A função Ackley (Figura 11b) é expressa por:

$$\chi(x_1, x_2) = -20e^{-\frac{1}{5}\sqrt{\frac{1}{2}(x_1^2+x_2^2)}} - e^{\frac{1}{2}[\cos(2\pi x_1)+\cos(2\pi x_2)]} + e + 20. \quad (2.52)$$

A função Ackley possui muitos mínimos locais isolados entre si, em contraposição a função Rosenbrock. Como todos métodos de otimização baseados no gradiente não garantem a convergência para um mínimo global, apenas garantem a convergência

para um mínimo qualquer, a função Ackley avalia a capacidade do método de otimização alcançar o mínimo global. De certa maneira, essa função avalia a acurácia do método de otimização. O mínimo global da função Ackley também pode ser obtido analiticamente e está em $x_1 = 0$ e $x_2 = 0$. É importante ressaltar que a escolha do modelo inicial, próximo o suficiente do mínimo global, é crucial para a convergência do método para o mínimo correto.

Os métodos de otimização citados estão implementados na biblioteca de otimização SEISCOPE (MÉTIVIER; BROSSIER, 2016). Serão analisadas as performances dos métodos de gradiente descendente, gradiente conjugado e LBFGS, através das funções de teste Rosenbrock e Ackley. Em todos os casos, o modelo inicial para a função Rosenbrock foi de $x_1 = -0,5$ e $x_2 = 0,5$, e para a função Ackley foi de $x_1 = -1,5$ e $x_2 = 0,5$. Os critérios de parada utilizados foram a tolerância em relação à função objetivo inicial:

$$\frac{\chi(\mathbf{m}_k)}{\chi(\mathbf{m}_0)} \leq 10^{-8}, \quad (2.53)$$

que garante que o método converge quando a função objetivo for muito pequena em comparação à função objetivo inicial. E a tolerância em relação a função objetivo anterior:

$$1 - \frac{\chi(\mathbf{m}_k)}{\chi(\mathbf{m}_{k-1})} \leq 10^{-4}, \quad (2.54)$$

que impõe que a função objetivo deve variar, significativamente, entre as iterações. Caso contrário a otimização é interrompida. E também, foi limitado o número máximo de iterações permitidas, igual a 10000.

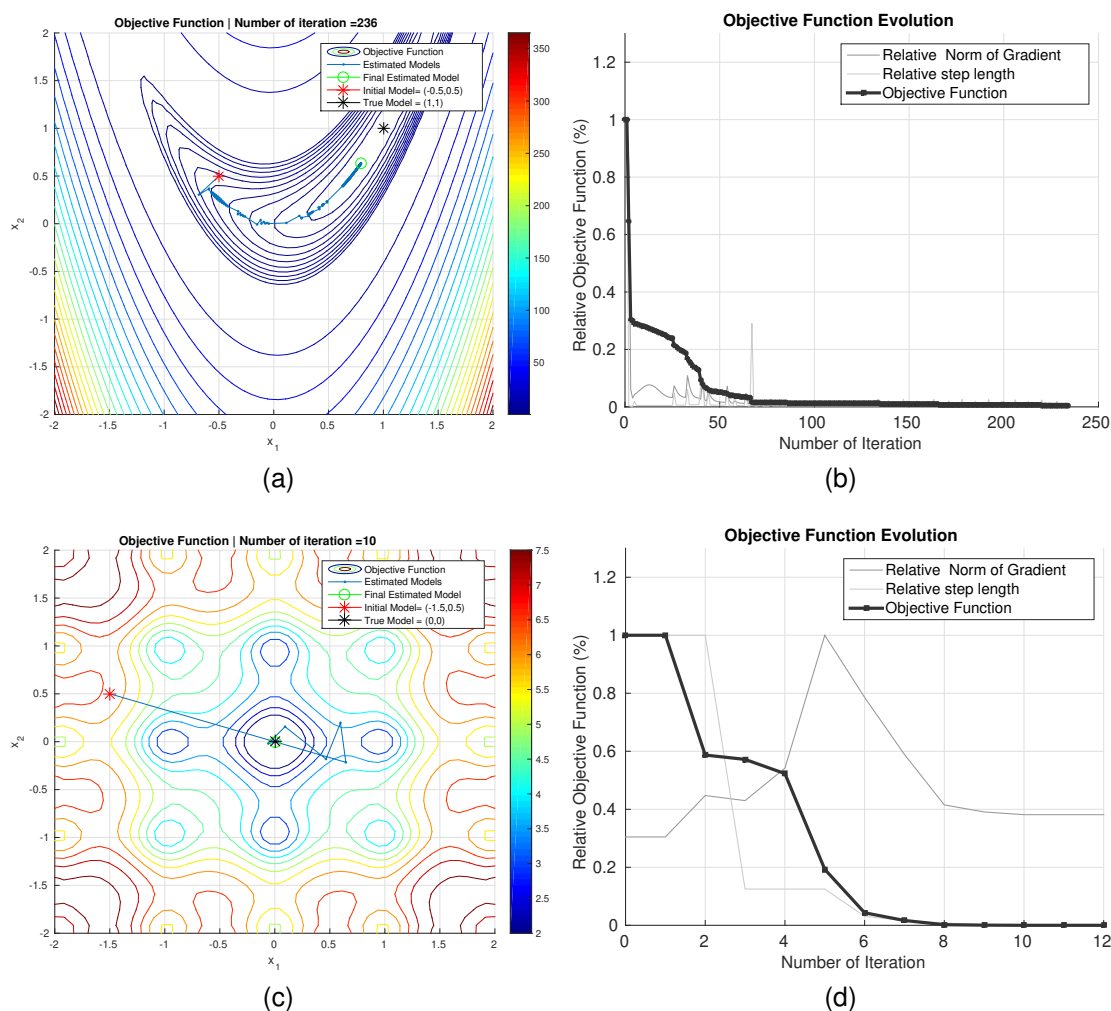
Gradiente descendente

Nos testes utilizando o método gradiente descendente (Figura 11), o mínimo global foi alcançado apenas no caso da função Ackley. O gradiente descendente não alcançou o mínimo global da função Rosenbrock (Figura 12a), pois o critério de parada, que não permite variações insignificantes da função objetivo interrompeu o algoritmo. Um segundo teste foi realizado, e foram necessárias 4966 iterações para alcançar o mínimo global. A função objetivo relativa para esse caso (Figura 12b) alcançou valores

próximos de zero, antes de 100 iterações, porém as atualizações não produziram variações significantes na função objetivo, interrompendo-a precocemente.

No caso da função Ackley, foram necessárias 10 iterações (Figura 12c) para atingir o mínimo global. Como a função Ackley possui os mínimos locais isolados, a função objetivo alcançou o mínimo rapidamente (Figura 12d), não sofrendo o mesmo problema que a função Rosenbrock.

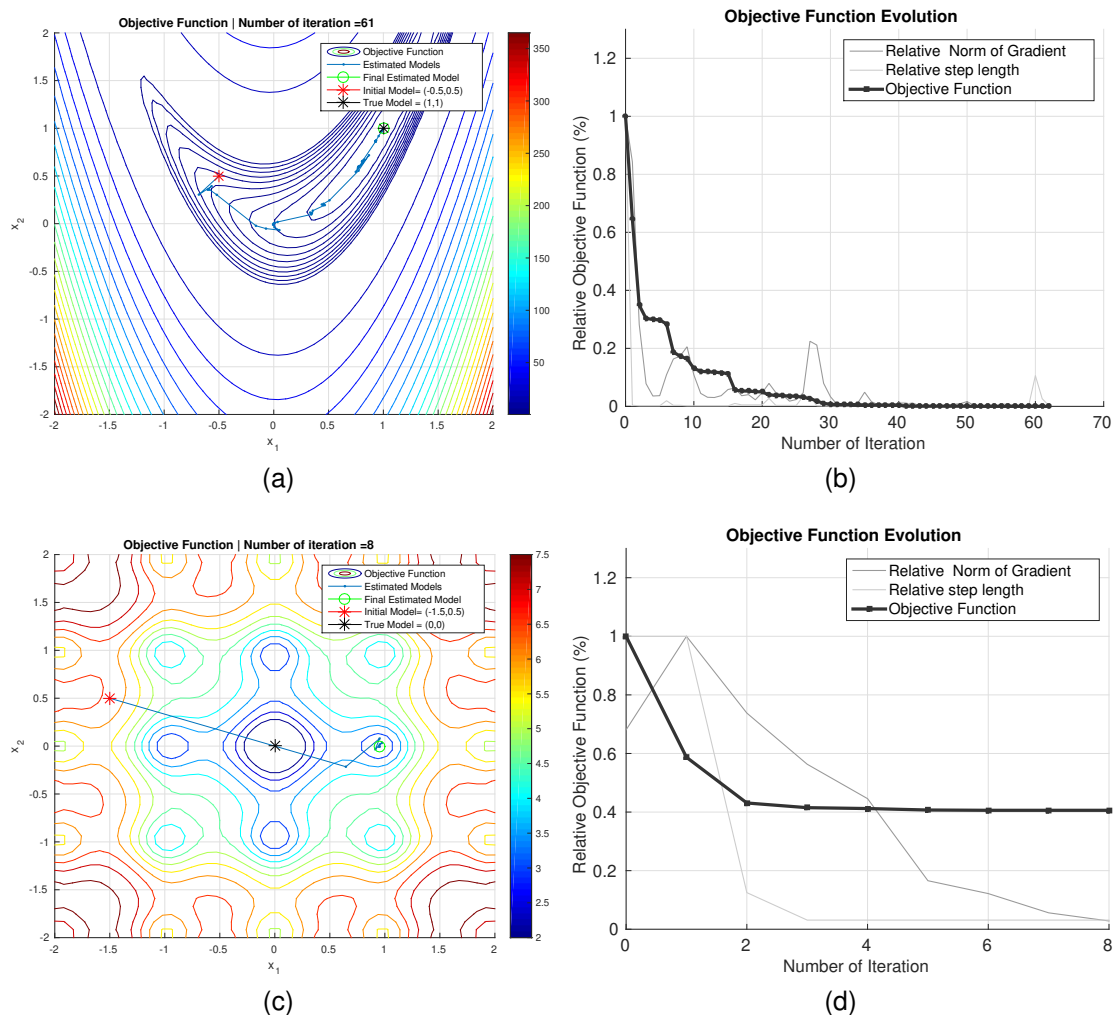
Figura 11 – Funções objetivo utilizadas nos testes de performance do método gradiente descendente. a) Atualização do modelo utilizando a função objetivo Rosenbrock. b) Evolução em valores relativos da função objetivo Rosenbrock. c) Atualização do modelo utilizando a função objetivo Ackley. d) Evolução em valores relativos da função objetivo Ackley.



Gradiente conjugado

O método gradiente conjugado alcançou o mínimo global da função Rosenbrock (Figura 13a), com os critérios de parada impostos, em 61 iterações, enquanto utilizou

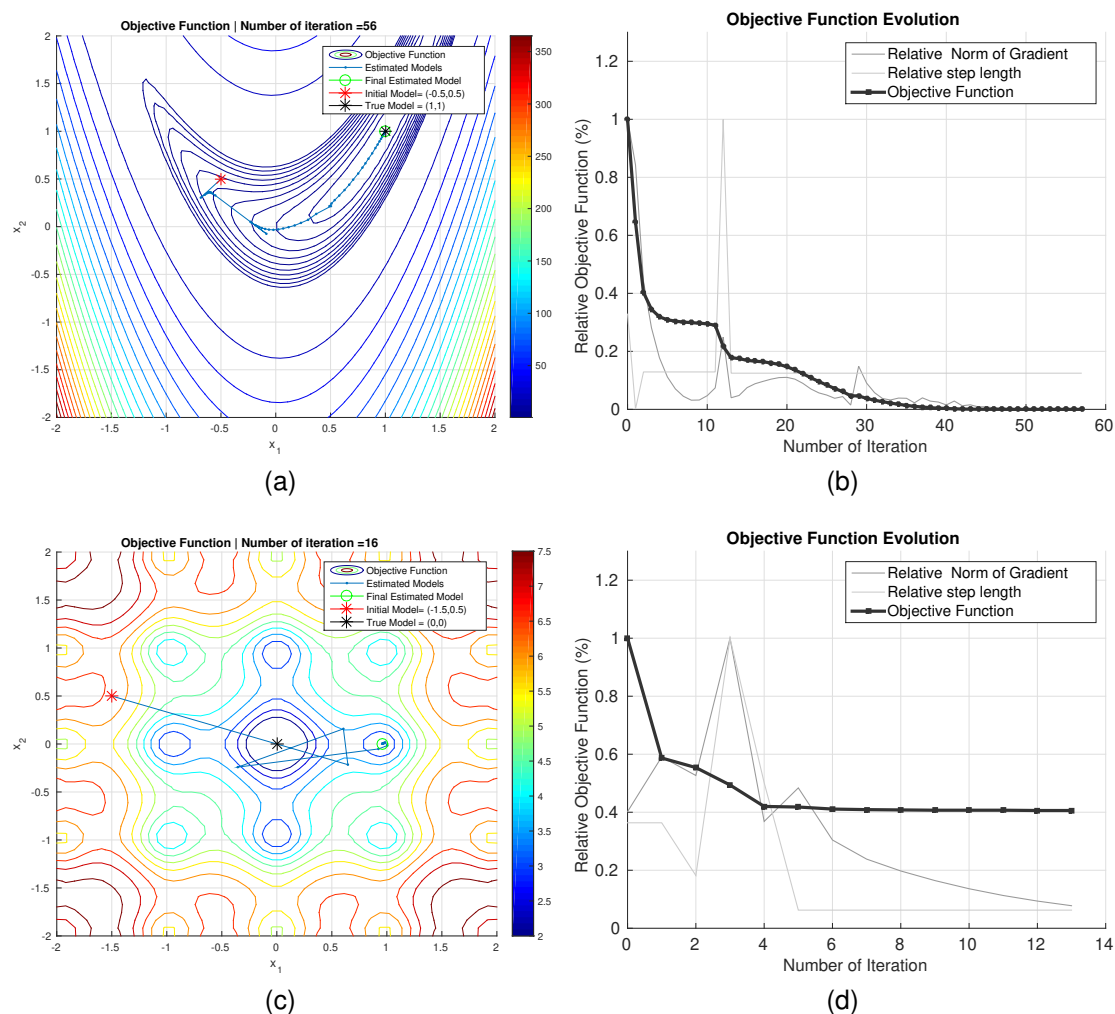
Figura 12 – Funções objetivo utilizadas nos testes de performance do método gradiente conjugado. a) Atualização do modelo utilizando a função objetivo Rosenbrock. b) Evolução em valores relativos da função objetivo Rosenbrock. c) Atualização do modelo utilizando a função objetivo Ackley. d) Evolução em valores relativos da função objetivo Ackley.



8 iterações para atingir um mínimo local da função Ackley (Figura 13c). É evidente a performance superior do método gradiente conjugado em relação ao gradiente descendente no caso da função Rosenbrock. A análise da função objetivo relativa (Figura 13b) confirma sua superioridade.

Entretanto, no caso da função *Ackley*, o método gradiente conjugado levou a um mínimo local. O critério de parada, que não permite variações insignificantes da função objetivo, fez o método ser interrompido rapidamente, não insistindo na busca de uma resposta errada. Caso não fosse aplicado esse critério de parada, o algoritmo gradiente conjugado utilizaria 120. Como é possível observar na Figura 13d, a função objetivo relativa fica distante do zero, indicando que o mínimo global não foi encontrado.

Figura 13 – Funções objetivo utilizadas nos testes de performance do método quasi-Newton LBFGS. a) Atualização do modelo utilizando a função objetivo Rosenbrock b) Evolução em valores relativos da função objetivo Rosenbrock. c) Atualização do modelo utilizando a função objetivo Ackley d) Evolução em valores relativos da função objetivo Ackley.



LBFGS

O método quasi-Newton LBFGS alcançou o mínimo da função Rosenbrock em 56 iterações (Figura 14a), mais eficiente que o gradiente conjugado. Mas teve um comportamento semelhante quando utilizada a função Ackley (Figura 14c), ficando aprisionado também em um mínimo local. Se não for aplicado o critério de parada que evita atualizações insignificantes da função objetivo são realizadas 38 iterações. Um número muito inferior em relação ao gradiente conjugado. As análises das funções objetivo relativas (Figuras 14b e 14d), permitem confirmar a melhor performance para o caso da função Rosenbrock e a falha em alcançar o mínimo global, no caso da função Ackley.

2.5 Cálculo do gradiente da função objetivo e a derivada de Fréchet

O cálculo do gradiente da função objetivo, que permite a estimativa da direção de atualização no espaço dos modelos (Equação 2.29), é uma das etapas que mais demanda recursos computacionais, junto com o cálculo do problema direto. Na literatura, a norma L2 geralmente é utilizada para o cálculo do gradiente da função objetivo, entretanto, nada impede que outra norma seja utilizada, como será apresentado a seguir.

Lailly (1983) e Tarantola (1984) foram os primeiros autores a propor o cálculo do gradiente da função objetivo aplicado ao problema sísmico de forma eficiente. Em problemas com poucos parâmetros a serem determinados, o gradiente da função objetivo pode ser calculado diretamente através da estimativa das derivadas de Fréchet, que possuem um alto custo computacional principalmente em caso com muitos parâmetros.

O modelo estimado pela FWI é obtido através da solução do problema de otimização, definido pelas Equações 2.35, 2.36 ou 2.44, respectivamente, os métodos de gradiente descendente, gradiente conjugado ou quasi-Newton LBFGS. Nesses casos, é necessário o cálculo do gradiente da função objetivo $\nabla\chi(\mathbf{m}_k)$ utilizando o modelo $\mathbf{m}_k$.

Utilizando a norma L2 (Equação 2.4), o cálculo do gradiente da função objetivo fica:

$$\begin{aligned}
 \nabla\chi_2(\mathbf{m}) &= \frac{\partial\nabla\chi_2(\mathbf{m})}{\partial\mathbf{m}} = \frac{\partial}{\partial\mathbf{m}} \frac{1}{2} \int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}|^2 d\mathbf{x}dt \\
 \nabla\chi_2(\mathbf{m}) &= \frac{1}{2} \int_T \int_V \frac{\partial}{\partial\mathbf{m}} |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}|^2 d\mathbf{x}dt \\
 \nabla\chi_2(\mathbf{m}) &= \frac{2}{2} \int_T \int_V \frac{\partial\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t)}{\partial\mathbf{m}} \cdot (\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}) d\mathbf{x}dt \\
 \nabla\chi_2(\mathbf{m}) &= \int_T \int_V \frac{\partial\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t)}{\partial\mathbf{m}} \cdot \Delta\mathbf{d} d\mathbf{x}dt,
 \end{aligned} \tag{2.55}$$

onde Δd é o resíduo (diferença entre o dado calculado e dado observado) e $\frac{\partial d_{cal}}{\partial \mathbf{m}}$ é a derivada de Fréchet (VIRIEUX; OPERTO, 2009). No caso acústico, os dados calculados são obtidos a partir da solução da equação da onda (algoritmo 1) para o modelo $\mathbf{m}$, assim, a derivada de Fréchet pode ser definida como a derivada do campo de pressão $p(\mathbf{x}, t)$, em relação a todos os pontos do modelo $\mathbf{m}$.

A aplicação da definição da derivada parcial às derivadas de Fréchet ilustra a demanda de esforços computacionais proibitivos. Considere o modelo $\mathbf{m}$, em duas dimensões, discretizado em $N = N_x \cdot N_z$ pontos. A derivada de Fréchet fica:

$$\frac{\partial p(\mathbf{m}; \mathbf{x}, t)}{\partial \mathbf{m}} = \lim_{\varepsilon \rightarrow 0} \frac{p(m_1, m_2, \dots, m_i + \varepsilon \Delta m_i, \dots, m_N; \mathbf{x}, t) - p(m_1, m_2, \dots, m_i, \dots, m_N; \mathbf{x}, t)}{\varepsilon} \quad (2.56)$$

onde o modelo $\mathbf{m}$ foi destacado para indicar que a derivada de Fréchet precisa resolver $N + 1$ problemas diretos para calcular a derivada do campo de pressão em cada ponto do modelo (PRATT; SHIN, 1998). A inviabilidade desse cálculo levou Lailly (1983) e Tarantola (1984) a aplicarem o método do estado adjunto, para calcular o gradiente da função objetivo, evitando a estimativa das derivadas de Fréchet.

2.6 O método do estado adjunto e os multiplicadores de Lagrange

O método do estado adjunto calcula o gradiente da função objetivo sem a necessidade das derivadas de Fréchet. Esse método minimiza um funcional (função objetivo) aumentado, também conhecido como método dos multiplicadores de Lagrange associado (PLESSIX, 2006). Para determinar $\nabla \chi(\mathbf{m})$, primeiramente, considere o funcional definido pela norma L2 (Equação 2.4), ou seja, um problema de minimização por mínimos quadrados:

$$\chi_2(\mathbf{m}) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{t=1}^{N_t} \sum_{r=1}^{N_r} |d_{cal}^s(\mathbf{m}; \mathbf{x}^r, t) - d_{obs}^s|^2,$$

onde os índices s , t e r estão relacionados a numeração das fontes, as amostras de tempo e a numeração dos receptores para cada fonte, respectivamente. N_s , N_t e N_r

são os números de fontes, os números de amostras de tempo por tiro e os números de receptores por tiro.

E a equação da onda acústica (Equação A.56) escrita como um operador:

$$\begin{aligned} \nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) &= w(\mathbf{x}, t) \\ \left[\nabla^2 - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} \right] \cdot p(\mathbf{x}, t) &= w(\mathbf{x}, t) \\ \mathcal{L}[\mathbf{m}] \cdot p &= w, \end{aligned} \quad (2.57)$$

onde o parâmetro do modelo a ser estimado é o modelo de velocidade $\mathbf{m} = v_p$.

De acordo com Plessix (2006), o problema de minimização, utilizando o método dos multiplicadores associado, assume que $\mathbf{m}$ é o mínimo da função objetivo $\chi_2(\mathbf{m})$ vinculado ao problema direto $\mathcal{L}[\mathbf{m}] \cdot p = w$, se os campos p e λ_L são pontos de sela da Lagrangeana associada $L(p, \lambda_L, \mathbf{m})$:

$$L(p, \lambda_L, \mathbf{m}) = \frac{1}{2} \int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}|^2 d\mathbf{x} dt - \lambda_L \cdot \int_T \int_V \Psi(p, \mathbf{m}) d\mathbf{x} dt \quad (2.58)$$

onde $\lambda_L = \lambda_L(\mathbf{x}, t)$ é o multiplicador de Lagrange, que equivale ao campo adjunto $p^\dagger(\mathbf{x}, t)$, e $\Psi(p, \mathbf{m})$ o problema direto vinculado, dado por:

$$\Psi(p, \mathbf{m}) = \mathcal{L}[\mathbf{m}] \cdot p - w. \quad (2.59)$$

Substituindo a Equação 2.59 na Equação 2.58, a Lagrangeana associada, ou funcional, que deve ser minimizado é escrito como:

$$L(p, \lambda_L, \mathbf{m}) = \frac{1}{2} \int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}|^2 d\mathbf{x} dt - \int_T \int_V \lambda_L \left[\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} - w \right] d\mathbf{x} dt. \quad (2.60)$$

Sabendo que a diferencial da Lagrangeana é dado por:

$$\delta L(p, \lambda_L, \mathbf{m}) = \left(\frac{\partial L}{\partial p}, \frac{\partial L}{\partial \lambda_L}, \frac{\partial L}{\partial \mathbf{m}} \right) (\delta p, \delta \lambda_L, \delta \mathbf{m}), \quad (2.61)$$

assim, o gradiente da Lagrangeana pode ser escrito como:

$$\nabla L(p, \lambda_L, \mathbf{m}) = \left(\frac{\partial L}{\partial p}, \frac{\partial L}{\partial \lambda_L}, \frac{\partial L}{\partial \mathbf{m}} \right) \quad (2.62)$$

Impondo que p e λ_L sejam pontos de sela, as derivadas da Lagrangeana associada com relação aos campos p e λ_L são iguais a zero: $\frac{\partial L}{\partial p} = 0$ e $\frac{\partial L}{\partial \lambda_L} = 0$. Assim, o gradiente da Lagrangeana será dado por (PLESSIX, 2006):

$$\nabla \chi(\mathbf{m}) = \nabla L(p, \lambda_L, \mathbf{m}) = \frac{\partial L}{\partial \mathbf{m}} \quad (2.63)$$

2.6.1 Derivada da Lagrangeana com relação aos parâmetros

A derivada da Lagrangeana L , com relação aos parâmetros do modelo $\mathbf{m}$, pode ser obtida aplicando a definição de derivada:

$$\frac{\partial L(p, \mathbf{m}, \lambda_L)}{\partial \mathbf{m}} = \lim_{\varepsilon \rightarrow 0} \frac{L(p, \lambda_L, \mathbf{m} + \varepsilon \delta \mathbf{m}) - L(p, \lambda_L, \mathbf{m})}{\varepsilon}. \quad (2.64)$$

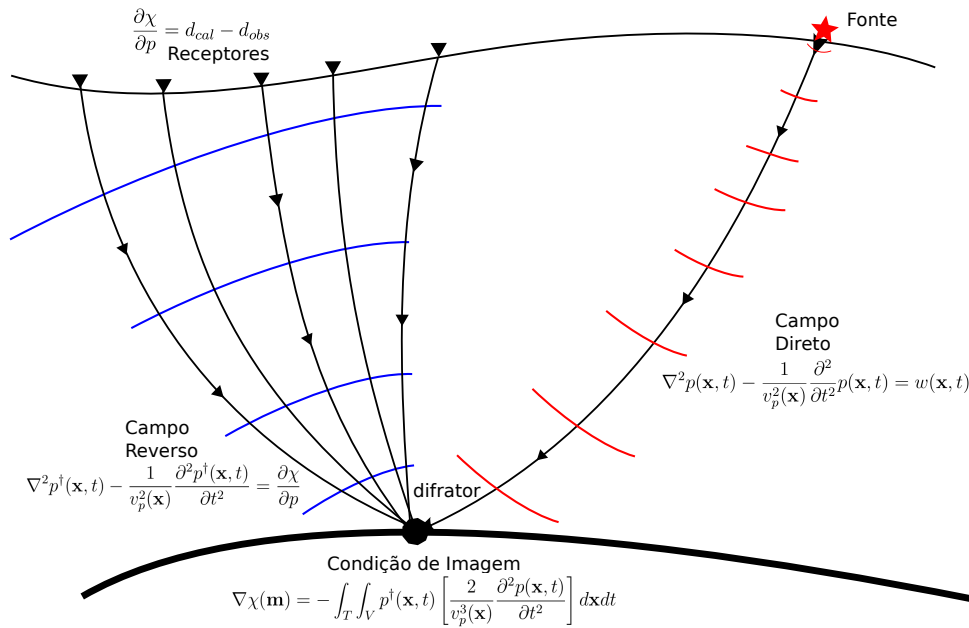
Note que, na aplicação da derivada parcial $\frac{\partial L}{\partial \mathbf{m}}$ os campos λ_L e p são independentes de $\mathbf{m}$. Por esse motivo, o termo correspondente aos dados calculados não contribui para o cálculo da derivada da Lagrangeana:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{m}} &= \frac{\partial}{\partial \mathbf{m}} \left[\frac{1}{2} \int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}|^2 d\mathbf{x} dt - \int_T \int_V \lambda_L(\mathbf{x}, t) \left[\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} - w \right] d\mathbf{x} dt \right] \\ \frac{\partial L}{\partial \mathbf{m}} &= \frac{\partial}{\partial \mathbf{m}} \left[\frac{1}{2} \int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}|^2 d\mathbf{x} dt - \int_T \int_V \lambda_L(\mathbf{x}, t) \left[\nabla^2 \overset{0}{p(\mathbf{x}, t)} - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 \overset{0}{p(\mathbf{x}, t)}}{\partial t^2} - \overset{0}{w} \right] d\mathbf{x} dt \right] \end{aligned}$$

$$\begin{aligned}
\frac{\partial L}{\partial \mathbf{m}} &= \frac{\partial}{\partial \mathbf{m}} \left[- \int_T \int_V \lambda_L(\mathbf{x}, t) \left[- \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} \right] d\mathbf{x} dt \right] \\
\frac{\partial L}{\partial \mathbf{m}} &= - \int_T \int_V \lambda_L(\mathbf{x}, t) \left[- \frac{2}{v_p^3(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} \right] d\mathbf{x} dt \\
\frac{\partial L}{\partial \mathbf{m}} &= \nabla \chi(\mathbf{m}) = - \int_T \int_V \lambda_L(\mathbf{x}, t) \left[\frac{2}{v_p^3(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} \right] d\mathbf{x} dt.
\end{aligned} \tag{2.65}$$

A Equação 2.65 será utilizada para a atualização dos parâmetros do modelo, por um dos métodos de otimização baseado no gradiente da função objetivo (Equações 2.35, 2.36 ou 2.44). O multiplicador de Lagrange λ_L é o campo adjunto $p^\dagger$ (apresentado na próxima seção), e o termo $\frac{\partial^2 p}{\partial t^2}$ é a segunda derivada temporal do campo calculado.

Figura 14 – Representação do método adjunto.



Fonte: Modificado de Bulcão (2004).

Tarantola (1984) apontou as semelhanças entre o algoritmo para o cálculo do gradiente da função objetivo com o método de migração, baseado no princípio de imageamento proposto por Claerbout (1971), da mesma maneira que a migração reversa no tempo por correlação cruzada, o cálculo do gradiente é obtido pela correlação de *lag* zero entre dois campos. O campo direto $\frac{\partial^2 p}{\partial t^2}$, obtido pela modelagem

do campo de pressão e aplicação da derivada segunda em relação ao tempo, e o campo reverso (campo adjunto), obtido através da modelagem reversa no tempo, utilizando as condições finais e a fonte adjunta (FICHTNER et al., 2008).

A Figura 14 ilustra o conceito de aplicação da condição de imagem de Claerbout (1971). As linhas vermelhas indicam a propagação do campo direto e as linhas azuis a propagação reversa no tempo do campo adjunto. Na posição do modelo onde existe um espalhador/difrator, a correlação de *lag* zero entre os campos formará uma imagem, ou no caso do gradiente, uma perturbação no espaço dos modelos.

2.6.2 Derivada da Lagrangeana com relação ao campo direto

A derivada da Lagrangeana com relação ao campo direto permitirá obter o campo adjunto $p^\dagger$, que é utilizado no cálculo do gradiente da função objetivo (Equação 2.65). Uma das condições para minimização do parâmetro $\mathbf{m}$ na Lagrangeana associada (Equação 2.58), apontado por Plessix (2006), é que a derivada da Lagrangeana com relação ao campo p seja zero $\frac{\partial L}{\partial p} = 0$. Assim, aplicando a derivada na Lagrangeana associada:

$$\begin{aligned} \frac{\partial L(p, \lambda_L, \mathbf{m})}{\partial p} &= \frac{\partial}{\partial p} \left[\frac{1}{2} \int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}|^2 d\mathbf{x} dt - \lambda_L \cdot \int_T \int_V \Psi(p, \mathbf{m}) d\mathbf{x} dt \right] = 0 \\ \underbrace{\frac{\partial}{\partial p} \left[\int_T \int_V \lambda_L \left(\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} - f \right) d\mathbf{x} dt \right]}_{\text{Termo O}} &= \underbrace{\frac{\partial}{\partial p} \chi(p)}_{\text{Termo F}}. \end{aligned} \quad (2.66)$$

Para resolver a Equação 2.66 qualquer função objetivo pode ser utilizada, por isso foi substituído o termo referente à norma L2 por uma função objetivo arbitrária $\chi(p)$. A função objetivo depende do campo direto p , pois os dados calculados dependem da equação da onda (algoritmo 2.57).

O termo O dará origem ao operador adjunto e o termo F, à fonte adjunta. Por questões de simplicidade, esses dois termos serão resolvidos separadamente. Em ambos os casos, na aplicação da derivada parcial será utilizada a definição de derivada.

Cálculo do termo O

Aplicando a definição de derivada:

$$\frac{\partial L(p, \mathbf{m}, \lambda_L)}{\partial p} = \lim_{\varepsilon \rightarrow 0} \frac{L(p + \varepsilon \delta p, \mathbf{m}, \lambda_L) - L(p, \mathbf{m}, \lambda_L)}{\varepsilon}, \quad (2.67)$$

e resolvendo apenas o termo O :

$$\begin{aligned} \frac{\partial}{\partial p} \left[\lambda_L \cdot \int_T \int_V \Psi(p, \mathbf{m}) d\mathbf{x} dt \right] &= \frac{\partial}{\partial p} \left[\int_T \int_V \lambda_L \left(\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} - f \right) d\mathbf{x} dt \right] \\ \frac{\partial}{\partial p} \left[\lambda_L \cdot \int_T \int_V \Psi(p, \mathbf{m}) d\mathbf{x} dt \right] &= \int_T \int_V \lim_{\varepsilon \rightarrow 0} \frac{\lambda_L \left[\left(\nabla^2 (p + \varepsilon \delta p) - \frac{1}{v_p^2} \frac{\partial^2 (p + \varepsilon \delta p)}{\partial t^2} \right) - \left(\nabla^2 p - \frac{1}{v_p^2} \frac{\partial^2 p}{\partial t^2} \right) \right]}{\varepsilon} d\mathbf{x} dt \\ \frac{\partial}{\partial p} \left[\lambda_L \cdot \int_T \int_V \Psi(p, \mathbf{m}) d\mathbf{x} dt \right] &= \int_T \int_V \lim_{\varepsilon \rightarrow 0} \frac{\lambda_L \left[\nabla^2 (\varepsilon \delta p) - \frac{1}{v_p^2} \frac{\partial^2 (\varepsilon \delta p)}{\partial t^2} \right]}{\varepsilon} d\mathbf{x} dt \\ \frac{\partial}{\partial p} \left[\lambda_L \cdot \int_T \int_V \Psi(p, \mathbf{m}) d\mathbf{x} dt \right] &= \int_T \int_V \lambda_L \nabla^2 \delta p - \lambda_L \frac{1}{v_p^2} \frac{\partial^2 \delta p}{\partial t^2} d\mathbf{x} dt. \end{aligned} \quad (2.68)$$

A estratégia para resolver esse problema é compreender que δp está presente também no termo F da Equação 2.66. Uma vez isolados, poderão ser fatorados e removidos da equação. Assim, no termo com derivada espacial da Equação 2.68 será usada a seguinte identidade do cálculo vetorial: $\mathbf{u} \cdot \nabla(\mathbf{v}) = \nabla(\mathbf{u} \cdot \mathbf{v}) - \nabla \mathbf{u} \cdot \mathbf{v}$, com $\mathbf{u} \rightarrow \lambda_L$ e $\mathbf{v} \rightarrow \nabla \delta p$. E analisando apenas o termo com derivada espacial:

$$\int_T \int_V \lambda_L \nabla^2 \delta p d\mathbf{x} dt = \int_T \int_V \nabla(\lambda_L \cdot \nabla \delta p) d\mathbf{x} dt - \int_T \int_V \nabla \lambda_L \cdot \nabla \delta p d\mathbf{x} dt. \quad (2.69)$$

Aplicando novamente a identidade $\mathbf{u} \cdot \nabla(\mathbf{v}) = \nabla(\mathbf{u} \cdot \mathbf{v}) - \nabla \mathbf{u} \cdot \mathbf{v}$ no segundo termo da equação anterior, mas utilizando $\mathbf{u} = \nabla \lambda_L$ e $\mathbf{v} = \delta p$, a equação anterior fica:

$$\int_T \int_V \lambda_L \nabla^2 \delta p d\mathbf{x} dt = \int_T \int_V \nabla(\lambda_L \cdot \nabla \delta p) d\mathbf{x} dt - \int_T \int_V \nabla(\nabla \lambda_L \cdot \delta p) d\mathbf{x} dt + \int_T \int_V \nabla(\nabla \lambda_L) \cdot \delta p d\mathbf{x} dt \quad (2.70)$$

Utilizando o teorema de Gauss $\int_V \nabla \mathbf{u} = \int_{\partial D} \mathbf{u} \cdot \vec{n} \cdot dS$, nos dois primeiros termos da equação anterior:

$$\int_T \int_V \lambda_L \nabla^2 \delta p d\mathbf{x} dt = \int_T \int_{\partial D} \lambda_L \cdot \nabla \delta p \cdot \vec{n} \cdot dS dt - \int_T \int_{\partial D} \nabla \lambda_L \cdot \delta p \cdot \vec{n} \cdot dS dt + \int_T \int_V \nabla^2 \lambda_L \cdot \delta p d\mathbf{x} dt, \quad (2.71)$$

Sabendo que a condição de contorno do problema direto é $p(\mathbf{x}, t)|_{\mathbf{x} \in \partial D} = 0$ e impondo a mesma condição de contorno para o campo λ_L :

$$\lambda_L(\mathbf{x}, t)|_{\mathbf{x} \in \partial D} = 0. \quad (2.72)$$

O termo com derivada espacial fica:

$$\begin{aligned} \int_T \int_V \lambda_L \nabla^2 \delta p d\mathbf{x} dt &= \int_T \int_{\partial D} \lambda_L \cdot \nabla \delta p \cdot \vec{n} \cdot dS dt - \int_T \int_{\partial D} \nabla \lambda_L \cdot \delta p \cdot \vec{n} \cdot dS dt + \int_T \int_V \nabla^2 \lambda_L \cdot \delta p d\mathbf{x} dt \\ \int_T \int_V \lambda_L \nabla^2 \delta p d\mathbf{x} dt &= \int_T \int_{\partial D} \lambda_L \cdot \nabla \delta p \cdot \vec{n} \cdot dS dt \xrightarrow{0} - \int_T \int_{\partial D} \nabla \lambda_L \cdot \delta p \cdot \vec{n} \cdot dS dt \xrightarrow{0} + \int_T \int_V \nabla^2 \lambda_L \cdot \delta p d\mathbf{x} dt \\ \int_T \int_V \lambda_L \nabla^2 \delta p d\mathbf{x} dt &= \int_T \int_V \nabla^2 \lambda_L \cdot \delta p d\mathbf{x} dt. \end{aligned} \quad (2.73)$$

Para isolar δp da Equação 2.68, no termo que possui a derivada temporal, será utilizada a seguinte identidade de integração: $\int_T \mathbf{u} \frac{\partial \mathbf{v}}{\partial t} dt = \mathbf{u} \mathbf{v} \Big|_{t=t_0}^{t=t_f} - \int_T \frac{\partial \mathbf{u}}{\partial t} \mathbf{v} dt$, com $\mathbf{u} = \lambda_L$ e $\mathbf{v} = \frac{1}{v_p^2(\mathbf{x})} \frac{\partial \delta p}{\partial t}$:

$$\int_T \int_V \lambda_L \frac{1}{v_p^2} \frac{\partial^2 \delta p}{\partial t^2} d\mathbf{x} dt = \int_V \lambda_L \frac{1}{v_p^2} \frac{\partial \delta p}{\partial t} \Big|_{t=t_0}^{t=t_f} d\mathbf{x} - \int_T \int_V \frac{\partial \lambda_L}{\partial t} \frac{1}{v_p^2} \frac{\partial \delta p}{\partial t} d\mathbf{x} dt \quad (2.74)$$

Aplicando, novamente, a mesma identidade no segundo termo da equação anterior

$$\int_T \mathbf{u} \frac{\partial \mathbf{v}}{\partial t} dt = \mathbf{u} \mathbf{v} \Big|_{t=t_0}^{t=t_f} - \int_T \frac{\partial \mathbf{u}}{\partial t} \mathbf{v} dt, \text{ mas com } \mathbf{u} = \frac{\partial \lambda_L}{\partial t} \text{ e } \mathbf{v} = \frac{1}{v_p^2(\mathbf{x})} \delta p:$$

$$\int_T \int_V \lambda_L \frac{1}{v_p^2} \frac{\partial^2 \delta p}{\partial t^2} d\mathbf{x} dt = \int_V \lambda_L \frac{1}{v_p^2} \frac{\partial \delta p}{\partial t} \Big|_{t=t_0}^{t=t_f} d\mathbf{x} - \int_V \frac{\partial \lambda_L}{\partial t} \frac{1}{v_p^2} \delta p \Big|_{t=t_0}^{t=t_f} d\mathbf{x} + \int_T \int_V \frac{\partial^2 \lambda_L}{\partial t^2} \frac{1}{v_p^2} \delta p d\mathbf{x} dt \quad (2.75)$$

As condições iniciais do problema direto são $p(\mathbf{x}, t = t_0) = \frac{\partial p(\mathbf{x}, t = t_0)}{\partial t} = 0$ e permitem eliminar os termos que dependem das condições iniciais na equação anterior. As condições finais para o campo λ_L são impostas, de modo que:

$$\lambda_L(\mathbf{x}, t = t_f) = \frac{\partial \lambda_L(\mathbf{x}, t = t_f)}{\partial t} = 0. \quad (2.76)$$

Assim, os termos que dependem dos tempos finais nos limites de integração na Equação 2.75 podem ser eliminados:

$$\int_T \int_V \lambda_L \frac{1}{v_p^2} \frac{\partial^2 \delta p}{\partial t^2} d\mathbf{x} dt = \int_T \int_V \frac{1}{v_p^2} \frac{\partial^2 \lambda_L}{\partial t^2} \cdot \delta p d\mathbf{x} dt, \quad (2.77)$$

permitindo reescrever o termo O (Equação 2.68) como:

$$\int_T \int_V \lambda_L \nabla^2 \delta p - \lambda_L \frac{1}{v_p^2} \frac{\partial^2 \delta p}{\partial t^2} d\mathbf{x} dt = \int_T \int_V \left[\nabla^2 \lambda_L - \frac{1}{v_p^2} \frac{\partial^2 \lambda_L}{\partial t^2} \right] \delta p \cdot d\mathbf{x} dt. \quad (2.78)$$

O núcleo da integração na Equação 2.78 é chamado de operador adjunto, pois possui a mesma forma do operador do problema direto. Permitindo determinar o campo λ_L com o mesmo operador de modelagem direta, respeitando as condições finais impostas ao campo λ_L , ou campo adjunto $p^\dagger = \lambda_L$.

De acordo com Fichtner (2011), a maior dificuldade para resolver o campo adjunto é a imposição, que no instante de tempo final, o campo adjunto seja nulo $p^\dagger(\mathbf{x}, t = t_f) = 0$ (Equação 2.76). Essa dificuldade é superada utilizando o mesmo operador de modelagem direta para obter o campo adjunto, fazendo as condições finais atuarem como condições iniciais e resolvendo reversamente no tempo.

2.6.3 A fonte adjunta

O termo F da Equação 2.66 é conhecido como fonte adjunta, que é obtido através da aplicação da derivada parcial da Lagrangeana associada L com relação ao campo p igual a zero. O termo correspondente à fonte adjunta é afetado apenas pela escolha da função objetivo χ .

A escolha da norma a ser utilizada no método adjunto implica apenas no tipo de fonte adjunta que será utilizada em conjunto com o operador adjunto, para determinar o campo adjunto $p^\dagger$. A aplicação da definição da derivada ao termo da fonte adjunta permitirá determinar a sua forma.

Norma L2

A maioria das aplicações utilizam a norma L2 (Equação 2.4) como função objetivo, o que significa resolver o problema por mínimos quadrados. É importante ressaltar que os dados calculados são obtidos através do campo direto p . Portanto, é possível escrever:

$$\frac{\partial \chi(p)}{\partial p} = \frac{\partial \chi_2}{\partial p}$$

$$\frac{\partial \chi(p)}{\partial p} = \frac{\partial}{\partial p} \left[\frac{1}{2} \int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}|^2 d\mathbf{x} dt \right]$$

$$\frac{\partial \chi(p)}{\partial p} = \frac{\partial}{\partial p} \left[\frac{1}{2} \int_T \int_V |p(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_R) - \mathbf{d}_{obs}|^2 d\mathbf{x} dt \right]$$

$$\frac{\partial \chi(p)}{\partial p} = \frac{1}{2} \int_T \int_V \lim_{\varepsilon \rightarrow 0} \frac{|(p + \varepsilon \delta p) \delta(\mathbf{x} - \mathbf{x}_R) - \mathbf{d}_{obs}|^2 - |p \delta(\mathbf{x} - \mathbf{x}_R) - \mathbf{d}_{obs}|^2}{\varepsilon} d\mathbf{x} dt$$

$$\frac{\partial \chi(p)}{\partial p} = \frac{1}{2} \int_T \int_V 2(p \delta(\mathbf{x} - \mathbf{x}_R) - \mathbf{d}_{obs}) \delta p \cdot d\mathbf{x} dt$$

$$\frac{\partial \chi(p)}{\partial p} = \int_T \int_V (\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}) \delta p \cdot d\mathbf{x} dt$$

$$\frac{\partial \chi(p)}{\partial p} = \int_T \int_V \Delta \mathbf{d} \delta p \cdot d\mathbf{x} dt. \quad (2.79)$$

Assim, a fonte adjunta para o caso da norma L2 é obtida através do sismograma residual $\Delta \mathbf{d}$.

Norma L1

A fonte adjunta utilizando a norma L1 como função objetivo pode ser obtida pela mesma estratégia. Aplicando a derivada da função objetivo definida pela norma L1

(Equação 2.6) com relação ao campo p :

$$\begin{aligned}
 \frac{\partial \chi(p)}{\partial p} &= \frac{\partial \chi_1}{\partial p} \\
 \frac{\partial \chi(p)}{\partial p} &= \frac{\partial}{\partial p} \left[\int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}| d\mathbf{x} dt \right] \\
 \frac{\partial \chi(p)}{\partial p} &= \frac{\partial}{\partial p} \left[\int_T \int_V |p(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_R) - \mathbf{d}_{obs}| d\mathbf{x} dt \right] \\
 \frac{\partial \chi(p)}{\partial p} &= \int_T \int_V \lim_{\varepsilon \rightarrow 0} \frac{|(p + \varepsilon \delta p) \delta(\mathbf{x} - \mathbf{x}_R) - d_{obs}| - |p \delta(\mathbf{x} - \mathbf{x}_R) - d_{obs}|}{\varepsilon} d\mathbf{x} dt.
 \end{aligned} \tag{2.80}$$

Do cálculo diferencial, a derivada do valor absoluto de uma função qualquer $f(x) = |x|$ é dada por:

$$\frac{df(x)}{dx} = \lim_{\varepsilon \rightarrow 0} \frac{|x + \varepsilon \delta x| - |x|}{\varepsilon} = \frac{x}{|x|} \delta x. \tag{2.81}$$

Dessa maneira, a fonte adjunta para a norma L1 é obtida pelo sinal do sismograma residual¹ (DIKPESE; TARANTOLA, 1999; PYUN; SON; SHIN, 2009; SCHUSTER, 2017) :

$$\begin{aligned}
 \frac{\partial \chi(p)}{\partial p} &= \int_T \int_V \frac{p(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_R) - d_{obs}}{|p(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_R) - d_{obs}|} \delta p d\mathbf{x} dt \\
 \frac{\partial \chi(p)}{\partial p} &= \int_T \int_V \frac{\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}}{|\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}|} \delta p d\mathbf{x} dt \\
 \frac{\partial \chi(p)}{\partial p} &= \int_T \int_V \frac{\Delta \mathbf{d}}{|\Delta \mathbf{d}|} \delta p d\mathbf{x} dt
 \end{aligned} \tag{2.82}$$

2.6.4 Campo adjunto

Reescrevendo a derivada da Lagrangeana com relação ao campo p (Equação 2.66) utilizando os resultados da Equação 2.78 e a fonte adjunta por norma L2 (Equação

¹ Na prática é incluído um termo de estabilização no denominador para evitar divisões por zero.

2.79), por exemplo, encontra-se:

$$\underbrace{\frac{\partial}{\partial p} \left[\int_T \int_V p^\dagger \left(\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} - f \right) d\mathbf{x} dt \right]}_{\text{Termo O}} = \underbrace{\frac{\partial}{\partial p} \chi(p)}_{\text{Termo F}}$$

$$\int_T \int_V \left[\nabla^2 p^\dagger - \frac{1}{v_p^2} \frac{\partial^2 p^\dagger}{\partial t^2} \right] \delta p \cdot d\mathbf{x} dt = \int_T \int_V \Delta \mathbf{d} \delta p \cdot d\mathbf{x} dt \quad (2.83)$$

$$\nabla^2 p^\dagger - \frac{1}{v_p^2} \frac{\partial^2 p^\dagger}{\partial t^2} = \Delta \mathbf{d},$$

e utilizando a norma L1 (Equação 2.82):

$$\nabla^2 p^\dagger - \frac{1}{v_p^2} \frac{\partial^2 p^\dagger}{\partial t^2} = \frac{\Delta \mathbf{d}}{|\Delta \mathbf{d}|}. \quad (2.84)$$

Assim, é possível escrever de forma mais geral que o campo adjunto pode ser obtido através da equação:

$$\nabla^2 p^\dagger(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p^\dagger(\mathbf{x}, t)}{\partial t^2} = \frac{\partial \chi}{\partial p}, \quad (2.85)$$

onde $p^\dagger(\mathbf{x}, t)$ é obtido pelo mesmo operador de modelagem direta, utilizando como termo fonte, a fonte adjunta $\frac{\partial \chi}{\partial p}$, determinada pela escolha da função objetivo. Devido às condições finais impostas para $p^\dagger$, $p^\dagger(\mathbf{x}, t = t_f) = \frac{\partial p^\dagger(\mathbf{x}, t = t_f)}{\partial t} = 0$, o operador adjunto deve ser calculado reversamente no tempo.

2.6.5 Derivada da Lagrangeana com relação ao campo adjunto

A minimização da função objetivo pelo método dos multiplicadores de Lagrange impõe que o campo direto p e o campo adjunto $p^\dagger$ (multiplicador de Lagrange λ_L) sejam pontos de sela. Assim, calculando a derivada da Lagrangeana com relação ao campo adjunto e igualando a zero:

$$\frac{\partial L}{\partial p^\dagger} = 0$$

$$\frac{\partial L}{\partial p^\dagger} = \frac{\partial}{\partial p^\dagger} \left[\chi(p) - \int_T \int_V p^\dagger \left[\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} - f \right] d\mathbf{x} dt \right] = 0$$

$$\frac{\partial L}{\partial p^\dagger} = \frac{\partial}{\partial p^\dagger} [\chi(p, \mathbf{m})] - \frac{\partial}{\partial p^\dagger} \left[\int_T \int_V p^\dagger \left[\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} - f \right] d\mathbf{x} dt \right] = 0$$

$$\frac{\partial L}{\partial p^\dagger} = \int_T \int_V \delta p^\dagger \left[\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} - f \right] d\mathbf{x} dt = 0$$

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} = w, \quad (2.86)$$

é obtida a equação de estado que coincide com o problema vinculado, a equação da onda acústica.

2.6.6 Reparametrização do problema direto e o cálculo de $\frac{\partial L}{\partial \mathbf{m}}$

A escolha dos parâmetros utilizados na solução do problema inverso (Equação 2.3) pode afetar drasticamente as propriedades de convergência do método de otimização, mesmo que os parâmetros sejam matematicamente equivalentes (TARANTOLA, 1986). Carneiro et al. (2018) mostraram que se o método de otimização utilizado para resolver o problema inverso (equação de atualização 2.35) for parametrizado pelo quadrado da vagarosidade:

$$\mathbf{m} = s(\mathbf{x}) = \frac{1}{v_p(\mathbf{x})^2}, \quad (2.87)$$

ao invés da velocidade de propagação $\mathbf{m} = v_p$, garante direções de atualização melhor escaladas na equação de atualização 2.35.

A equação da onda acústica (Equação A.56) reescrita em termos do quadrado da vagarosidade s , fica:

$$\nabla^2 p(\mathbf{x}, t) - s(\mathbf{x}) \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) = w(\mathbf{x}, t). \quad (2.88)$$

O cálculo da direção de atualização (Equação 2.63) do problema de otimização será dado por:

$$\frac{\partial L}{\partial \mathbf{m}} = \frac{\partial}{\partial \mathbf{m}} \left[- \int_T \int_V p^\dagger(\mathbf{x}, t) \left[\cancel{\nabla^2 p(\mathbf{x}, t)}^0 - s(\mathbf{x}) \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} - \cancel{w}^0 \right] d\mathbf{x} dt \right] \quad (2.89)$$

$$\frac{\partial L}{\partial \mathbf{m}} = \nabla_{\chi_s}(\mathbf{m}) = \int_T \int_V p^\dagger(\mathbf{x}, t) \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} d\mathbf{x} dt,$$

reescrevendo a equação de atualização, utilizando a vagarosidade ao quadrado $s(\mathbf{x})$:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \nabla_{\mathbf{m}} \chi(\mathbf{m}) \quad (2.90)$$

$$s(\mathbf{x})_{k+1} = s(\mathbf{x})_k - \alpha_k \nabla_{s\chi}(\mathbf{m}).$$

Agora, utilizando a definição da vagarosidade ao quadrado (Equação 2.87):

$$\begin{aligned} \frac{1}{v_p(\mathbf{x})_{k+1}^2} &= \frac{1}{v_p(\mathbf{x})_k^2} - \alpha_k \nabla_{s\chi}(v_p) \\ \frac{1}{v_p(\mathbf{x})_{k+1}^2} &= \frac{1 - v_p(\mathbf{x})_k^2 \alpha_k \nabla_{s\chi}(\mathbf{m})}{v_p(\mathbf{x})_k^2} \\ v_p(\mathbf{x})_{k+1}^2 &= \frac{v_p(\mathbf{x})_k^2}{1 - v_p(\mathbf{x})_k^2 \alpha_k \nabla_{s\chi}(\mathbf{m})} \\ v_p(\mathbf{x})_{k+1} &= \frac{v_p(\mathbf{x})_k}{\sqrt{1 - v_p(\mathbf{x})_k^2 \alpha_k \nabla_{s\chi}(\mathbf{m})}} \end{aligned} \quad (2.91)$$

Para pequenas perturbações é possível aproximar essa função por uma expansão

em série de Taylor até a primeira ordem:

$$\frac{1}{\sqrt{1+x}} \approx 1 - \frac{x}{2}, \quad (2.92)$$

desse modo, a Equação 2.91 fica:

$$v_p(\mathbf{x})_{k+1} \approx v_p(\mathbf{x})_{k+1} = v_p(\mathbf{x}) - \frac{v_p(\mathbf{x})_k^3}{2} \alpha_k \nabla_s \chi(\mathbf{m}). \quad (2.93)$$

Portanto, a escolha do parâmetro do modelo $\mathbf{m}$ define a direção de atualização no espaço do modelo. Quando o problema direto é parametrizado pela velocidade de propagação $\mathbf{m} = v_p$, o gradiente da função objetivo $\nabla_{v_p} \chi(\mathbf{m})$ é dado pela Equação 2.65. Porém, se o problema for parametrizado pelo quadrado da vagarosidade o gradiente da função objetivo pode ser escrito como a seguir, mesmo se o problema direto for resolvido utilizando a velocidade:

$$\frac{\partial L}{\partial \mathbf{m}} = \nabla \chi(\mathbf{m}) = - \int_T \int_V p^\dagger(\mathbf{x}, t) \left[\frac{v_p^3(\mathbf{x})}{2} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} \right] d\mathbf{x} dt. \quad (2.94)$$

2.6.7 A Lagrangeana regularizada por *Total Variation*

A regularização do problema inverso introduz informações a priori, para restringir o espaço de soluções. Portanto, a regularização é aplicada adicionando um termo regularizador $R(\mathbf{m})$, multiplicado pelo parâmetro de regularização λ_R , a função objetivo escolhida $\chi(\mathbf{m})$ (Equação 2.16). Assim, dentro do contexto do método do estado adjunto, a regularização afeta o cálculo da direção de atualização dos modelos, através do cálculo do gradiente da Lagrangeana $\frac{\partial L}{\partial \mathbf{m}}$. Escrevendo a Lagrangeana associada (Equação 2.58), utilizando a norma L2 (Equação 2.4) como função objetivo e o termo de regularização *Total Variation* (Equação 2.20):

$$L(p, \lambda_L, \mathbf{m}) = \chi_2(\mathbf{m}) + \lambda_R R_{TV}(\mathbf{m}) - \lambda_L \cdot \int_T \int_V \Psi(p, \mathbf{m}) d\mathbf{x} dt$$

$$L(p, \lambda_L, \mathbf{m}) = \frac{1}{2} \int_T \int_V |\mathbf{d}_{cal}(\mathbf{m}; \mathbf{x}, t) - \mathbf{d}_{obs}|^2 d\mathbf{x}dt + \lambda_R \int_V \sqrt{(\nabla_{\mathbf{x}} \mathbf{m}(\mathbf{x}))^2} d\mathbf{x} - \lambda_L \int_T \int_V \Psi(p, \mathbf{m}) d\mathbf{x}dt \quad (2.95)$$

É importante destacar que a escolha do parâmetro de regularização λ_R controla a influência da regularização na solução do problema inverso. Escolhas de valores altos para λ_R tornarão a função objetivo menos relacionada com o problema original que utiliza a comparação entre dados calculados e dados observados. Por outro lado, a escolha de valores baixos para λ_R diminuirá o efeito do termo de regularização, tornando a função objetivo apenas dependente dos dados (SOUZA, 2018).

O cálculo da derivada da Lagrangeana regularizada (Equação 2.95), com relação aos parâmetros do modelo, é muito semelhante ao cálculo da derivada anterior (Equação 2.65), com exceção do termo de regularização. A aplicação da definição da derivada (Equação 2.64) ao termo regularizador fica (ANAGAW; SACCHI, 2011):

$$\begin{aligned} \frac{\partial R(\mathbf{m})}{\partial \mathbf{m}} &= \frac{\partial}{\partial \mathbf{m}} \int_V \sqrt{(\nabla_{\mathbf{x}} \mathbf{m}(\mathbf{x}))^2} d\mathbf{x} \\ \frac{\partial R(\mathbf{m})}{\partial \mathbf{m}} &= \int_V \frac{\partial}{\partial \mathbf{m}} |\nabla_{\mathbf{x}} \mathbf{m}(\mathbf{x})| d\mathbf{x}, \end{aligned} \quad (2.96)$$

Utilizando, novamente, a definição da derivada do valor absoluto de uma função (Equação 2.81) é possível encontrar a componente da direção de atualização, responsável pela regularização:

$$\frac{\partial R(\mathbf{m})}{\partial \mathbf{m}} = \int_V \frac{\nabla_{\mathbf{x}} \mathbf{m}}{|\nabla_{\mathbf{x}} \mathbf{m}|} d\mathbf{x}. \quad (2.97)$$

Assim, a direção de atualização regularizada por *Total Variation* é dada por:

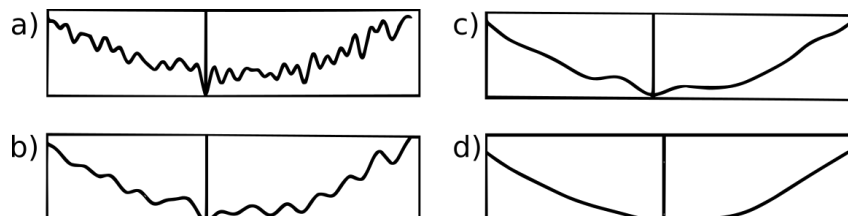
$$\frac{\partial L}{\partial \mathbf{m}} = \nabla \chi(\mathbf{m}) = - \left(\int_T \int_V p^\dagger \left[\frac{2}{v_p^3(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} \right] d\mathbf{x}dt + \lambda_R \int_V \frac{\nabla_{\mathbf{x}} \mathbf{m}}{|\nabla_{\mathbf{x}} \mathbf{m}|} d\mathbf{x} \right), \quad (2.98)$$

onde o primeiro termo é o gradiente da função objetivo sem regularização, e o segundo termo é a derivada do termo de regularização $R(\mathbf{x})$, com relação aos parâmetros do modelo $\mathbf{m}$.

2.7 O problema de salto de ciclo e a abordagem multiescala

Devido à natureza oscilatória dos dados sísmicos, o resíduo, que é utilizado no cálculo da função objetivo, também tende a possuir um comportamento oscilatório. Logo, a função objetivo possuirá comportamento altamente oscilatório, como consequência da não-linearidade dos dados sísmicos (MATIAS, 2016). Dessa forma, é natural esperar que a função objetivo possua uma grande quantidade de mínimos locais (Figura 15a).

Figura 15 – Representação da função objetivo e seu comportamento oscilatório para escalas diferentes. (a)-(d) ilustram a mesma função objetivo com escalas crescentes. A função objetivo possui menos mínimos locais quanto maior for a escala.



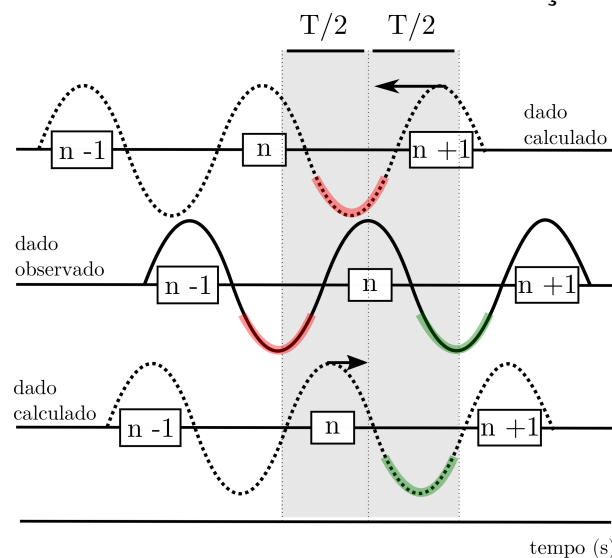
Fonte: Modificado de Bunks et al. (1995).

Entretanto, mesmo a escolha de um modelo de velocidades inicial próximo ao mínimo global não garantirá a convergência da função objetivo, para um modelo de propriedades com sentido geológico. Uma das causas desse problema pode ser explicada pelo salto de ciclo (*cycle skipping*).

Segundo Virieux e Operto (2009), esse problema ocorre quando existe uma diferença de fase maior que meio período do comprimento de onda (Figura 16), entre o dado calculado e observado, fazendo com que o método de otimização ajuste o dado calculado a um ciclo do traço sísmico observado com defasagem de um ou mais comprimentos de onda. Como consequência, o acúmulo dos ajustes fora de fase, entre dado calculado e observado, pode inserir no modelo de propriedades estruturas

espúrias sem sentido geológico ou, no pior dos casos, estruturas que fazem sentido geologicamente, mas não condizem com a realidade.

Figura 16 – Esquema que representa o problema de salto de ciclo. Quando existe uma diferença de fase maior que meio período do comprimento de onda, os ciclos do sinal são ajustados de forma errada (indicação em vermelho) pelo algoritmo de otimização. Isso provoca uma diminuição da função objetivo porém estima o modelo utilizando uma informação equivocada.



Fonte: Modificado de Virieux e Operto (2009)

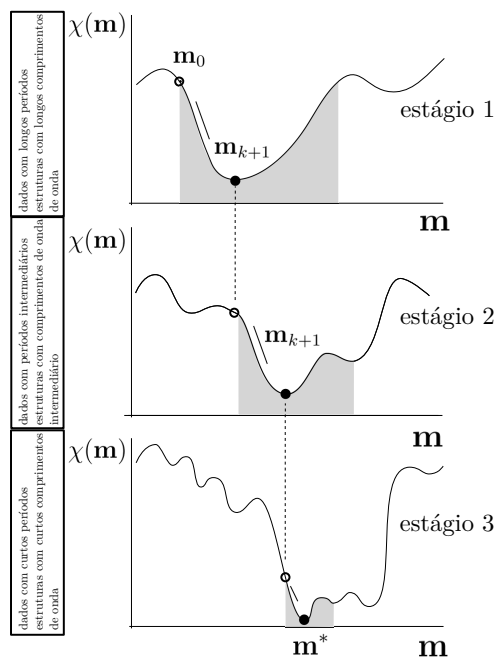
Para contornar o problema de salto de ciclo, Bunks et al. (1995) propuseram a abordagem multiescala. Essa abordagem consiste em decompor o problema em diferentes escalas, de modo que nas maiores escalas a função objetivo possua um comportamento menos oscilatório (Figura 15d), permitindo que o algoritmo de otimização local se aproxime do mínimo global da função objetivo mais consistentemente.

A abordagem multiescala pode ser aplicada em diversos domínios: na frequência, no afastamento fonte-receptor, em uma janela de tempo do dado observado, entre outras. Devido à natureza dos dados sísmicos, a abordagem multiescala mais utilizada é aplicada no domínio da frequência.

Nessa abordagem são otimizadas as baixas frequências (maiores escalas) no início e progressivamente são otimizadas as frequências mais altas, até uma determinada frequência de corte. Nas baixas frequências espera-se que o dado sísmico possua menos oscilações, assim o dado terá um menor conteúdo de frequência, porém, a função objetivo terá um comportamento com menos oscilações

(Figura 15d). Isso contribui para que o método de otimização convirja para um valor mais próximo do mínimo global. Com o aumento progressivo do conteúdo de frequência, a função objetivo tende a convergir para o mínimo global, evitando o problema do salto de ciclo (Figura 17).

Figura 17 – Representação do comportamento da função objetivo para as diferentes escalas.



Fonte: Modificado de Fichtner (2011).

Assim, quando utilizada a equação da onda no domínio do tempo, para resolver o problema direto e a abordagem multiescala do domínio da frequência, é necessário decompor o dado nas frequências de corte escolhidas. Por exemplo, se o dado observado contém frequências de até 30 Hz, esses dados precisam ser filtrados utilizando a primeira frequência de corte escolhida para a abordagem multiescala, e calcular os dados utilizando a mesma frequência de corte inicial. O algoritmo de otimização é aplicado para encontrar o mínimo da função objetivo utilizando esse conteúdo de frequência. Após a convergência da função objetivo, os dados observados são filtrados novamente com uma frequência de corte maior. E os dados calculados são gerados com o novo conteúdo de frequência, e o algoritmo de otimização é aplicado novamente. Esse processo é repetido até alcançar a máxima frequência de corte escolhida (Figura 17).

2.8 Estimativa da *wavelet*

A estimativa da *wavelet* é uma importante etapa no método FWI. Sua estimativa apropriada permite gerar dados calculados que alcançam um melhor ajuste aos dados observados, resultando em melhores propriedades de convergência do método de otimização utilizado para resolver o problema de inversão não-linear. Na FWI no domínio do tempo, a *wavelet* estimada pode ser decomposta em faixas de frequência para a aplicação da abordagem multiescala (BUNKS et al., 1995; PRATT, 1999; VIRIEUX; OPERTO, 2009; BROSSIER; OPERTO; VIRIEUX, 2009).

Existem várias formas de estimar a *wavelet* fonte a partir dos dados observados (SEARS; BARTON; SINGH, 2010). A maioria dos métodos utiliza o modelo convolucional para extrair a assinatura da fonte, a partir das seções sísmicas. Esses métodos precisam de uma série de refletividade, que normalmente é obtida utilizando informações de perfis de poços. Quando não existe nenhuma informação de poço, a série de refletividade é assumida como uma variável aleatória não correlacionada (WANG, 2016).

A *wavelet* fonte também pode ser estimada a partir de dados pré-empilhados. Schuster (2017) apresentou diversos métodos para estimativa da fonte. Entre eles estão o empilhamento das reflexões do fundo do mar utilizando apenas os afastamentos fonte-receptor mais curtos e a seleção e integração da onda direta. Por outro lado, Pratt (1999) realizou a estimativa da assinatura da fonte no domínio da frequência, obtendo a solução exata para estimativa da fonte dentro do contexto de mínimos quadrados, dada por:

$$w(\omega) = \frac{D_{cal}(\omega)^T D_{obs}(\omega)}{D_{cal}(\omega)^T D_{cal}(\omega)^*} \quad (2.99)$$

onde D_{cal} e D_{obs} são as transformadas de Fourier do dado calculado e dado observado. Os índices T e $*$ representam os operadores transposto e complexo conjugado.

Entretanto, a estimativa da *wavelet* fonte nesse trabalho foi obtida resolvendo o problema de minimização (Equação 2.3) onde a *wavelet* fonte $w(t)$ é o parâmetro a ser encontrado pelo método de otimização. Esse problema foi resolvido utilizando o método gradiente descendente (Equação 2.35). O gradiente da função objetivo, que

fornece a direção de atualização no espaço dos modelos, é calculado pelo método adjunto (Equação 2.63) utilizando a norma L2 (Equação 2.4). Nos testes realizados, a comparação entre as *wavelets* estimadas pela equação 2.99 e pelo problema de minimização forneceram resultados similares.

Para a aplicação dessa abordagem, é selecionado somente o traço mais próximo da fonte de cada sismograma observado e a primeira chegada. Por questões de simplicidade e custo computacional, o traço escolhido é comparado ao dado calculado através da equação da onda 1D, também discretizada por diferenças finitas. A velocidade da água (1500 m/s) foi usada como modelo de velocidades para simular a primeira chegada.

A função objetivo é calculada e usada para obter o gradiente da função objetivo. A *wavelet* estimada é filtrada nas mesmas faixas de frequência que o dado observado, e aplicada ao processo de FWI, que estima o modelo de velocidade a partir de um modelo de velocidade inicial. Essa abordagem permite a estimativa da *wavelet* no contexto da FWI e adiciona-lá ao fluxo de trabalho de inversão do modelo de velocidade. Além disso, torna o experimento numérico mais realista.

2.8.1 Método adjunto para estimativa da *wavelet*

De acordo com Schuster (2017), a inversão da *wavelet* fonte pode ser vista como uma FWI unidimensional, que pode ser resolvida pelo método gradiente descendente onde, a direção de atualização é encontrada através do método adjunto.

Para a estimativa da *wavelet* fonte, foi usada a equação da onda acústica 1D (Equação A.56, com $x = z$) para solução do problema direto. A *wavelet* w é escolhida como parâmetro a ser otimizado, dessa maneira, a função objetivo utilizando a norma L2 pode ser definida como:

$$\chi(w) = \frac{1}{2} \int_T \|\mathbf{d}_{obs}(z_r, t) - \mathbf{d}_{cal}(z_r, t)\|^2 dt, \quad (2.100)$$

onde os dados observados $\mathbf{d}_{obs}$, gerados sinteticamente com a equação da onda acústica 2D, foram comparados aos dados calculados $\mathbf{d}_{cal}$ gerados pela equação da onda 1D. A posição onde o campo é registrado é dado por z_r .

O método dos multiplicadores de Lagrange (Seção 2.6) é utilizado para minimizar a função objetivo (Equação 2.100) e encontrar seu mínimo vinculado ao problema direto. Se o problema direto for escrito como $\mathcal{L}[w].p = w(t)$, a Lagrangeana sem restrição é dada por:

$$L(p, p^\dagger, w) = \frac{1}{2} \int_T \|\mathbf{d}_{obs}(z_r, t) - \mathbf{d}_{cal}(z_r, t)\|^2 dt - \int_T p^\dagger (\mathcal{L}[w].p - w(t)) dt, \quad (2.101)$$

onde $p^\dagger$ é o multiplicador de Lagrange ou campo adjunto.

O gradiente da Lagrangeana, considerando a *wavelet* fonte como parâmetro a ser invertido é dado por:

$$\nabla L(p, p^\dagger, w) = \left(\frac{\partial L}{\partial p}, \frac{\partial L}{\partial p^\dagger}, \frac{\partial L}{\partial w} \right). \quad (2.102)$$

Como o mínimo da Lagrangeana é um ponto de sela, é possível impor que:

$$\frac{\partial L}{\partial p^\dagger} = 0, \quad (2.103)$$

e é encontrada a equação de estado:

$$\frac{1}{v_p^2(z)} \frac{\partial^2}{\partial t^2} p(z, t) - \nabla^2 p(z, t) = w(z, t). \quad (2.104)$$

E impondo que:

$$\frac{\partial L}{\partial p} = 0, \quad (2.105)$$

é encontrado a equação adjunta:

$$\frac{1}{v_p^2(z)} \frac{\partial^2}{\partial t^2} p^\dagger(z, t) - \nabla^2 p^\dagger(z, t) = \Delta d. \quad (2.106)$$

que permite determinar $p^\dagger$ em todas as posições e instantes de tempo.

Por fim, a direção de atualização que minimiza a função objetivo é encontrada derivando a Lagrangeana associada em relação a w :

$$\nabla\chi(w) = \nabla L(p, p^\dagger, w) = \frac{\partial L}{\partial w} = p^\dagger(z_r, t). \quad (2.107)$$

O gradiente da função objetivo (Equação 2.107) é usado em conjunto com o método gradiente descendente (Equação 2.35) para a estimativa da *wavelet*:

$$w_{k+1} = w_k - \alpha_k \nabla\chi(w), \quad (2.108)$$

onde w_{k+1} é a *wavelet* estimada, w_k é a *wavelet* atual (a *wavelet* inicial w_0 é um parâmetro de entrada do algoritmo de otimização), α_k é o tamanho do passo em cada iteração e $\nabla\chi(w)$ é a perturbação na *wavelet* em cada iteração.

Parte II

Metodologia, Resultados e Discussões

3 Metodologia

Nos experimentos numéricos realizados, todos os dados observados foram obtidos sinteticamente, ou seja, o mesmo operador que gera os dados calculados foi usado para a obtenção dos dados observados, com a diferença do modelo de velocidade utilizado. No caso dos dados observados, foi utilizado o modelo de referência, ou modelo verdadeiro, e no caso dos dados calculados, os modelos iniciais ou estimados. Esse processo ficou conhecido como “crime de inversão” (WIRGIN, 2004). O objetivo desse experimento pouco realista visa evitar os diversos problemas que surgem ao se trabalhar com dados reais, como por exemplo, o pré condicionamento dos dados para a aplicação da FWI. É importante ressaltar que a FWI acústica possui um problema fundamental em sua concepção, como apontou Weglein (2013), pois usa os dados errados, a equação errada e o modelo da Terra errado, uma vez que a aproximação acústica ignora diversos fenômenos ondulatórios.

Nos testes apresentados na seção 4 foram avaliados os métodos de otimização gradiente descendente, gradiente conjugado e quasi-Newton LBFGS. Nesses experimentos numéricos, para cada método de otimização foram comparados os resultados de inversão usando a função objetivo definida como a norma L2 (Equação 2.4) e também a função objetivo definida como uma norma L1 (Equação 2.6). E ainda, foram feitos testes incluindo a regularização por *total variation*.

Nessa seção são apresentados os modelos de velocidades utilizados nos testes numéricos. Também é abordada a preparação dos modelos iniciais e como os dados são registrados.

Além disso, alguns detalhes sobre a linguagem de programação FORTRAN utilizada no desenvolvimento dos algoritmos de modelagem e inversão, e sua integração com a biblioteca de paralelização MPI e a biblioteca de otimização SEISCOPE. Além disso, é apresentada a interface gráfica desenvolvida em Matlab para controle de qualidade dos resultados das modelagens e inversões.

3.1 Preparação do modelo inicial

O modelo inicial deve estar próximo ao modelo verdadeiro para garantir a convergência do método de otimização baseado no gradiente. Os modelos iniciais utilizados no processo de inversão simulam os modelos de velocidade construídos durante a etapa do processamento ou construídos através de uma inversão tomográfica (WANG et al., 2019; GEISLINGER et al., 2019; WELLINGTON et al., 2015).

Para simular tais modelos, o modelo de referência foi suavizado usando a biblioteca *Seismic Unix* (STOCKWELL, 2002). Para garantir a preservação dos tempos de trânsito dos dados sísmicos modelados, foi suavizada a vagariedade ($\frac{1}{v_p}$) do modelo de referência e, em seguida, convertida em velocidade.

A diferença relativa entre o valor quadrático médio (RMS - *Root mean square*) do modelo verdadeiro, o modelo inicial e estimado:

$$\Delta v_{rms} \% = \frac{v_{rms}^{ref} - v_{rms}}{v_{rms}^{ref} - v_{rms}^0} \times 100, \quad (3.1)$$

foi utilizado como métrica da qualidade do modelo. Ou seja, se essa diferença for igual a zero, o modelo analisado é idêntico ao modelo de referência.

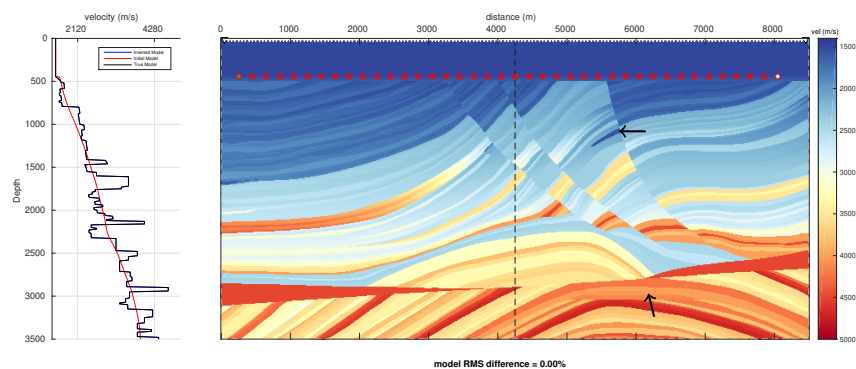
3.2 Modelo Marmousi

O modelo Marmousi (Figura 18) utilizado nos experimento é um trecho do modelo Marmousi II, desenvolvido por Versteeg (1994) e atualizado por Martin, Wiley e Marfurt (2006). O modelo Marmousi foi criado em 1988 pelo Instituto Francês de Petróleo (IFP), com a finalidade de testar algoritmos de imageamento em estruturas geológicas complexas.

O modelo Marmousi possui fortes variações laterais de velocidade, além disso, apresenta falhas e anomalias de alta velocidade, e também, anomalias de baixa velocidade, situações geológicas que dificultam a construção do modelo de velocidade pelo métodos convencionais, como por exemplo a tomografia sísmica. As setas indicam

duas anomalias no modelo de velocidade que podem ser interpretadas como rochas reservatórios.

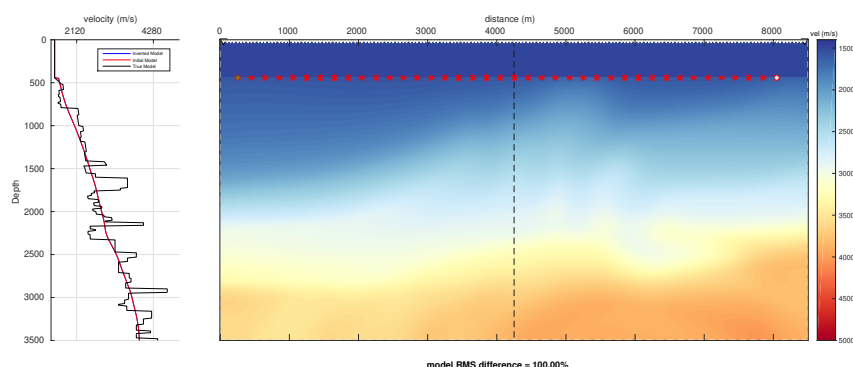
Figura 18 – Modelo de referência Marmousi. Número de amostras $N_x = 851$ e $N_z = 351$ com espaçamento entre os pontos da malha de 10 m. A linha pontilhada indica a posição do perfil apresentado na lateral. As setas indicam anomalias de baixa velocidade.



O modelo Marmousi II possui originalmente 13601 amostras horizontais e 2801 amostras verticais, com espaçamento entre os pontos da malha de 1,25 m. Nos testes realizados nesse trabalho foi selecionada a região de maior complexidade geológica no modelo Marmousi. O modelo foi re-amostrado para 851 amostras na direção horizontal e 351 amostras na direção vertical. O espaçamento de malha utilizado foi de 10 m.

A Figura 19 apresenta o modelo de velocidade inicial, baseado no modelo Marmousi. Esse modelo de velocidade foi obtido através da suavização do modelo verdadeiro (Figura 18). Foi utilizado o programa *smooth2* do pacote *Seismic Unix* com parâmetros $r_1 = r_2 = 50$. Esse modelo é utilizado como modelo inicial no processo de inversão.

Figura 19 – Modelo inicial criado a partir da suavização do modelo de referência Marmousi (Figura 18).



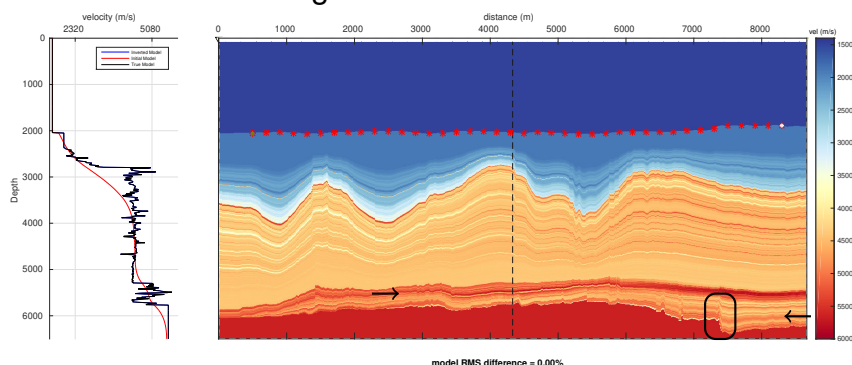
3.3 Modelo baseado no Campo de Búzios

Como os resultados do método de inversão fornecem respostas distintas de acordo com a região de estudo, é importante aplicá-los direcionados a área de interesse. Assim, a construção de modelos conceituais que representam as principais dificuldades encontradas na região de estudo, como por exemplo, as espessas camadas de sal estratificado, grandes variações laterais de velocidades, falhas geológicas, entre outros, são importantes para a validação dos algoritmos de inversão.

Nesse trabalho, foi realizada a construção de um modelo de velocidade conceitual, utilizando o contexto do pré-sal brasileiro na região do campo de Búzios na Bacia de Santos. A Figura 20 apresenta o modelo conceitual construído para o campo de Búzios.

A seta a esquerda indica a as rochas reservatórios no pré-sal e a seta a direita indica a base dos corpos salinos. Uma falha geológica foi simulada no modelo de velocidade, indicada pelo retângulo.

Figura 20 – Modelo de velocidade conceitual baseado nos dados pós empilhamento do campo de Búzios. Número de amostras $N_x = 865$ e $N_z = 651$ com espaçamento entre os pontos da malha de 10 m. A linha pontilhada indica a posição do perfil lateral. As setas indicam a base do sal e as rochas reservatórios. O retângulo indica uma falha simulada artificialmente.



3.3.1 Contexto geológico

A bacia de Santos vem se consolidando no cenário petrolífero nacional, através da constante oferta de blocos exploratórios e de investimentos de pesquisa, por parte de companhias nacionais e estrangeiras. Tradicionalmente, tais investimentos vêm sendo aplicados em fronteiras exploratórias em águas profundas e ultraprofundas, buscando reservatórios em arenitos do Cretáceo Superior.

Recentemente, a seção pré-sal (indicada pelas setas) vem ganhando destaque, devido às descobertas nos reservatórios carbonáticos. O Campo de Búzios está localizado na porção central da Bacia de Santos, a cerca de 180 km da costa do município do Rio de Janeiro, em lâmina d'água de aproximadamente 1900 m de profundidade.

Os reservatórios do Campo de Búzios se situam entre 5.000 e 6.0000 m de profundidade, e são representados pelas coquinas da Fm. Itapema e os microbialitos da Fm. Barra Velha, os quais estão sotopostos a espessa camada evaporítica referente a Fm. Ariri, composta majoritariamente por Halita, tendo uma espessura variando da ordem de centenas de metros até mais de 2 km (ANP, 2016).

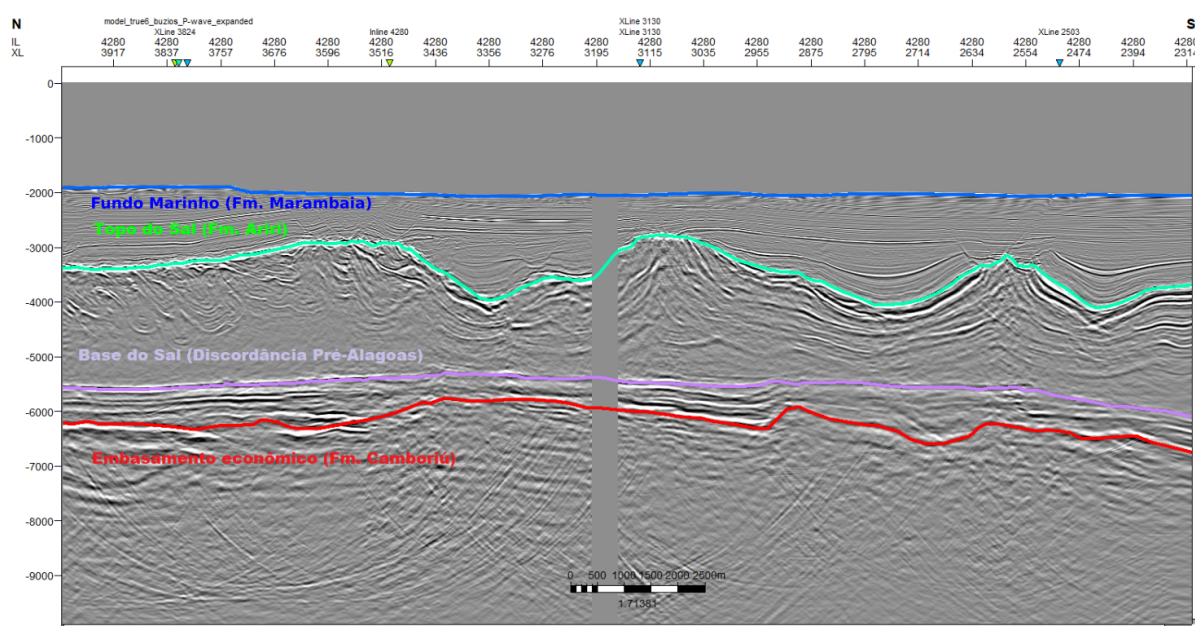
3.3.2 Construção do modelo de velocidades conceitual

Para a construção do primeiro modelo conceitual baseado, nas informações do campo de Búzios, foram utilizados o cubo de dados sísmicos 3D pós empilhamento migrado em profundidade e os perfis sônicos de quatro poços, dentro da região do levantamento (COSTA et al., 2019) .

A interpretação dos dados sísmicos foi realizada ao longo de todo o cubo sísmico, e foram considerados apenas quatro horizontes principais: Formação Marambaia (fundo marinho), Formação Ariri (topo do sal), Discordância Pré-Alagoas (base do sal) e a Formação Camboriú (embasamento econômico). Todos os horizontes são bem definidos na sísmica, com exceção da Formação Camboriú, que foi mapeada com a ajuda do atributo sísmico *TecVa* (BULHÕES; AMORIM, 2005).

Todos os marcadores foram localizados a partir do perfil composto de cada poço sendo usados para a interpretação sísmica ao longo de todo o cubo sísmico, gerando as principais geometrias do modelo geológico utilizado na construção do modelo de velocidade conceitual.

Figura 21 – Seção sísmica e os horizontes interpretados das principais formações utilizada para criação do modelo geológico conceitual do campo de Búzios. O horizonte azul é o fundo marinho (Fm. Marambaia), em verde é o topo do sal (Fm. Ariri), o horizonte roxo é a Discordância Pré-Alagoas e o horizonte vermelho é a Fm. Camboriú.



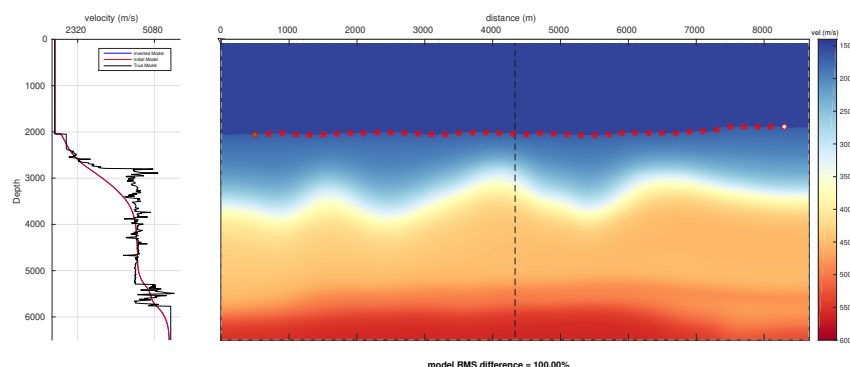
A figura 21 apresenta a seção selecionada para a construção do modelo conceitual 2D do campo de Búzios. A ausência de dados na região central da seção foi provocada por uma falha no arquivo de dados sísmicos. Nessa região, os horizontes foram interpolados. Entretanto, novos dados já foram solicitados para solucionar esse problema, assim os modelos que serão desenvolvidos em trabalhos futuros não sofrerão com o mesmo problema.

Após a interpretação dos horizontes, foi feita a extrapolação dos dados do perfil sônico, de modo que os valores de velocidades respeitassem a geometria dos horizontes interpretados. As regiões com profundidades maiores que 6500 m foram ignoradas, com o objetivo de reduzir o número total de amostras do modelo conceitual.

Assumiu-se que a região abaixo do reservatório possui velocidade constante igual à última amostra do perfil sônico, igual à 5668 m/s. O modelo gerado possui um espaçamento de malha de 5 m e 1729 amostras na horizontal e 1301 amostras na vertical. Para diminuir o custo computacional do cálculo do problema direto e inverso, o modelo conceitual foi re-amostrado, reduzindo o número de amostras. As dimensões finais do modelo conceitual foram de 865 amostras na horizontal e 651 amostras na vertical com um espaçamento de malha de 10 m.

A Figura 22 apresenta o modelo inicial, baseado no modelo conceitual de Búzios. Esse modelo de velocidade foi obtido através da suavização do modelo verdadeiro (Figura 20) com o programa *smooth2* utilizando os parâmetros $r_1 = r_2 = 50$. Esse modelo é utilizado como modelo inicial no processo de inversão.

Figura 22 – Modelo de velocidade inicial criado a partir da suavização do modelo de referência Búzios (Figura 20)



3.4 Os dados calculados

Os dados calculados $\mathbf{d}_{cal}^s(\mathbf{m}; \mathbf{x}^r, t)$ são obtidos resolvendo o problema direto (Equação 2.1), que nesse caso é a equação da onda acústica (Equação A.56). Assim, o operador $\mathcal{L}$ resolve a equação da onda e registra os campos nas posições dos receptores $\mathbf{x}^r$ para cada fonte na posição $\mathbf{x}_f$.

No caso acústico, o parâmetro do modelo $\mathbf{m}$ é o campo de velocidade de propagação das ondas acústicas $v_p(\mathbf{x})$. Assim, os dados calculados são obtidos através das seguintes etapas:

Algorithm 1: Dados Calculados

1 - Solução da equação da onda por diferenças finitas para cada $\mathbf{x}_f$;

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) = w(\mathbf{x}_f, t);$$

2 - Registro dos campos de pressão na posição dos receptores $\mathbf{x}^r$;

$$\mathbf{d}_{cal}^s(\mathbf{m}; \mathbf{x}^r, t) = p(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}^r).$$

3.5 Linguagem de programação

A linguagem de programação FORTRAN foi utilizada nesse trabalho para obter a solução da equação da onda acústica. Além disso, essa linguagem também foi utilizada para o desenvolvimento do algoritmo de cálculo do gradiente da função objetivo. A linguagem foi escolhida devido à sua alta performance, em problemas multidimensionais. As etapas de modelagem direta e cálculo do gradiente da função objetivo são as etapas que mais demandam esforços computacionais, porém, são altamente paralelizáveis.

Os algoritmos foram desenvolvidos de modo integrado às bibliotecas de paralelização MPI e a biblioteca de otimização SEISCOPE, descritas abaixo. Os algoritmos de modelagem sísmica e inversão sísmica foram construídos de forma que se adaptam aos diferentes arranjos de máquinas disponíveis no laboratório de pesquisa. O algoritmo é capaz tanto de rodar em uma única máquina com diversos processadores ou em um *cluster* de computadores utilizando todos os processadores disponíveis. Os algoritmos também foram adaptados para serem executados em

GPU's (Unidades de Processamento Gráfico)

3.6 Paralelização

Mesmo em casos bidimensionais, a demanda por processamento para a resolução do problema direto é grande. Levando em consideração toda a estratégia de inversão do campo de onda completo, a implementação serial convencional, onde cada tarefa é executada sequencialmente, é impraticável. Portanto, uma estratégia de implementação em paralelo é fundamental, ou seja, uma implementação que permita a execução de tarefas simultaneamente.

A solução da equação da onda no domínio do tempo, através método das diferenças finitas, permite duas estratégias de paralelização: a decomposição do domínio computacional em subdomínios e à paralelização das diversas fontes utilizadas na modelagem direta e reversa.

O primeiro paradigma de paralelização é recomendado para o caso em que o número de processadores disponíveis é muito superior ao número de tiros utilizados. Nessa situação, o domínio computacional é dividido em diversos subdomínios onde cada processador atualiza o campo de onda em um subdomínio. Nessa abordagem é necessário introduzir uma camada de troca de informação dos campos de onda entre os processadores (BOHLEN, 2002).

Quando o número de processadores disponíveis não é muito superior ao número de tiros, deve ser aplicado o paradigma de distribuição de tiros. Nessa estratégia cada processador disponível resolve um problema direto e inverso por tiro (VIRIEUX; OPERTO, 2009), o que torna o tempo de processamento da FWI proporcional ao número de processadores disponíveis.

Existem diversas bibliotecas que implementam essas funcionalidades. Nesse trabalho foi utilizada a biblioteca MPI (*Message Passing Interface*) para a paralelização dos tiros, por causa de sua simplicidade e fácil implementação junto a linguagem de programação FORTRAN, para a distribuição de tarefas em um *cluster* de computadores.

A adaptação para GPU's aplicou a paralelização no cálculo do Laplaciano ∇^2 ,

que percorre todos os pontos do campo de pressão para calcular a marcha no tempo (Equação B.13 ou B.14). A paralelização foi realizada através dos usos das diretivas de programação OpenAcc da NVIDIA.

3.7 A biblioteca de otimização SEISCOPE

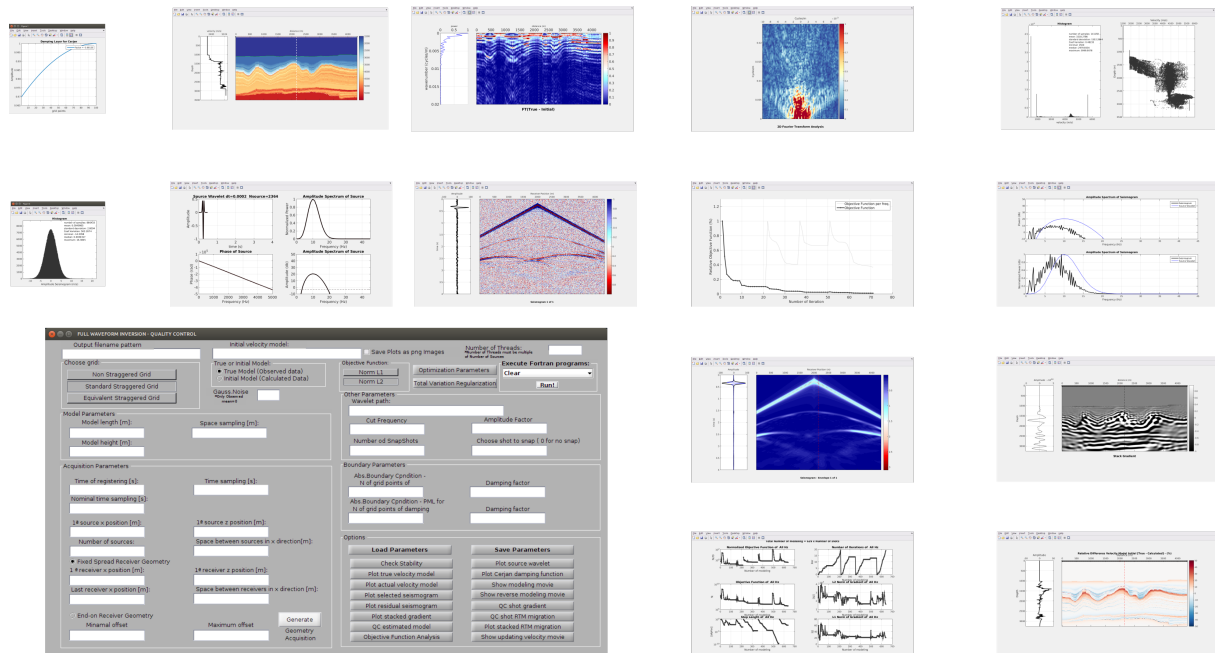
A biblioteca de otimização SEISCOPE (MÉTIVIER; BROSSIER, 2016) foi utilizada para a solução do problema de otimização não-linear. A biblioteca SEISCOPE foi projetada para possuir uma fácil integração com os diferentes problemas de otimização estudados pela geofísica. Essa biblioteca é formada por um conjunto de rotinas que podem ser facilmente integradas com os algoritmos que resolvem problemas de otimização de larga escala.

As rotinas estão organizadas de modo que cada método de otimização (gradiente descendente, gradiente conjugado, LBFGS) possa ser utilizado facilmente, evitando a complexidade associada a implementação. Todos os métodos compartilham da mesma estratégia de busca linear, garantindo a convergência e a diminuição constante da função objetivo. A estratégia de busca linear implementada garante que as condições de Wolfe sejam respeitadas.

3.8 Interface para controle de qualidade

Foi desenvolvida uma interface gráfica (Figura 23) para controlar os diversos aspectos do algoritmo de inversão, tais como os parâmetros de modelagem direta e reversa. Além disso, essa interface gráfica é utilizada para o controle de qualidade dos resultados parciais. Sendo possível acompanhar a atualização do modelo de velocidade, a variação do gradiente da função objetivo utilizado no método de otimização, além de permitir a visualização dos sismogramas e evolução das funções objetivos.

Figura 23 – Interface gráfica para controle de qualidade



4 Resultados

Nessa seção são apresentados os resultados dos testes de inversão utilizando o modelo Marmousi e o modelo Búzios. Os resultados são discutidos na seção seguinte. Para cada caso são apresentados exemplos da evolução no tempo do campo de pressão (*snapshot*), assim como os campos de pressão registrados na superfície (sismogramas). Os exemplos de dados residuais, calculados a partir da diferença entre os dados observados e dados calculados, também são apresentados.

São exibidos exemplos do gradiente da função objetivo utilizando as duas parametrizações do problema direto, na velocidade e no quadrado da vagarosidade. Exemplos do gradiente utilizando a norma L1 também são exibidos nessa seção. O efeito da aplicação da regularização por *Total Variation* no gradiente também é apresentado para cada modelo.

Os modelos de velocidades estimados pelos métodos de otimização gradiente descendente, gradiente conjugado e quasi-Newton LBFGS são apresentados utilizando o modelo Marmousi e o modelo baseado nos dados do campo de Búzios. Também foram realizados testes utilizando a Norma L1 e a regularização por *total variation*. Os métodos que forneceram os melhores resultados foram utilizados nos testes incluindo no fluxo de inversão a estimativa da *wavelet*.

A análise dos números de ondas recuperadas pela FWI é uma forma de avaliar a qualidade do modelo de velocidade estimado pelos diferentes métodos de otimização. Após a estimativa dos modelos de velocidades foram analisados os números de onda recuperados em parte dos resultados. Por fim, as imagens geradas pela RTM utilizando o modelo de velocidade inicial foi comparada as imagens geradas utilizando os modelos estimados pela FWI.

4.1 Modelagem sísmica

A modelagem sísmica é responsável pela simulação de uma aquisição sísmica. Nesse processo são gerados os sismogramas utilizados nas diversas etapas da FWI. Nos

testes realizados nessa seção foi utilizada uma wavelet Ricker, com frequência de corte de 30 Hz. O mesmo operador de modelagem sísmica é utilizado no cálculo do gradiente da função objetivo, modificando a fonte utilizada, que no caso do cálculo do gradiente é função do resíduo dos dados (fonte adjunta) e além disso, a modelagem é calculada reversamente no tempo.

4.1.1 Modelo Marmousi

Nessa seção serão apresentados alguns exemplos dos resultados da modelagem sísmica utilizando o modelo de referência Marmousi (Figura 18) e o modelo inicial (Figura 19). A geometria de aquisição utilizada simula um levantamento sísmico utilizando OBNs. Também foi utilizado o princípio da reciprocidade para trocar as posições das fontes com os receptores e reduzir o número de modelagens. As posições das fontes estão indicadas em vermelho e as posições dos receptores estão indicadas em branco nas Figuras dos modelos de velocidades. Os parâmetros de aquisição utilizados nos testes, a seguir, estão descritos na tabela 1:

Tabela 1 – Parâmetros de aquisição utilizados nos exemplos dessa seção.

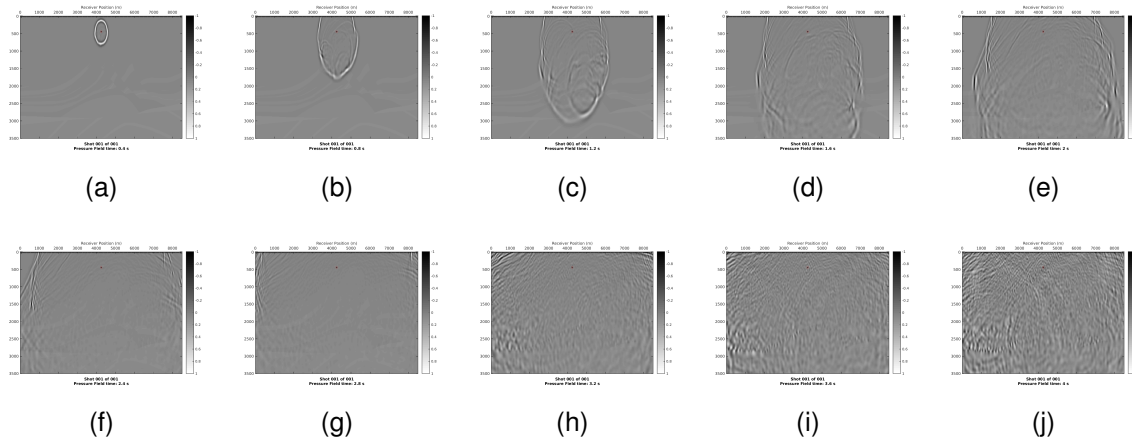
Parâmetro	Valor
Número de fontes	40
Espaçamento entre fontes	200 m
Número de receptores	169
Espaçamento entre receptores	50 m
Tempo de registro	8,0 s
Taxa de Amostragem	$4,0 \cdot 10^{-4}$ s
Fonte sísmica	wavelet Ricker
Frequência de corte	30 Hz

Instantâneos (*Snapshots*)

Os *snapshots* apresentados na Figura 24 são imagens da configuração do campo de pressão, para diferentes instantes de tempo, utilizando o modelo de referência Marmousi (Figura 18).

A visualização dos *snapshot* permite realizar um controle de qualidade para verificar o correto funcionamento do algoritmo de modelagem. E ainda, permite visualizar o comportamento das frentes de ondas no modelo de velocidade utilizado. É

Figura 24 – *Snapshots*. Distribuição das amplitudes do campo de pressão para diferentes instantes de tempo utilizando o modelo Marmousi.

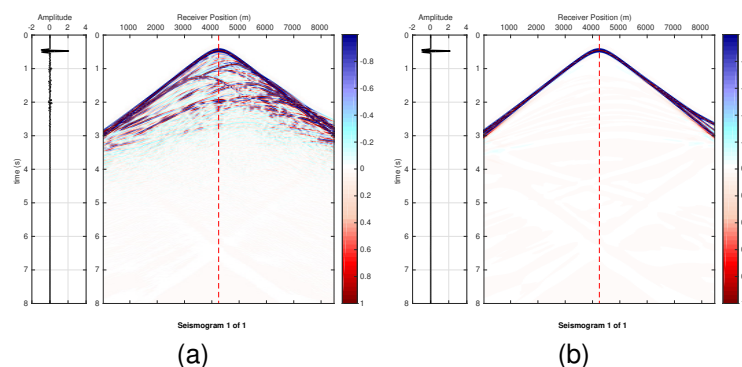


possível perceber que as frentes de onda iluminam as partes mais profundas do modelo.

Sismogramas

Exemplos de sismogramas gerados utilizando o modelo de referência e inicial são apresentados na Figura 25. Esses sismogramas são equivalentes a um sismograma organizado no domínio do receptor para uma aquisição OBN.

Figura 25 – Sismogramas na posição central do modelo a) Dado observado utilizando o modelo de referência Marmousi. b) Dado calculado utilizando o modelo inicial. A linha pontilhada indica a posição do perfil lateral.

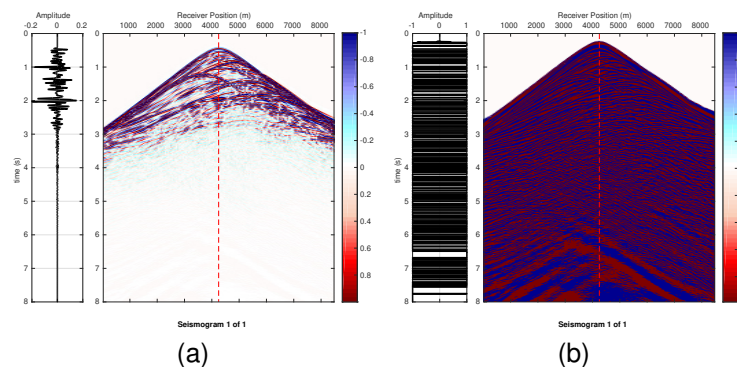


Nos dados calculados utilizando o modelo inicial (Figura 26b) somente é possível identificar claramente as primeiras chegadas e a onda direta. Por outro lado, no dado observado (Figura 26a) as diversas reflexões provocadas pelo modelo de referência podem ser observadas.

Sismograma residual

A Figura 27a apresenta o sismograma residual obtido através da diferença entre o sismograma observado (Figura 26a) e o sismograma calculado (Figura 26b). O resíduo é utilizado como fonte adjunta na modelagem reserva no tempo (Equação 2.79) quando a função objetivo utilizando a norma L2 é escolhida. Quando a norma L1 é usada, a fonte adjunta se torna o sinal do resíduo (Figura 27b), dado pela Equação 2.82. Para evitar divisão por zero no cálculo da fonte adjunta é somado um termo estabilizador ($1.0 \cdot 10^{-8}$) no denominador.

Figura 26 – a) Sismograma residual, foram utilizados o modelo de referência e o modelo inicial. b) Sinal do sismograma residual usado como fonte adjunta para o caso da norma L1.



4.1.2 Modelo Búzios

Da mesma forma que a seção anterior, serão apresentados alguns resultados da modelagem utilizando o modelo de referência Búzios (Figura 20) e o modelo inicial (Figura 22). Novamente foi simulada uma aquisição OBN e utilizado o princípio da reciprocidade para trocar a posição das fontes com os receptores. As posições das fontes estão indicadas em vermelho e as posições dos receptores estão indicadas em branco nas Figuras dos modelos. A tabela 2 apresenta os parâmetros de aquisição utilizados:

Snapshots

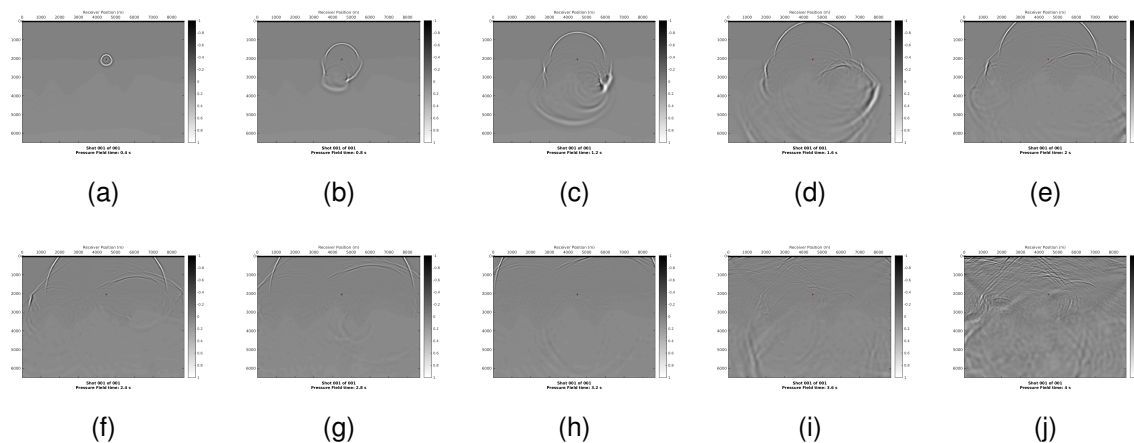
Na Figura 27 são apresentados as imagens do campo de pressão (*snapshots*) para o modelo de referencia Búzios (Figura 20) para vários instantes de tempo. A fonte

Tabela 2 – Parâmetros de aquisição utilizados nos exemplos dessa seção

Parâmetro	Valor
Número de fontes	40
Espaçamento entre fontes	200 m
Número de receptores	887
Espaçamento entre receptores	10 m
Tempo de registro	8,0 s
Amostragem	$4,0 \cdot 10^{-4}$ s
Fonte sísmica	<i>wavelet</i> Ricker
Frequência de corte	30 Hz

utilizada foi uma Ricker (Equação A.57) com frequência de corte de 30Hz.

Figura 27 – Snapshots. Distribuição das amplitudes do campo de pressão para diferentes instantes de tempo utilizando o modelo Búzios.

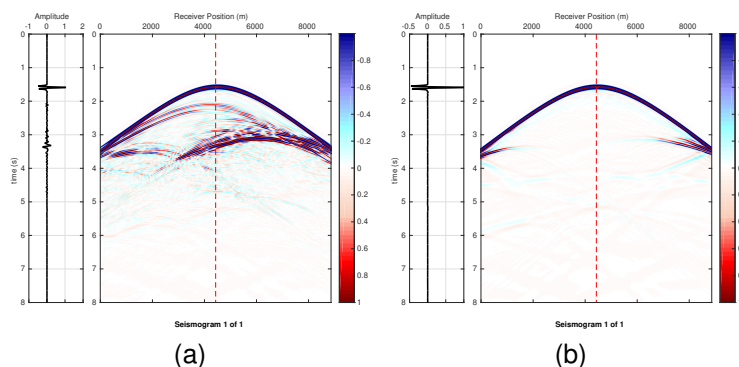


Nesse exemplo, é possível perceber que a maior parte da energia da frente de onda é refletida pelo topo da camada de sal (Fm. Ariri) devido ao grande contraste de velocidade. Porém, a região abaixo das camadas evaporíticas também é alcançada pelas frentes de onda.

Sismogramas

A Figura 28 apresenta um exemplo de sismograma observado e calculado obtidos utilizando o modelo Búzios (Figura 20) e o modelo inicial (Figura 22), respectivamente. Esses sismogramas são equivalentes a um sismograma organizado no domínio do receptor para uma aquisição OBN.

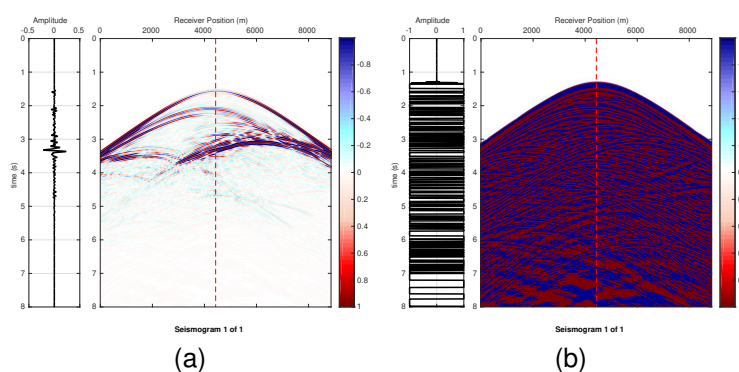
Figura 28 – a) Dado gerado utilizando o modelo de referência Búzios. b) Sismograma gerado utilizando o modelo inicial.



Sismograma residual

O sismograma residual, dado pela diferença entre os sismogramas observado e calculado para o modelo Búzios é apresentado na Figura 30a. O sinal do sismograma residual (Figura 30b) é usado como fonte adjunta para o caso da função objetivo ser definida como norma L1. É importante ressaltar que foi adicionado um valor pequeno no denominador para evitar divisão por zero.

Figura 29 – a) Simograma residual para o modelo inicial de Búzios. b) Sinal do resíduo usado como fonte adjunta no caso da norma L1.



4.2 Gradiente da função Objetivo

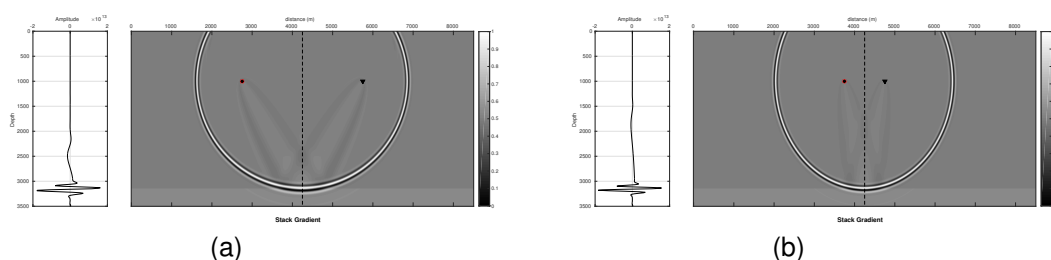
O cálculo do gradiente da função objetivo (Equação 2.65) permite estimar a direção de atualização utilizada nos métodos de otimização (Equações 2.30, 2.35, 2.36 e 2.44). Entretanto, para obter a convergência para o mínimo da função objetivo foi necessário condicionar seu gradiente de forma adequada.

Primeiramente, foi imposto que o modelo de velocidade não deve ser atualizado na camada de água. Para isso, o gradiente foi definido como zero na camada de água de todos os modelos. Em segundo lugar, quando aplicada a equação de atualização, o gradiente é escalado de forma que permita variações no modelo de velocidade de até 1% da velocidade máxima. Essas estratégias garantem que as perturbações no modelo de velocidade não provoquem instabilidades numéricas.

O gradiente da função objetivo, para um único par fonte receptor, foi chamado de caminho da onda (*wavepath*) por (WOODWARD, 1992). O caminho da onda pode ser usado como controle de qualidade do operador gradiente. A Figura 30 apresenta dois exemplos de caminho da onda para um modelo de velocidade com uma perturbação na parte mais profunda (o gradiente está sobreposto ao modelo de velocidade), a velocidade de fundo é de 2000 m/s e a perturbação possui velocidade de 2200 m/s.

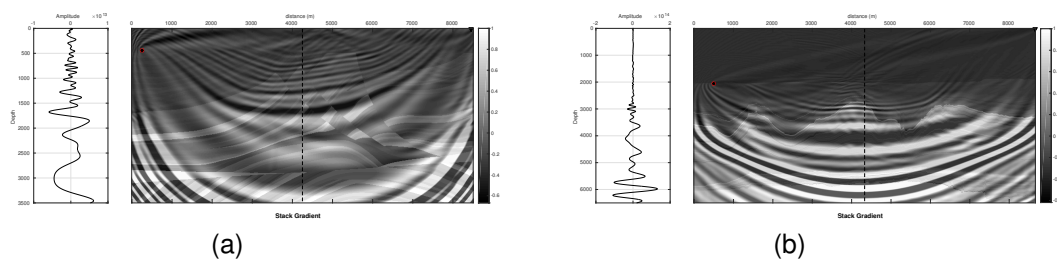
A elipse externa é formada pelo evento existente no resíduo dos dados, esse termo é o responsável pelo mapeamento dos refletores pelo gradiente. As estruturas no interior da elipse em forma de orelhas de coelho (*ear rabbit*) são formadas quando o campo adjunto é calculado em um modelo de velocidade incluindo a perturbação (ALKHALIFAH, 2014; HUANG; SCHUSTER, 2014). As Figuras 31a e 31b são exemplos de caminhos da onda para afastamentos de 3000 m e 1000 m, respectivamente.

Figura 30 – Gradiente da função objetivo para um único par fonte receptor, caminho da onda. Para *offsets* a) 3000 m e b) 1000 m.



O caminho da onda pode ser usado para o controle de qualidade da inversão. É possível determinar a região de iluminação limite determinada pela máxima distância entre fonte e receptor. Esse controle de qualidade foi realizado para os modelos Marmousi e Búzios com seus respectivos máximos afastamento fonte-receptor. Os gradientes estão sobrepostos aos modelos de velocidade. A wavelet fonte utilizada em todos os testes nessa seção foi uma Ricker, com frequência de corte 30 Hz.

Figura 31 – Gradiente da função objetivo para um único par fonte receptor, caminho da onda. a) Modelo Marmousi b) Modelo Búzios.

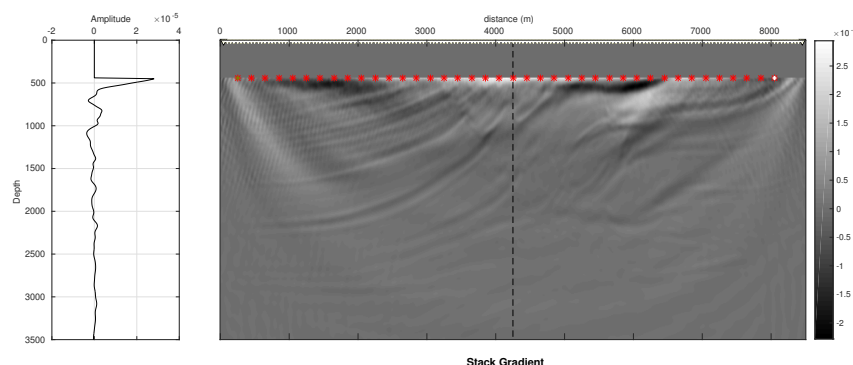


4.2.1 Modelo Marmousi

Os gradientes das funções objetivo utilizando todas as fontes e receptores para o Modelo Marmousi são apresentados nessa seção. O modelo de velocidade utilizado para o cálculo do campo adjunto foi o modelo de velocidade inicial (Figura 19).

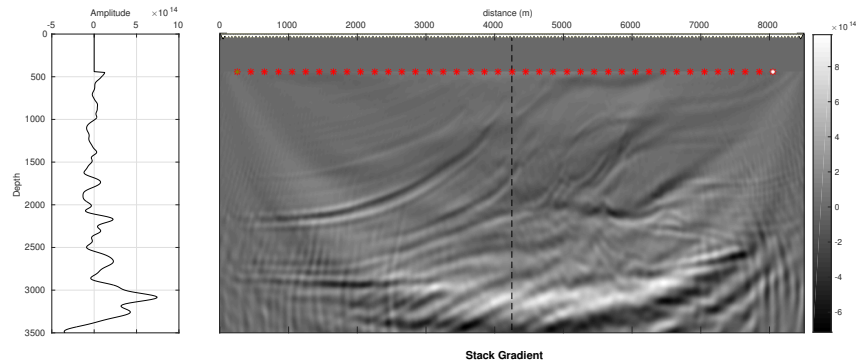
Foram comparados os gradientes utilizando as diferentes parametrizações do problema inverso. Quando o cálculo do gradiente utiliza a parametrização na velocidade (Figura 32), dada pela Equação 2.65, o gradiente não ilumina bem as partes mais profundas do modelo.

Figura 32 – Gradiente utilizando o modelo de velocidade inicial e parametrizado na velocidade.



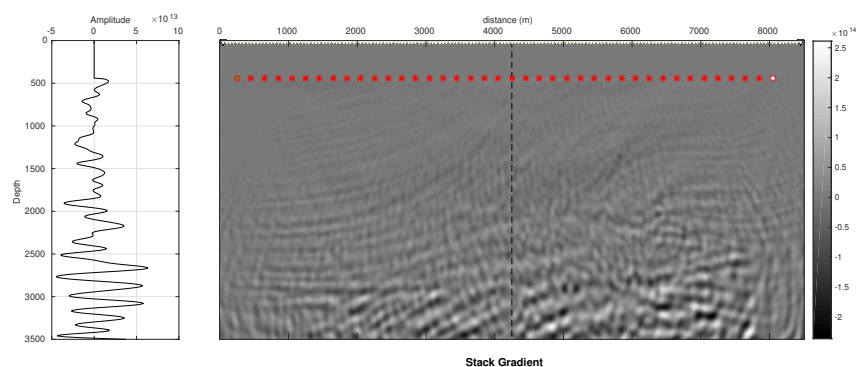
Em contrapartida, quando o gradiente é parametrizado no quadrado da vagarosidade (Figura 33), calculado através da Equação 2.94, o gradiente possui uma iluminação melhor nas partes mais profundas. As partes mais rasas também são bem iluminadas por essa parametrização.

Figura 33 – Gradiente utilizando o modelo de velocidade inicial e parametrizado no quadrado da vagarosidade.



Ambos os gradientes, apresentados nas Figuras 32 e 33, foram calculados utilizando uma fonte adjunta baseada na norma L2. A Figura 34 apresenta o gradiente da função objetivo parametrizado no quadrado da vagarosidade utilizando uma fonte adjunta baseada na norma L1. Nesse caso, o gradiente possui uma maior quantidade de artefatos porém é possível identificar as principais feições recuperadas pelo gradiente.

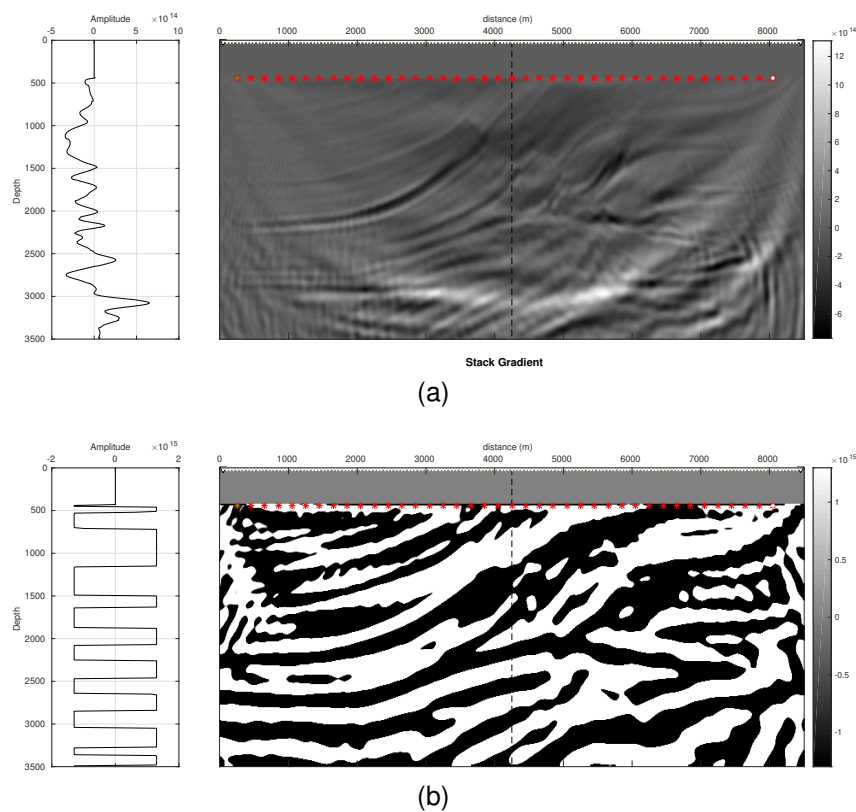
Figura 34 – Gradiente utilizando o modelo de velocidade inicial e parametrizado no quadrado da vagarosidade. A fonte adjunta utilizada foi baseada na norma L1.



Um exemplo de termo de regularização adicionado ao cálculo do gradiente da função objetivo (Equação 2.98) é apresentado na Figura 36a. A regularização por *total variation*, apresentada nesse exemplo, utilizou o modelo de velocidade estimado pelo método gradiente descendente incluindo a regularização por *total variation* após a primeira iteração. Como essa regularização utiliza a diferença entre o modelo

estimado e o modelo inicial a cada cálculo do gradiente, o termo de regularização só se torna diferente de zero quando o modelo estimado é diferente do modelo inicial, o que acontece após a primeira iteração FWI. A Figura 36b apresenta o termo de regularização adicionado ao gradiente. A análise do termo de regularização sugere que as principais feições geológicas do modelo são recuperadas, incluindo falhas e dobras.

Figura 35 – a) Gradiente incluindo o termo de regularização por *total variation*. O modelo de velocidade utilizado foi atualizado por uma iteração do algoritmo de inversão. b) Termo de regularização *total variation* após a primeira iteração.

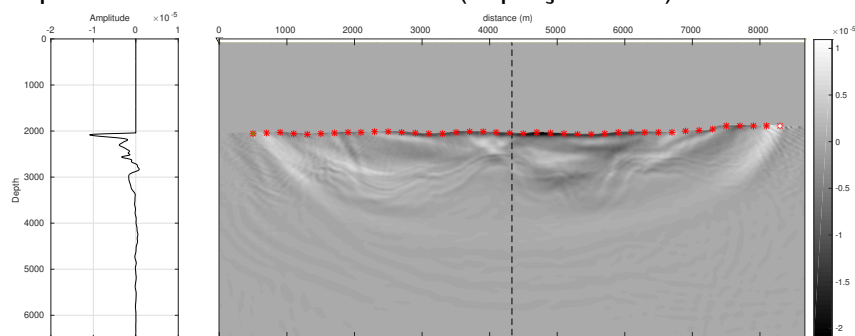


4.2.2 Modelo Búzios

Os gradientes das funções objetivo utilizando todas as fontes e receptores para o modelo Búzios são apresentados nessa seção. Para o cálculo do gradiente foi utilizado o modelo de velocidade inicial (Figura 22).

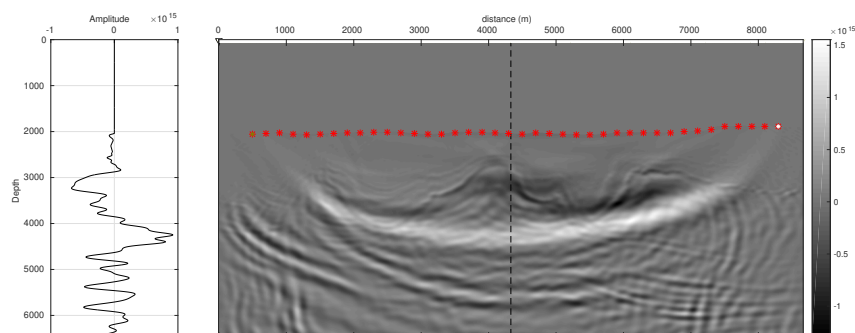
O gradiente da função objetivo parametrizado na velocidade (Equação 2.65), para o modelo Búzios, é apresentado na Figura 36. Neste caso, é possível perceber a dificuldade de iluminação das partes mais profundas do modelo. A parametrização do problema utilizando a velocidade privilegia a atualização nas partes mais rasas do modelo.

Figura 36 – Gradiente utilizando o modelo de velocidade inicial. Nesse caso o gradiente foi parametrizado na velocidade (Equação 2.65).



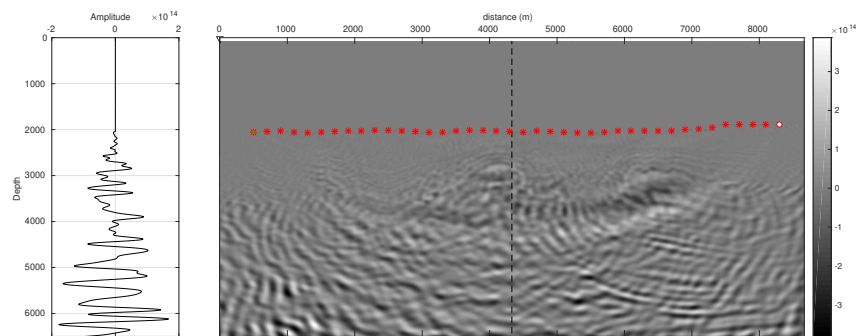
O gradiente da função objetivo parametrizado no quadrado da vagarosidade (Equação 2.94) é apresentado na Figura 37. Assim como no caso do modelo Marmousi, o gradiente da função objetivo é capaz de iluminar as partes mais profundas do modelo. Apesar de não ser possível identificar nitidamente a base do sal no gradiente utilizando o modelo inicial, após algumas atualizações do modelo de velocidade o gradiente é capaz de capturar as feições abaixo da camada de sal.

Figura 37 – Gradiente da função objetivo parametrizado na vagarosidade utilizando o modelo de velocidade inicial.



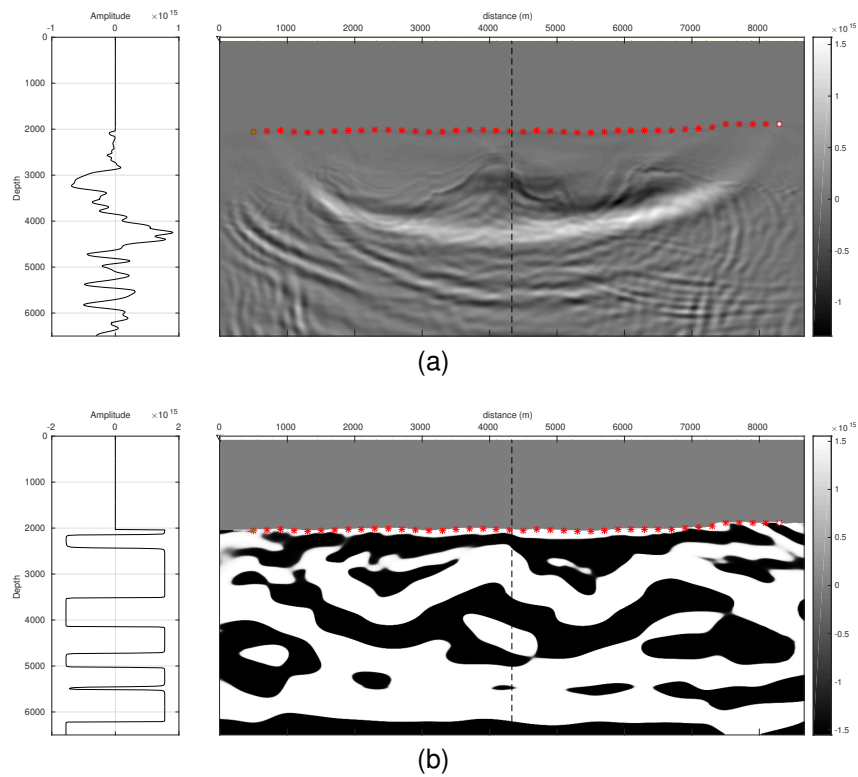
O gradiente da função objetivo utilizando a fonte adjunta baseada na norma L1 e parametrizado na vagarosidade é apresentado na Figura 38. Neste caso, o gradiente apresentou uma grande quantidade de artefatos entretanto, é possível identificar feições semelhantes ao caso utilizando a norma L2.

Figura 38 – Gradiente parametrizado na vagarosidade com a norma L1 utilizando o modelo de velocidade inicial.



A Figura 40a apresenta um exemplo de gradiente incluindo o termo de regularização aplicada ao modelo de Búzios, obtido através da Equação 2.97. Nesse exemplo, o modelo de velocidade utilizado foi após a primeira iteração do algoritmo de inversão, pois o termo de regularização somente é diferente de zero quando o modelo estimado é diferente do modelo inicial. O termo de regularização para a primeira atualização do modelo de Búzios (Figura 40b) apresente estruturas coerentes com o modelo de referência que deu origem aos dados observados. É importante destacar que o termo de regularização é recalculado a cada atualização do modelo de velocidade.

Figura 39 – a) Gradiente da função objetivo incluindo o termo de regularização por *total variation*. b) Termo de regularização *total variation* após a primeira atualização do modelo de velocidade.



4.3 Inversão modelo Marmousi

Nessa seção serão apresentados os resultados dos experimentos numérico para estimativa do modelo de velocidade, através dos métodos de otimização apresentados na seção 2.4. A posição das fontes e dos receptores estão indicadas nos modelos de velocidades em vermelho e branco, respectivamente. A linha vertical pontilhada na região central do modelo indica a posição do perfil de velocidade apresentado na lateral, em todos os modelos. O perfil do modelo inicial (linha vermelha) e do modelo de referência (linha preta) são apresentados no mesmo gráfico. Os parâmetros da aquisição utilizados nesses experimentos são apresentados na tabela 3. Vale ressaltar que o princípio da reciprocidade foi utilizado para detonar a fonte na posição dos receptores e registrar o campo de onda na posição dos tiros para reduzir o tempo de processamento.

Tabela 3 – Parâmetros de aquisição utilizados nos exemplos dessa seção

Parâmetro	Valor
Número de fontes	40
Espaçamento entre fontes	200 m
Número de receptores	169
Espaçamento entre receptores	50 m
Tempo de registro	8,0 s
Amostragem	$4,0 \cdot 10^{-4}$ s
Fonte sísmica	<i>wavelet</i> Ricker
Frequência de corte	12 -21 Hz

A abordagem multiescala no domínio da frequência utilizou as frequências de 0-12, 0-15, 0-18, 0-21 Hz. A escolha das frequências de corte utilizadas na inversão foi baseada no critério de não-dispersão (Equação B.18) e também no tempo de processamento.

Nesses experimentos, os dados observados são gerados com a mesma Ricker utilizada para gerar os dados calculados. O modelo de velocidade inicial utilizado nos experimentos foi apresentado na Figura 19. O gradiente da função objetivo foi parametrizado no quadrado da vagarosidade. E o modelo estimado ao final de cada faixa de frequência é utilizado como entrada da frequência subsequente.

Os parâmetros de otimização utilizados em cada faixa de frequência estão descritos na tabela 4. Os parâmetros de otimização são os mesmos para todos os testes. A tolerância para os limites superior e inferior das velocidades permitidas impõe

a seguinte condição para o modelo estimado de acordo com os parâmetros escolhidos: $1000 + 0,01 \leq v_p \leq 6000 + 0,01$. Além disso, foram fixada 10 iterações para cada faixa de frequência.

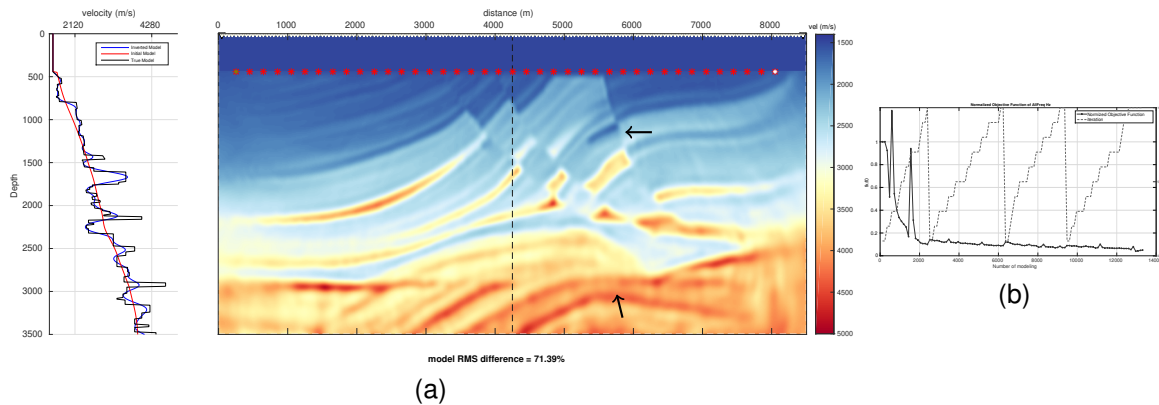
Tabela 4 – Parâmetros de otimização utilizados na FWI

Parâmetro	Valor
Número máximo de iterações	10
Tolerância critério de parada	$1,0 \times 10^{-5}$
Maior velocidade permitida	6000,00 m/s
Menor velocidade permitida	1000,00 m/s
Tolerância para os limites de velocidade	0,01 m/s

4.3.1 Método gradiente descendente

A Figura 41a apresenta o resultado para a inversão utilizando o método gradiente descendente. A diferença RMS relativa entre o modelo de referência e o modelo estimado foi de 71,38 %. As falhas, dobras e anomalias de baixa velocidade foram recuperadas pelo método. A evolução da função objetivo é apresentada na Figura 41b. O comportamento da função objetivo indica que o modelo convergiu para um mínimo. Somente na primeira banda frequência, a função objetivo já alcançou valores menores que 20%.

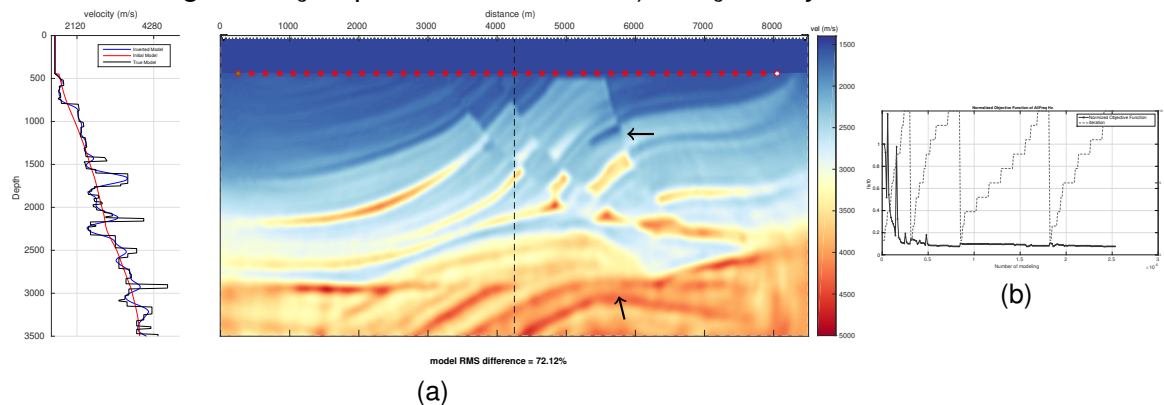
Figura 40 – a) Modelo estimado utilizando o método gradiente descendente. b) Função objetivo.



Regularização por *total variation*

A Figura 42a apresenta os resultados da inversão utilizando o método gradiente descendente incluindo a regularização por *total variation*. O parâmetro de regularização utilizado foi $\lambda_R = 0,01$. Esse valor foi obtido após a realização de diversos testes e escolhendo o melhor resultado. Nos demais testes foram utilizados o mesmo parâmetro de regularização. A diferença RMS relativa foi ligeiramente maior que o caso anterior, 72,12%. A função objetivo (Figura 42b) indica que um mínimo foi alcançado, entretanto, utilizou mais modelagens que o caso sem aplicação da regularização.

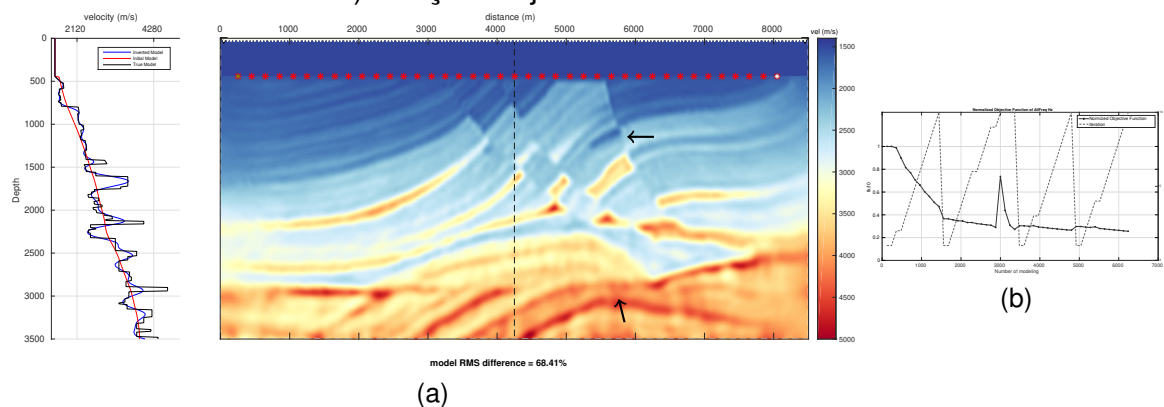
Figura 41 – a) Modelo estimado utilizando o método gradiente descendente incluindo a regularização por *total variation*. b) Função objetivo.



Norma L1

A Figura 43a apresenta o resultado da inversão utilizando o método gradiente descendente utilizando a norma L1. Nesse caso, o modelo estimado ficou mais próximo do modelo de referência obtendo uma diferença RMS relativa de 68,41%. Além de conseguir recuperar as principais estruturas do modelo, esse método teve um resultado melhor nas bordas do modelo. A função objetivo, apresentada na Figura 43b, se comportou de maneira muito diferente dos casos utilizando a norma L2. Esse fato se deve a forma de cálculo da função objetivo, que utiliza apenas a diferença absoluta do resíduo dos dados (Equação 2.7). Vale ressaltar que a utilização da Norma L1 utilizou menos modelagem para atingir o mínimo da função objetivo.

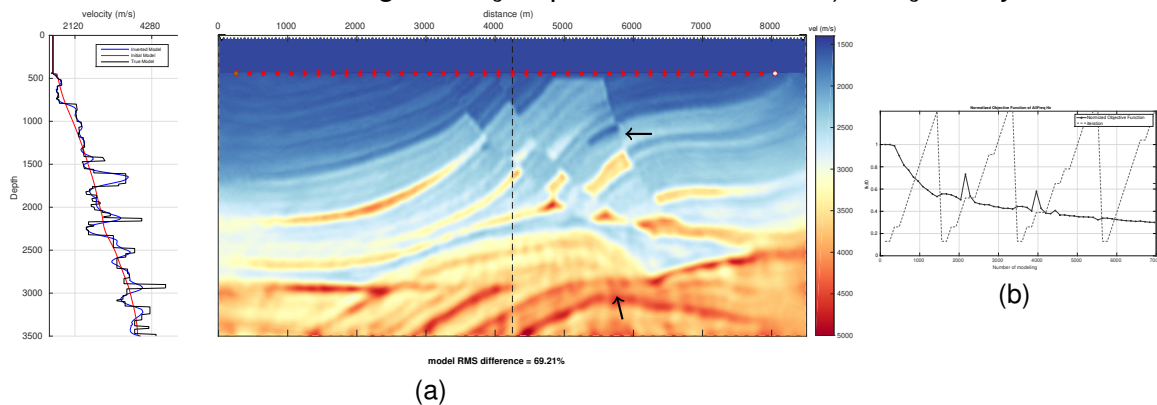
Figura 42 – a) Modelo estimado utilizando o método gradiente descendente utilizando a norma L1. b) Função Objetivo



4.3.1.1 Norma L1 + Regularização por *total variation*

A combinação entre a utilização da norma L1 com a regularização por *total variation* é apresentada na Figura 44a, que obteve uma diferença RMS relativa um pouco maior que anterior, 69,21%. A função objetivo para esse caso (Figura 44b) também aponta que esse resultado foi ligeiramente pior, utilizando mais modelagem e alcançando um mínimo da função objetivo maior que o caso utilizando somente a norma L1.

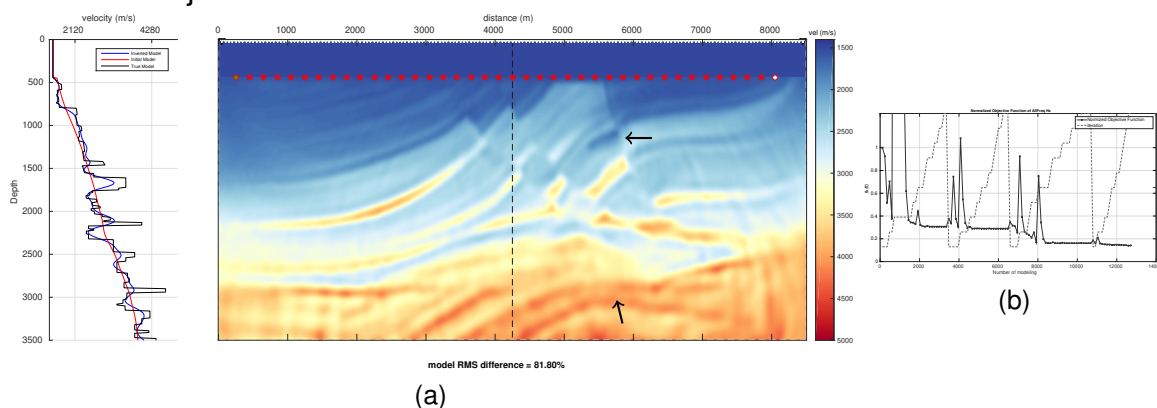
Figura 43 – a) Modelo estimado utilizando o método gradiente descendente incluindo a norma L1 e a regularização por *total variation*. b) Função objetivo.



4.3.2 Gradiente conjugado

O resultado para o método gradiente conjugado é apresentado na Figura 45a. Nesse caso, o modelo estimado foi pior que os obtidos pelo método gradiente descendente, como indica a diferença RMS relativa de 81,90%. Porém esse método utilizou menos modelagens que o método gradiente descende, como indica a Figura 45b.

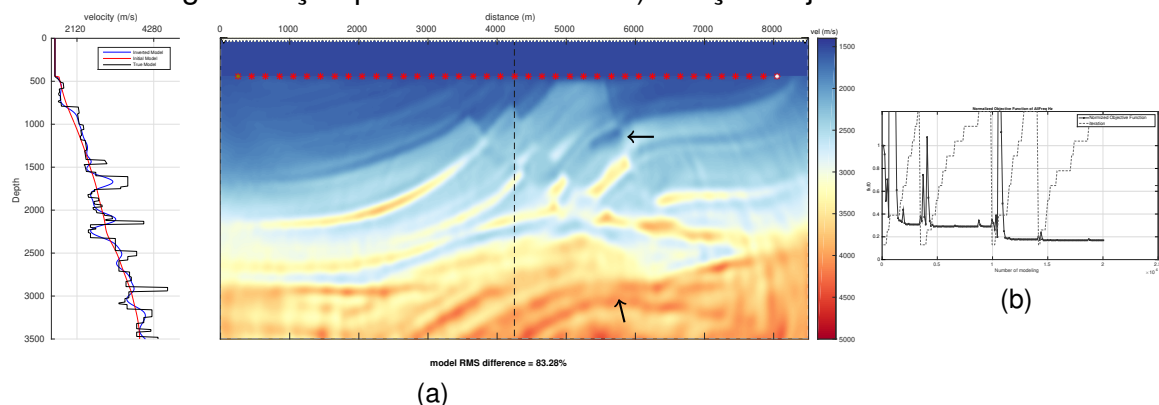
Figura 44 – a) Modelo estimado utilizando o método Gradiente Conjugado. b) Função objetivo.



regularização por *total variation*

A Figura 46a o resultado para o método gradiente conjugado incluindo a regularização por *total variation*, a diferença RMS relativa foi de 83,28%. Como no caso do gradiente descendente, a regularização por *total variation* foi ligeiramente pior quando comparada sem a aplicação da regularização. Inclusive aumentando o número de modelagem necessárias para a convergência, como indica a função objetivo (Figura 46b).

Figura 45 – Modelo estimado utilizando o método Gradiente Conjugado incluindo a regularização por *total variation*. b) Função objetivo.



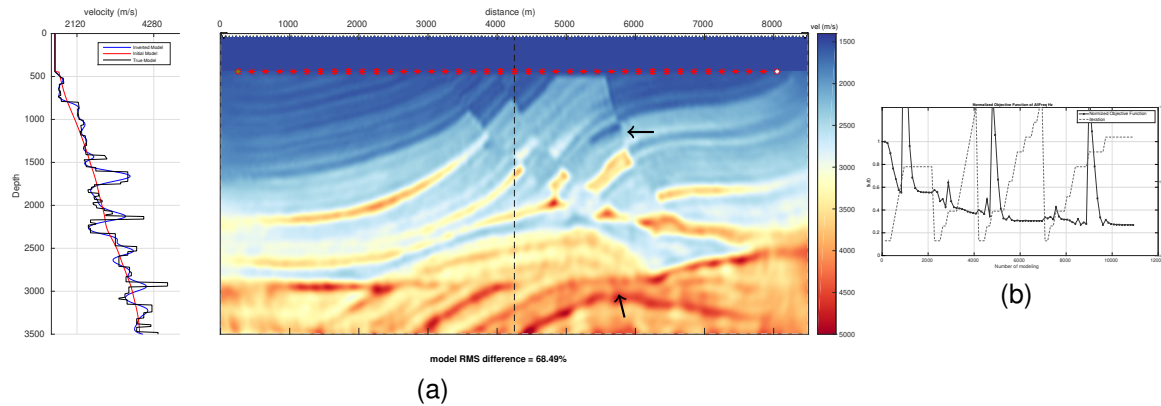
Norma L1

O método gradiente conjugado com a utilização da norma L1 (Figura 47a) obteve uma diferença RMS relativa de 68,49%, um resultado melhor que o caso utilizando a norma L2. A função objetivo para esse caso (Figura 47b) mostra que foram realizadas menos modelagens. Nesse caso, para a primeira e última banda de frequência não foram atingidas 10 iterações pois o valor da função objetivo foi igual a avaliação anterior, o que interrompe o algoritmo de inversão para a banda de frequência avaliada. Esse procedimento foi realizado para evitar cálculos desnecessários que não modificam o modelo de velocidade.

4.3.2.1 Norma L1 + Regularização por *total variation*

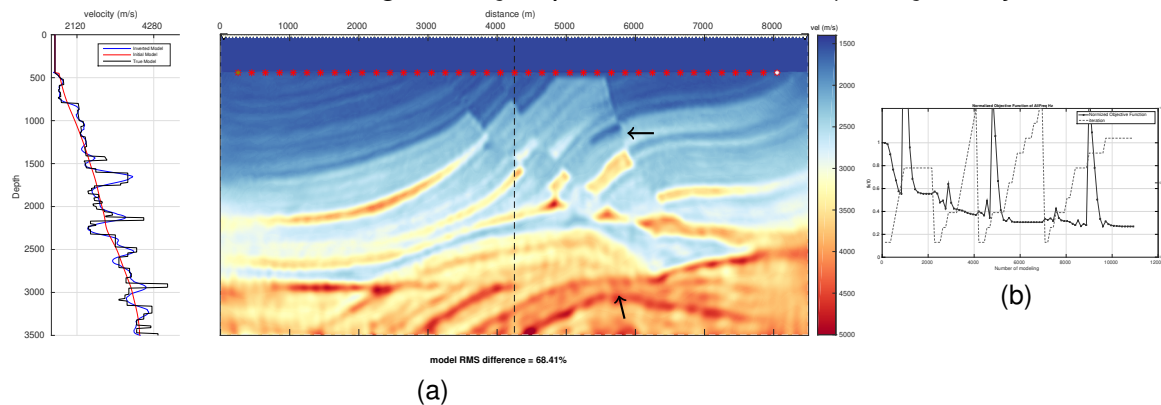
A Figura 48a apresenta o resultado para a inversão combinando a norma L1 com a regularização por *total variation*. Nesse caso, a diferença RMS relativa foi de 68,41%, praticamente a mesma que o caso anterior. O valor final da função objetivo e o número

Figura 46 – a) Modelo estimado utilizando o método Gradiente Conjugado incluindo a norma L1. b) Função objetivo.



total de modelagens (Figura 48b) sugere que a inclusão da regularização não influencia significativa o resultado.

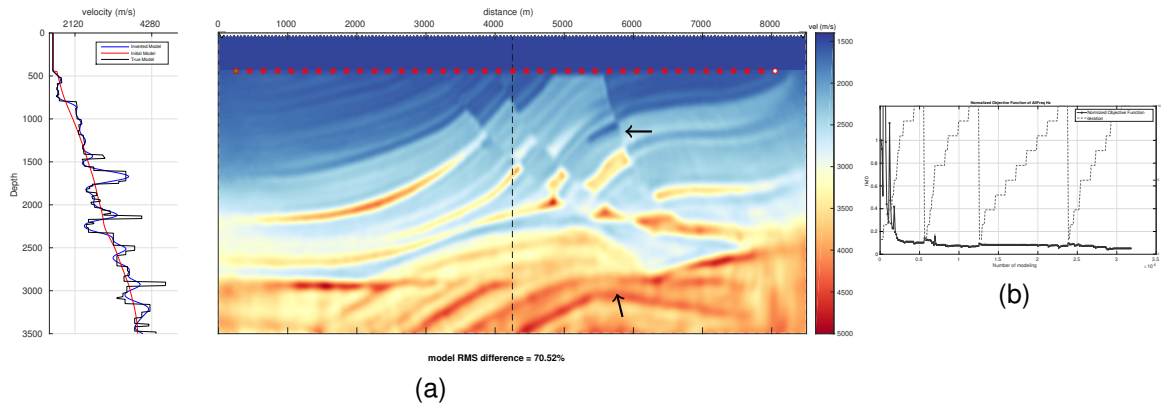
Figura 47 – a) Modelo estimado utilizando o método Gradiente Conjugado incluindo a norma L1 e a regularização por *total variation*. b) Função objetivo.



4.3.3 Método LBFGS

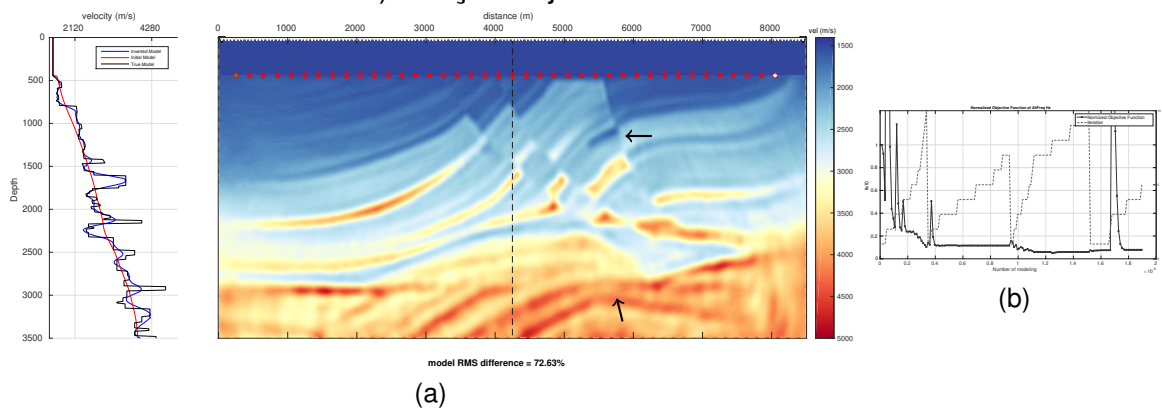
O resultado para o método LBFGS (Figura 49b) alcançou uma diferença RMS relativa de 70,52%, menor que os casos utilizando o gradiente descendente e gradiente conjugado. Entretanto, a Figura 49b com a função objetivo mostra que foram necessárias mais modelagens, contrariando a expectativa de um método mais eficiente.

Figura 48 – a) Modelo estimado utilizando o método LBFGS. b) Função objetivo.



Regularização por *total variation*

O resultado alcançado incluindo a regularização por *total variation*, apresentado na Figura 50a, obteve uma diferença RMS relativa de 72,63%, um pouco pior que o caso sem aplicação da regularização. Por outro lado, foram necessárias menos modelagens, como é possível observar na Figura 50b que também apresenta o comportamento da função objetivo. As bandas de frequências que não atingiram 10 iterações foram interrompidas pois o valor da função objetivo não se alterou após o cálculo da nova direção de atualização.

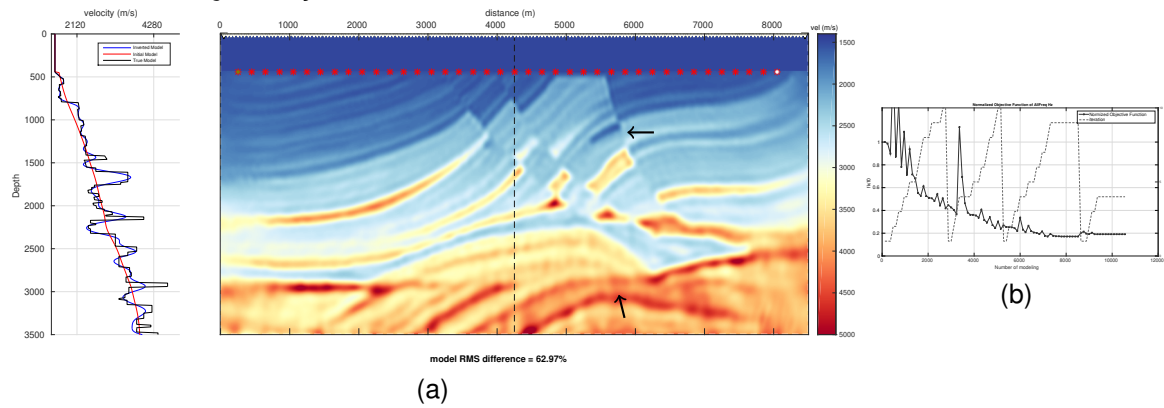
Figura 49 – Modelo estimado utilizando o método LBFGS incluindo a regularização por *total variation*. b) Função objetivo.

Norma L1

O resultado para o método LBFGS em conjunto com a norma L1 é apresentado na Figura 51a. Nesse caso, a diferença RMS relativa foi de 62,97%, o melhor resultado entre os testes com o modelo Marmousi. A função objetivo (Figura 51b) mostra que

foram utilizadas menos iterações quando aplicada a norma L1 em conjunto com o método LBFGS.

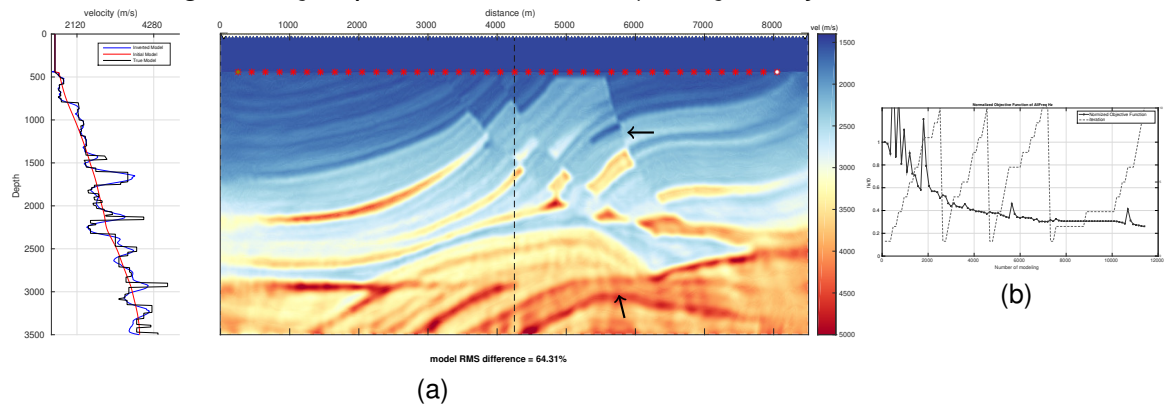
Figura 50 – Modelo estimado utilizando o método LBFGS incluindo a norma L1. b) Função objetivo.



Norma L1 + Regularização por *total variation*

O resultado para o método LBFGS utilizando a norma L1 e a regularização por *total variation* é apresentado na Figura 52a. Esse resultado foi ligeiramente pior que o caso utilizando apenas a norma L1, alcançando uma diferença RMS relativa de 64,31%. Além disso, a função objetivo (Figura 52b) alcançou um mínimo um pouco maior que o caso anterior.

Figura 51 – Modelo estimado utilizando o método LBFGS incluindo a norma L1 e a regularização por *total variation*. b) Função objetivo.



4.4 Inversão modelo Búzios

Nessa seção serão apresentados os resultados da inversão, utilizando os diferentes métodos de otimização abordados nos referenciais teóricos, aplicado ao modelo conceitual para o campo de Búzios (Figura 20). O modelo inicial, utilizado em todos os experimentos numéricos, foi apresentado na Figura 22. A posição das fontes e dos receptores estão indicadas nos modelos de velocidades estimados, em vermelho e branco, respectivamente. Vale ressaltar que também foi utilizado o princípio da reciprocidade para trocar a posição das fontes com os receptores. A linha vertical pontilhada na região central do modelo indica a posição do perfil de velocidade do modelo estimado (linha azul), do modelo inicial (linha vermelha) e do modelo de referência (linha preta), apresentados na lateral de cada resultado. Os parâmetros de aquisição, utilizados nessa seção, estão apresentados na Tabela 5.

Tabela 5 – Parâmetros de aquisição utilizados nos exemplos dessa seção

Parâmetro	Valor
Número de fontes	40
Espaçamento entre fontes	200 m
Número de receptores	887
Espaçamento entre receptores	10 m
Tempo de registro	8,0 s
Amostragem	$4,0 \cdot 10^{-4}$ s
Fonte sísmica	<i>wavelet</i> Ricker
Frequência de corte	6 - 21 Hz

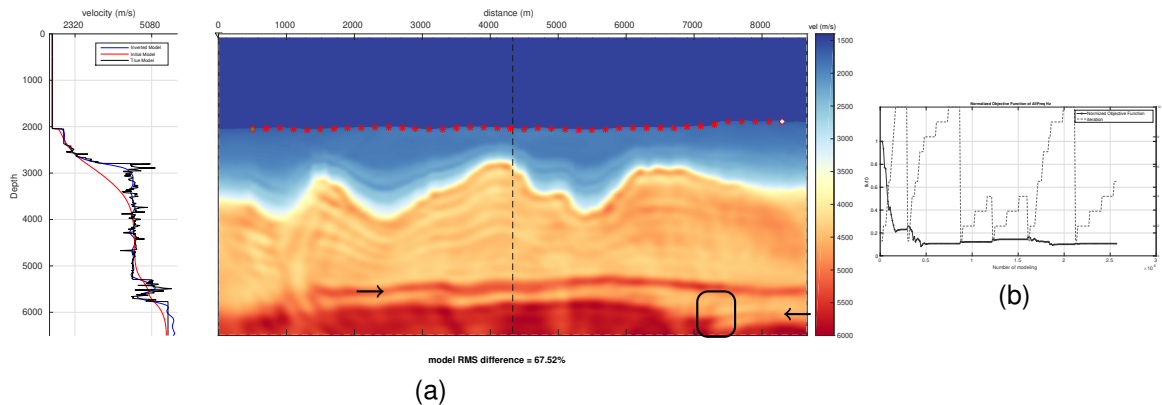
A abordagem multiescala no domínio da frequência utilizou as frequências 0-6, 0-9, 0-12, 0-15, 0-18 e 0-21 Hz. Da mesma forma que os testes anteriores, as frequências de corte foram definidas com base no critério de dispersão e estabilidade e também no tempo de processamento. Além disso, a mesma *wavelet* utilizada para a geração dos dados observados foi usada para obtenção dos dados calculados. E também os mesmos parâmetros de otimização apresentados na tabela 4 foram usado.

4.4.1 Gradiente Descendente

A Figura 53a apresenta o resultado para inversão usando o método gradiente descendente. Nesse caso, a diferença RMS relativa foi de 67,52%. No modelo estimado é possível identificar as principais características do modelo Búzios, utilizado

para a geração dos dados observados. A função objetivo apresentada na Figura 53b mostra que o algoritmo de inversão alcançou um mínimo. Assim como em alguns casos nos testes com o modelo Marmousi, algumas bandas de frequência não realizaram as 10 iterações previstas pois o valor da função objetivo se repetiu em uma avaliação subsequente.

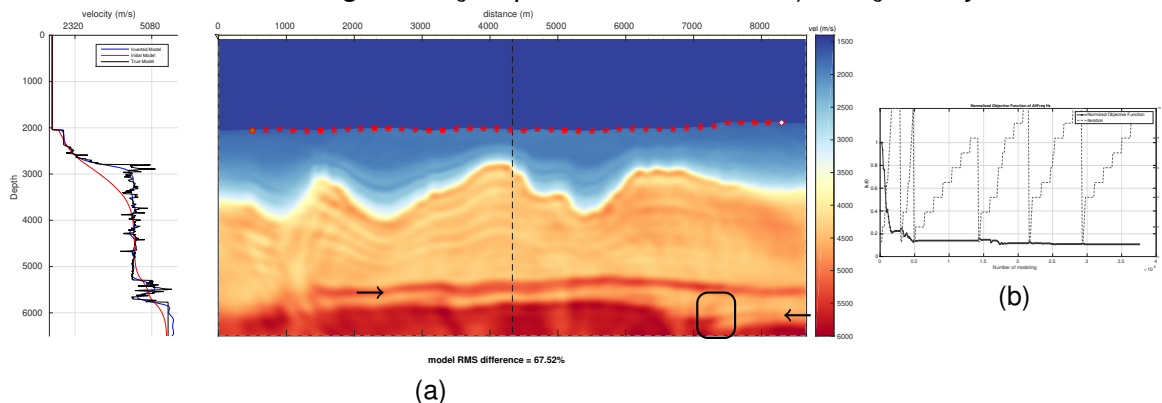
Figura 52 – a) Modelo estimado utilizando o método gradiente descendente. b) Função objetivo.



Regularização por *total variation*

O modelo estimado utilizando a regularização por *total variation* (Figura 54a) alcançou um resultado praticamente idêntico ao caso anterior. Além disso, o gráfico da função objetivo (Figura 54b) mostra que foram realizadas mais modelagens do que o caso sem aplicação da regularização.

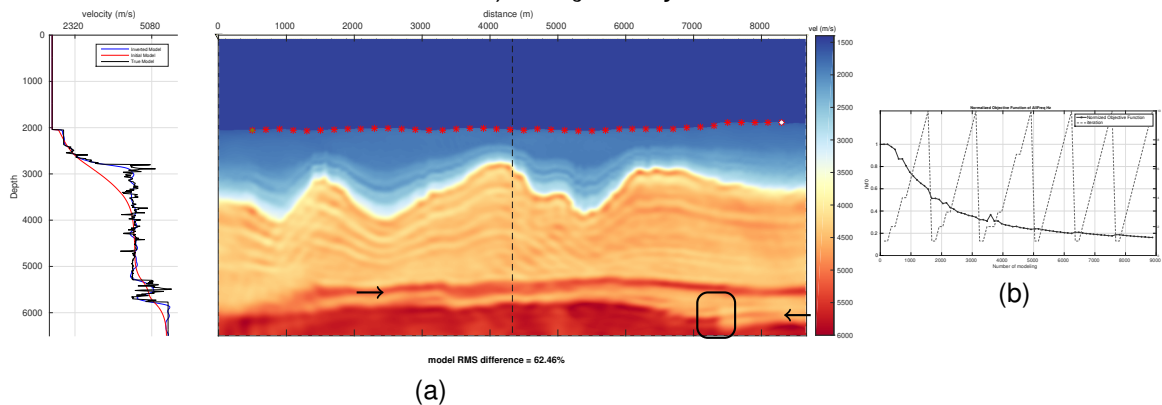
Figura 53 – Modelo de velocidade estimado pelo método gradiente descendente incluindo a regularização por *total variation*. b) Função objetivo.



Norma L1

A Figura 55a apresenta o resultado para o método gradiente descendente utilizando a norma L1, que alcançou uma diferença RMS relativa de 62,48%. Assim como nos caso com o modelo Marmousi, a aplicação da norma L1 alcançou uma diferença relativa menor que a aplicação da norma convencional. A função objetivo para a norma L1 (Figura 55b) tem um comportamento diferente quando comparado a norma L2 devido a forma como é feita a sua avaliação. Além disso, foram necessárias menos modelagens para a função objetivo alcançar o mínimo.

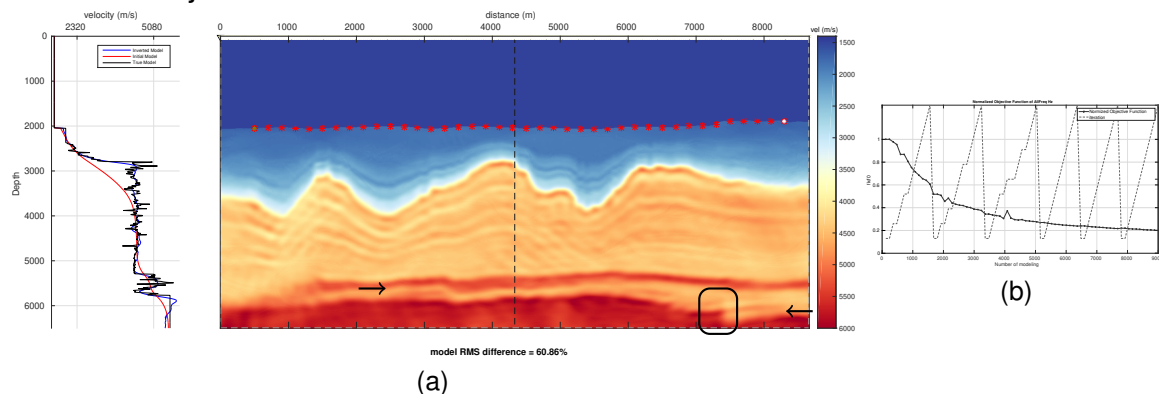
Figura 54 – a) Modelo de velocidade estimado pelo método gradiente descendente utilizando a norma L1. b) Função objetivo.



Norma L1 + Regularização por *total variation*

Nesse caso, a combinação da norma L1 com a regularização por *total variation*, apresentada na Figura 56a, forneceu uma diferença RMS relativa menor (60,86%). Ao contrário do que ocorreu quando foi utilizada a norma L1 para o modelo Marmousi. Por outro lado, a função objetivo (Figura 56b) teve praticamente o menos número de modelagens.

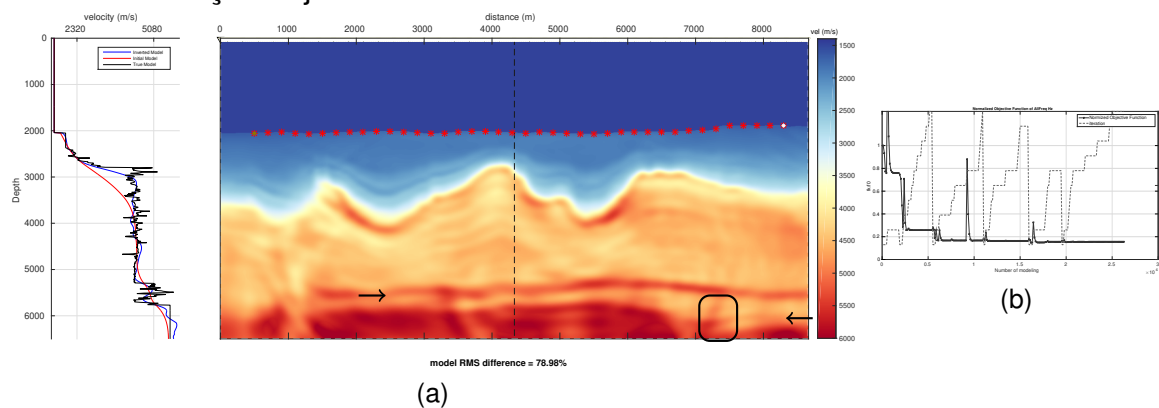
Figura 55 – Modelo de velocidade estimado pelo método Gradiente Descendente incluindo a norma L1 e a regularização por *total variation*. b) Função objetivo.



4.4.2 Método gradiente conjugado

A Figura 57a mostra o resultado para o caso da aplicação do método gradiente conjugado. A diferença RMS relativa alcançada por esse método foi de 78,98%, pior que todos os caso utilizando o gradiente descendente. A função objetivo apresentada na Figura 57b mostra que o método foi capaz de alcançar um mínimo.

Figura 56 – a) Modelo de velocidade estimado pelo método gradiente conjugado. b) Função objetivo.

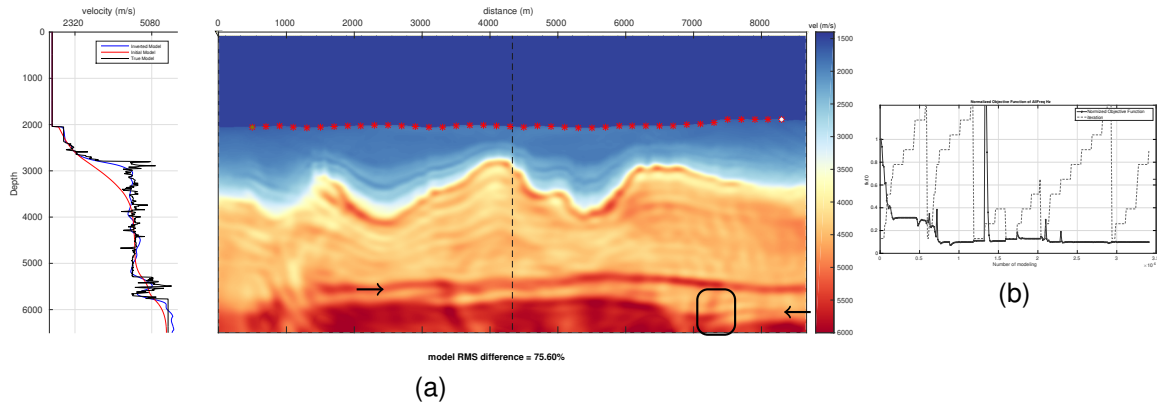


Regularização por *total variation*

Quando a regularização por *total variation* foi utilizada em conjunto com o método gradiente conjugado, forneceu um resultado melhor como mostra a Figura 58a alcançando uma diferença RMS relativa de 75,60%. Em contrapartida, o gráfico da

função objetivo (Figura 58b) mostra que foram necessárias mais modelagens para alcançar esse resultado.

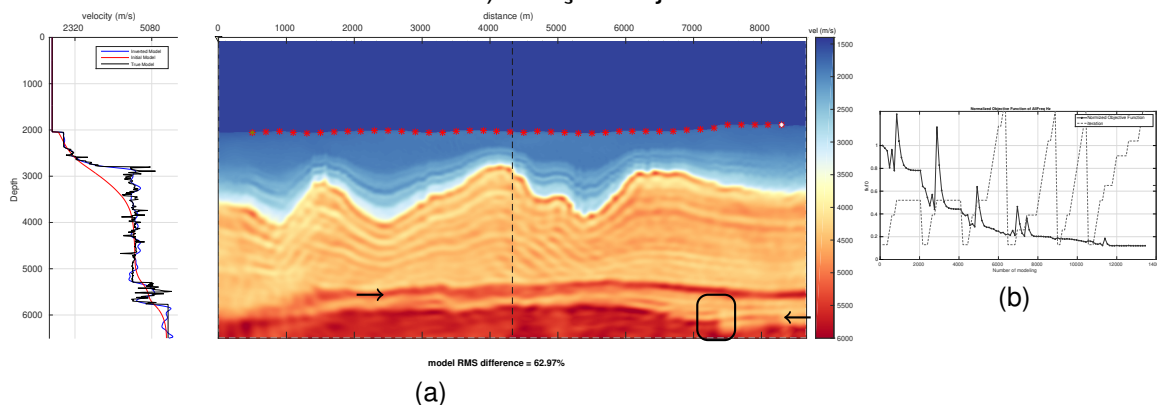
Figura 57 – a) Modelo de velocidade estimado pelo método gradiente conjugado incluindo a regularização por *total variation*. b) Função objetivo.



Norma L1

Mais uma vez a utilização da norma L1 forneceu melhores resultados que a aplicação da norma convencional, como indica a Figura 59a. A diferença RMS relativa para esse caso foi de 62,97%. Além disso, a função objetivo utilizou menos modelagens como indica a Figura 59b.

Figura 58 – a) Modelo de velocidade estimado pelo método gradiente conjugado incluindo a norma L1. b) Função objetivo.

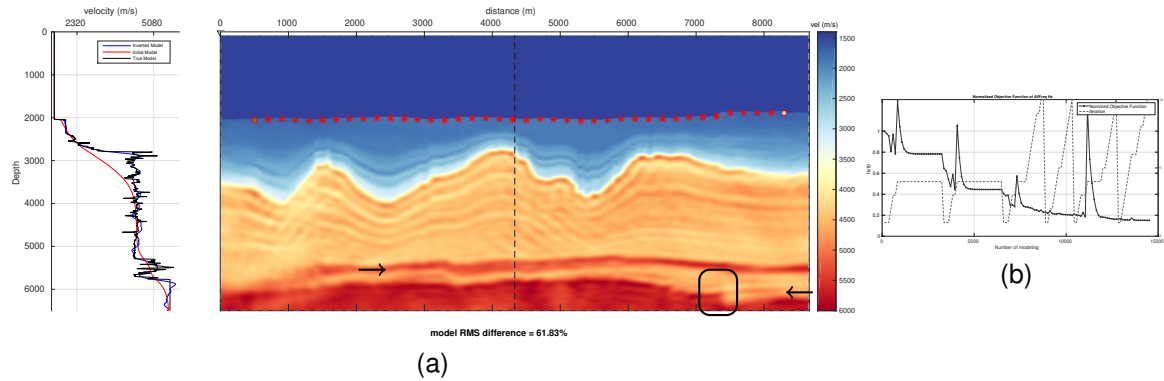


Norma L1 + Regularização por *total variation*

A combinação entre a norma L1 e a regularização por *total variation* para o método gradiente conjugado (Figura 60a) obteve uma diferença relativa menor que somente

usando a norma L1. Por outro lado, utilizou um pouco mais de modelagens que o caso usando somente a norma L1 como indica a Figura 60b.

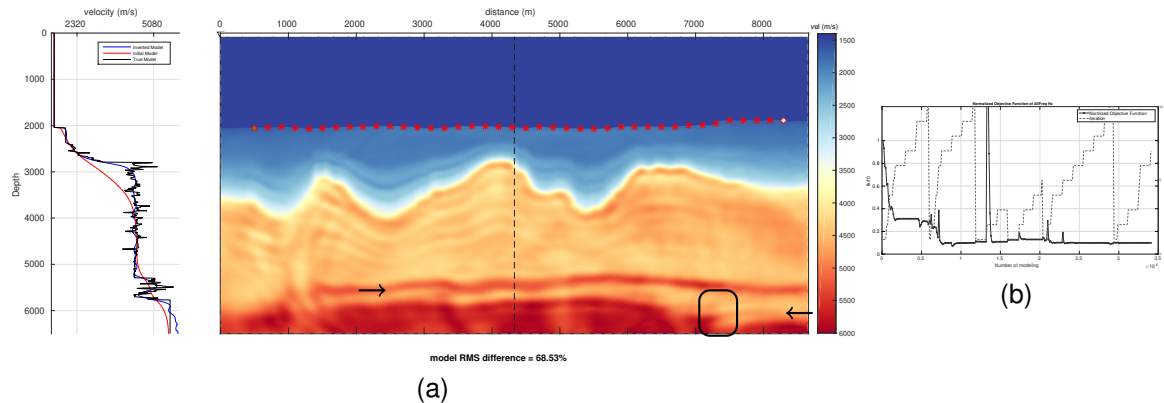
Figura 59 – a) Modelo de velocidade estimado pelo método gradiente conjugado incluindo a norma L1 e a regularização por *total variation*. b) Função objetivo.



4.4.3 Método LBFGS

A Figura 61a mostra o resultado para o método LBFGS, que alcançou uma diferença RMS relativa de 68,53%. Sua função objetivo (Figura 61b) indica que um mínimo foi alcançado.

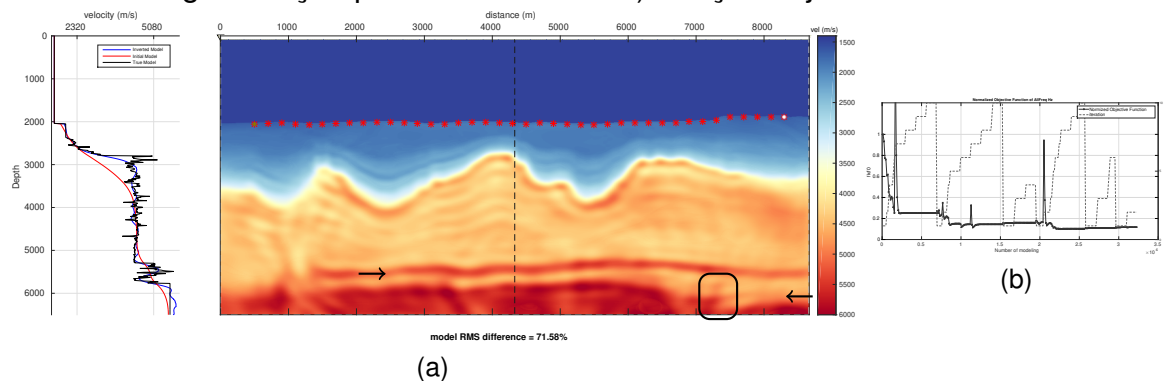
Figura 60 – a) Modelo de velocidade estimado pelo método LBFGS. b) Função objetivo.



Regularização por *total variation*

A aplicação por *total variation* não contribuiu para a redução da diferença relativa (71,58%) como mostra a Figura 62a. Assim como no caso anterior, a função objetivo (Figura 62b) alcançou um mínimo.

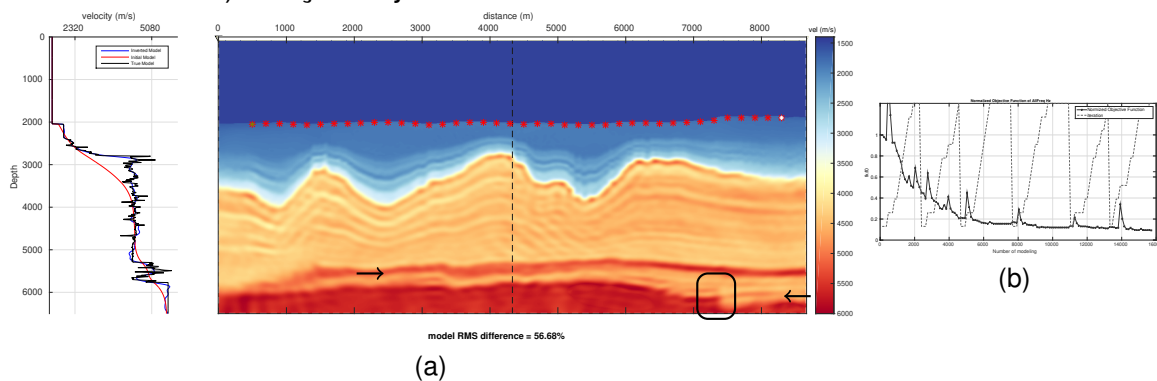
Figura 61 – a) Modelo de velocidade estimado pelo método LBFGS incluindo a regularização por *total variation*. b) Função objetivo.



Norma L1

Seguindo a tendência dos testes anteriores, a inversão utilizando a norma L1 apresentada na Figura 63a apresentou uma diferença relativa menor que os casos usando a norma L2. O gráfico da função objetivo (Figura 63b) nos tra que um mínimo foi alcançado e o método utilizou menos modelagens que os testes anteriores.

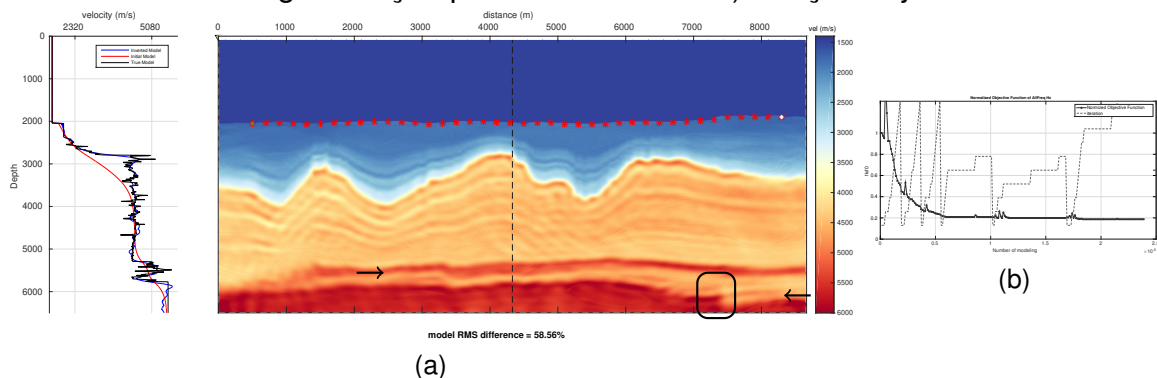
Figura 62 – a) Modelo de velocidade estimado pelo método LBFGS incluindo a norma L1. b) Função objetivo.



Norma L1 + Regularização por *total variation*

O resultado combinando a norma L1 com a regularização por *total variation*, apresentado na Figura 64a mostra que um resultado similar ao caso aplicando somente a norma L1. Entretanto, a inclusão da regularização por *total variation* aumentou o número total de modelagens necessárias (Figura 64b)

Figura 63 – a) Modelo de velocidade estimado pelo método LBFGS incluindo a norma L1 e a regularização por *total variation*. b) Função objetivo.



4.5 Estimativa da wavelet no contexto da FWI

Para a aplicação da inversão da *wavelet* fonte a um conjunto de dados 2D são necessários alguns ajustes para melhorar a qualidade da *wavelet* estimada. Primeiramente, deve-se selecionar o receptor mais próximo da fonte para cada tiro. Em seguida, comparar o traço escolhido com o traço calculado pela equação da onda 1D.

Em segundo lugar, a seleção (*picking*) da primeira chegada no dado observado é feita para a aplicação do algoritmo de inversão. Em terceiro lugar, a escolha de uma *wavelet* inicial que produza um dado calculado que possua fase similar ao dado observado. A *wavelet* inicial utilizada nesses experimentos numéricos foi a *wavelet* Ricker com frequência de corte de 45 Hz.

O campo adjunto $p^\dagger(z, t)$ (Equação 2.106) foi calculado com o mesmo operador por diferenças finitas para resolver a equação da onda 1D. Entretanto, a modelagem foi feita reversamente no tempo e o dado residual $\Delta d(z_{rec}, t) = \mathbf{d}_{obs}(z_{rec}, t) - \mathbf{d}_{cal}(z_{rec}, t)$ é a fonte adjunta. O campo adjunto registrado na posição do receptor $p^\dagger(z = z_{rec}, t)$ é a direção de atualização do problema de otimização não-linear, dado pela Equação 2.108.

Após o método de otimização gradiente descendente alcançar o mínimo, a fase da *wavelet* estimada é forçada ser zero e sua amplitude é reescalada para cada faixa de frequência usada na FWI. De modo que o dado calculado pelo operador da equação da onda 2D, usando a *wavelet* estimada como fonte, forneça uma amplitude semelhante ao dado observado. Esses ajustes implicam em sismogramas calculados mais semelhantes aos dados observados proporcionando melhores taxas de convergência ao algoritmo de FWI.

Tabela 6 – Parâmetros de aquisição utilizados nos exemplos dessa seção

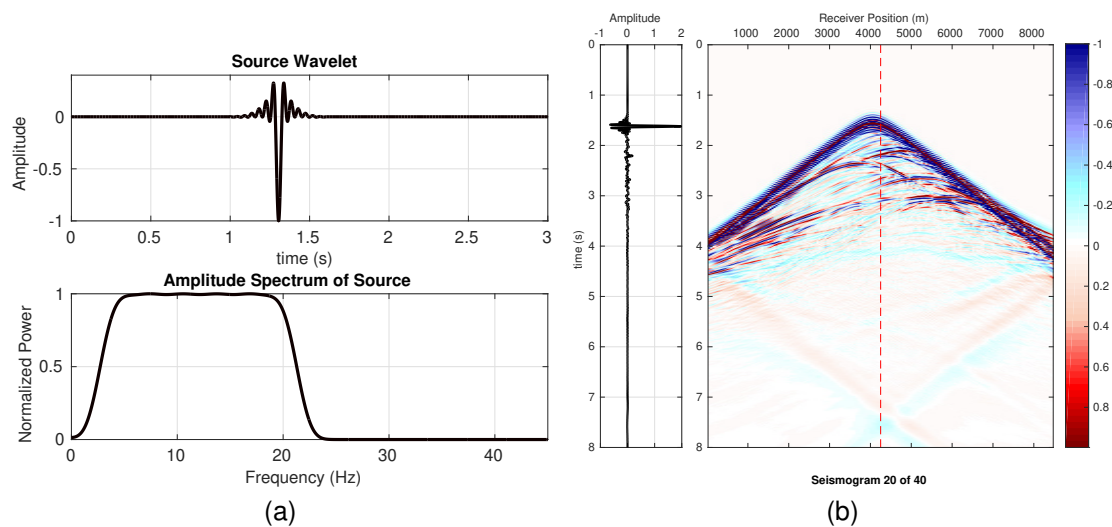
Parâmetro	Valor
Número de fontes	40
Espaçamento entre fontes	200 m
Número de receptores	169
Espaçamento entre receptores	50 m
Tempo de registro	8,0 s
Amostragem	$4,0 \cdot 10^{-4}$ s
Fonte sísmica	<i>wavelet</i> Bspline
Banda de frequência	2 - 22 Hz

4.5.1 Modelo Marmousi

Dados observados

Os dados observados foram gerados, sinteticamente, usando uma *wavelet* com um amplo espectro de amplitude (Figura 65a), simulando as aquisições sísmicas mais recentes que registram dados com mais qualidade. A *wavelet* utilizada foi calculada através da Equação A.58, apresentada no apêndice. Os parâmetros utilizados foram $m = 20$, $p = 2.0$, $q = 22$ e $f_b = 20$. A Figura 65b apresenta um exemplo de dados gerados sinteticamente, utilizando o modelo Marmousi (Figura 18). Os parâmetros de aquisição são apresentados na Tabela 6. As posições das fontes e dos receptores estão indicadas em vermelho e branco, respectivamente, no modelo estimado (Figura 68a).

Figura 64 – a) *Wavelet* wide Band B-spline (CAO; HAN, 2011), usada para gerar os dados observados. b) Sismograma sintético gerado usando a *wavelet* BSpline.

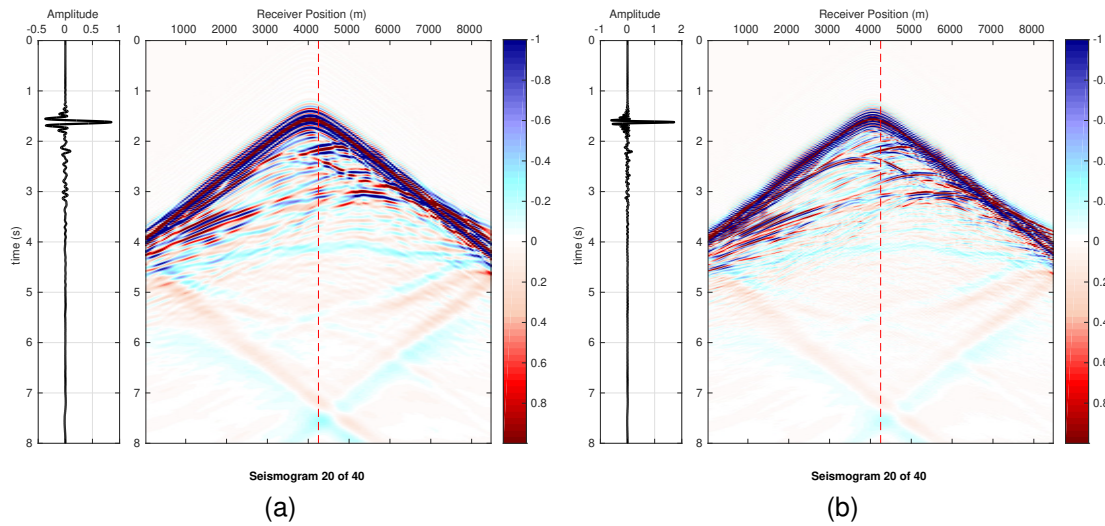


Filtragem dos dados observados

Com o objetivo de aplicar a FWI no domínio no tempo, os dados foram decompostos em diferentes faixas de frequência para a aplicação da estratégia multiescala, e assim, mitigar os problemas de salto de ciclos. Nesse experimento os dados observados foram filtrados com um filtro passa baixa trapezoidal de fase zero (*sufilter* (STOCKWELL, 2002)) nas faixas de frequências de 0-12 Hz, 0-15 Hz, 0-18 Hz e 0-21 Hz para a aplicação da

FWI. Nas Figuras 66a e 66b são apresentados dois exemplos de sismogramas filtrados, com as frequências de corte de 12 Hz e 21 Hz, respectivamente.

Figura 65 – Filtro passa baixa aplicado ao dado observado com frequência de corte de a) 12 Hz e b) 21 Hz. A linha pontilhada indica a posição do traço apresentado a esquerda.

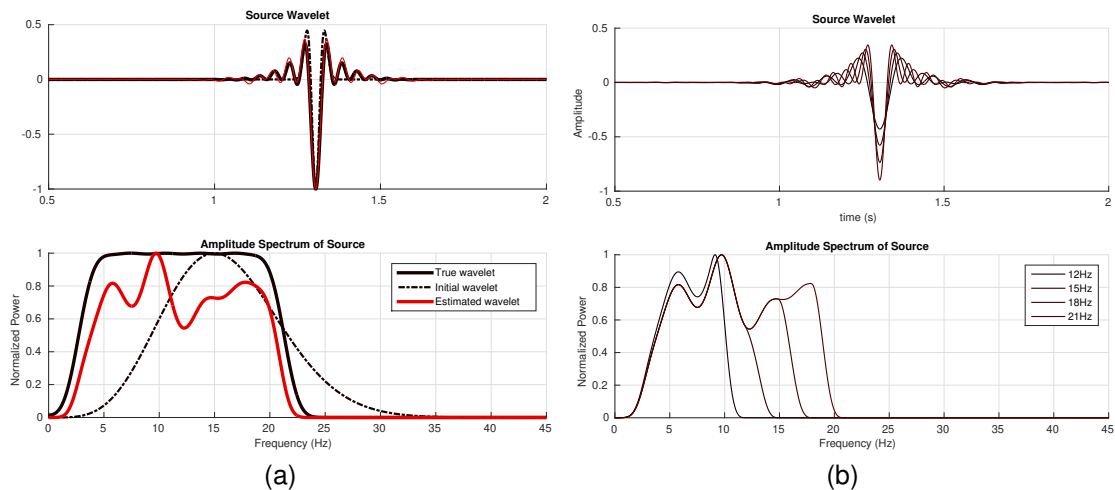


Wavelet estimada

Para a estimativa da wavelet foram utilizados 7 sismogramas, os tiros 5, 10, 15, 20, 25, 30 e 35, e selecionado os traços mais próximos da fonte. A *wavelet* estimada é dada pela média das *wavelets* estimadas em cada sismograma selecionado. O dado calculado inicial foi obtido através da equação da onda 1D, utilizando uma *wavelet* Ricker com frequência de corte de 45Hz como fonte. Foi usado um modelo de velocidade homogêneo com $v_p = 1500$ m/s.

Apenas a primeira chegada do traço mais próximo da fonte foi utilizada na estimativa da *wavelet*. A *wavelet* inicial é apresentada na Figura 67a, a *wavelet* verdadeira e a *wavelet* estimada pela Equação de otimização 2.108, também são apresentadas na Figura 67a. A *wavelet* estimada foi filtrada nas faixas de frequências utilizadas no algoritmo de FWI e estão apresentadas na Figura 67b. Foram utilizadas as frequências de corte de 12, 15, 18 e 21 Hz. Vale ressaltar que o mesmo filtro utilizado para filtrar os dados observados (Figura 65) foi aplicado para a filtrar a *wavelet* estimada.

Figura 66 – a) Comparação entre a *wavelet* inicial, *wavelet* fonte e *wavelet* estimada. b) Decomposição da wavelet estimada nas faixas de frequências utilizadas na inversão.

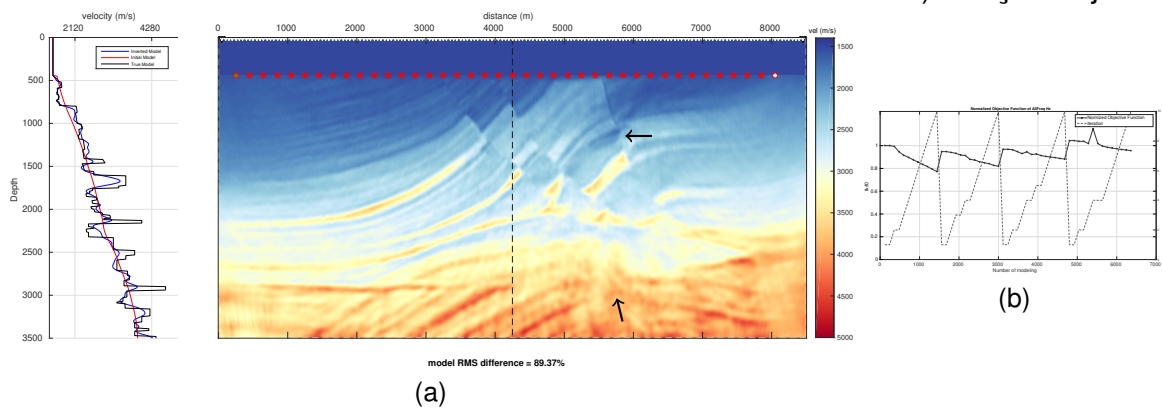


Inversão modelo Marmousi com a *wavelet* estimada

A *wavelet* estimada (Figura 66) decomposta nas faixas de frequências citadas acima, foi utilizada para a estimativa do modelo de velocidade, através do método gradiente descendente combinado com a norma L1. A diferença relativa RMS obtida para esse caso foi de 89,37%.

As incertezas provocadas pela introdução da *wavelet* estimada dificultaram a convergência do modelo para um mínimo mais próximo do mínimo global. Apesar da função objetivo não convergir como nos casos em que a fonte era conhecida, as principais feições do modelo Marmousi podem ser identificadas no modelo estimado da Figura 68a

Figura 67 – a) Modelo de velocidade estimado pelo método gradiente descendente utilizando a *wavelet* estimada. A norma L1 foi utilizada. b) Função objetivo.



4.5.2 Modelo Búzios

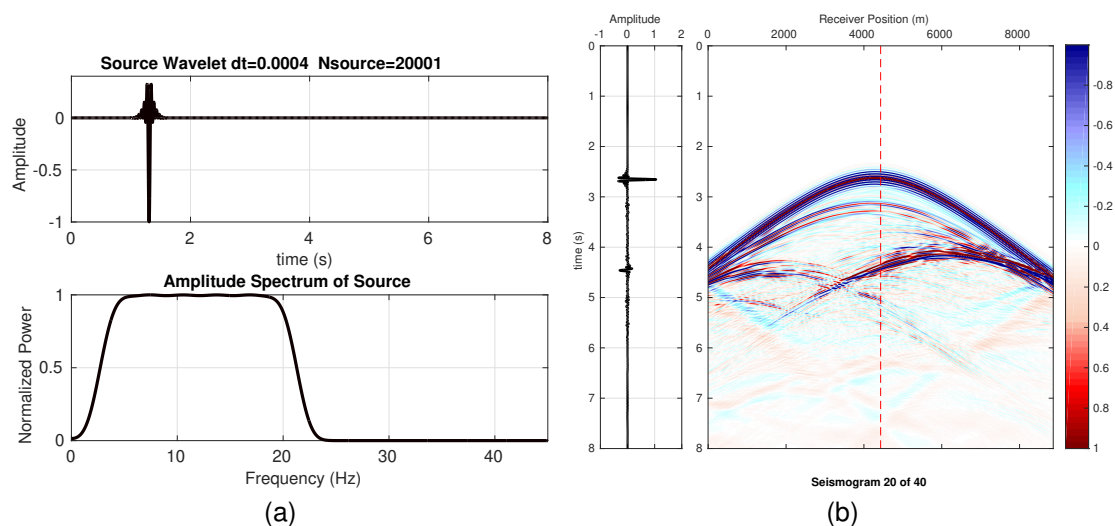
Dados observados

Os dados observados foram gerados com os mesmos parâmetros da *wavelet* BSpline que o exemplo anterior. A Figura 68 apresenta a *wavelet* utilizadas nesse experimento e a Figura 69b apresenta um exemplo de sismograma observado. A posição das fontes e dos receptores estão indicadas no modelo estimado (Figura 72a) em vermelho e branco, respectivamente. A Tabela 7 apresenta os principais parâmetros de aquisição.

Tabela 7 – Parâmetros de aquisição utilizados nos exemplos dessa seção

Parâmetro	Valor
Número de fontes	40
Espaçamento entre fontes	200 m
Número de receptores	887
Espaçamento entre receptores	10 m
Tempo de registro	8,0 s
Amostragem	$4,0 \cdot 10^{-4}$ s
Fonte sísmica	<i>wavelet Bspline</i>
Banda de frequência	2 - 22 Hz

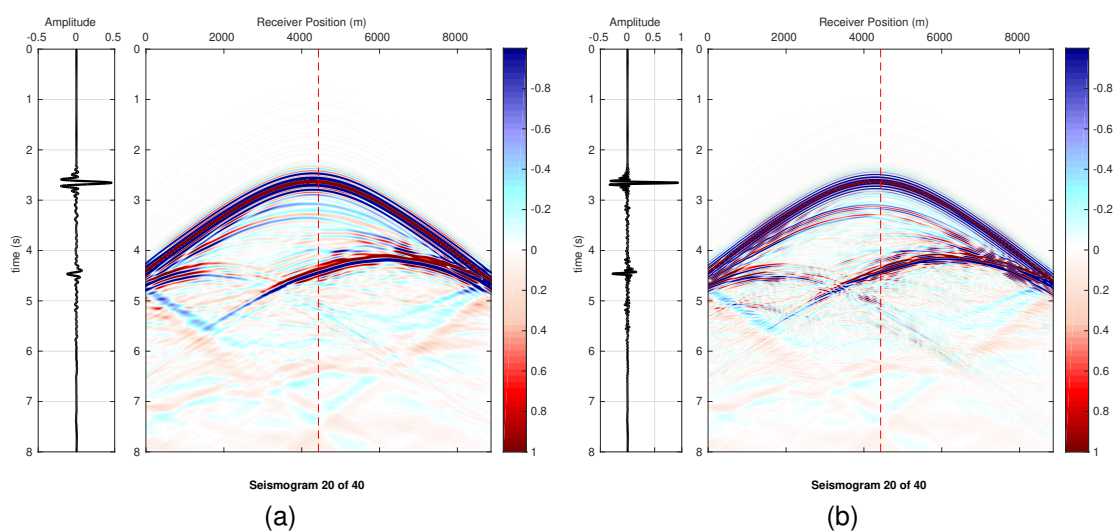
Figura 68 – a) *Wavelet* BSpline utilizada para a geração dos dados observados. b) Exemplo de sismograma sintético usando a *wavelet* BSpline.



Filtragem dos dados observados

As Figuras 70a e 70b apresentam dois exemplos de sismogramas observados filtrados, com as frequências de corte de 12 Hz e 21 Hz, respectivamente. Os dados observados foram filtrados nas seguintes bandas de frequência para a aplicação da abordagem multiescala na aplicação da FWI: 0-6 Hz, 0-9 Hz, 0-12 Hz, 0-15 Hz, 0-18 Hz, 0-21 Hz, 0-24 Hz, 0-27 Hz e 0-30 Hz.

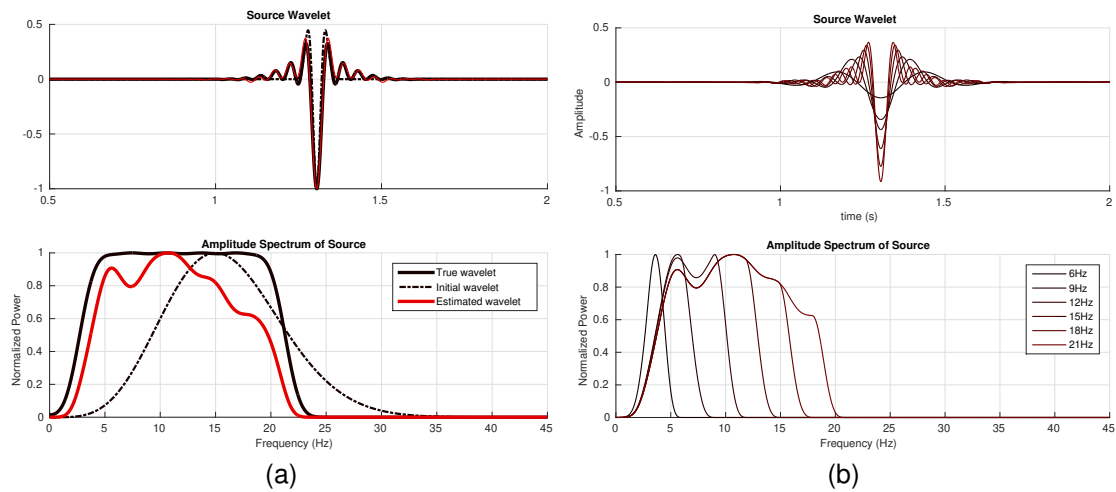
Figura 69 – Exemplos de dois sismogramas observados filtrados com frequência de corte de a) 12 Hz e b) 21 Hz.



Wavelet estimada

Da mesma forma que o exemplo anterior, para a estimativa da *wavelet* fonte somente foram utilizados os traços mais próximos da fonte dos tiros 5, 10, 15, 20, 25, 30 e 35. A *wavelet* estimada é apresentada na Figura 71a, assim como, a *wavelet* verdadeira e a *wavelet* inicial. A Figura 71b apresenta a *wavelet* estimada decomposta nas faixas de frequência utilizada na FWI.

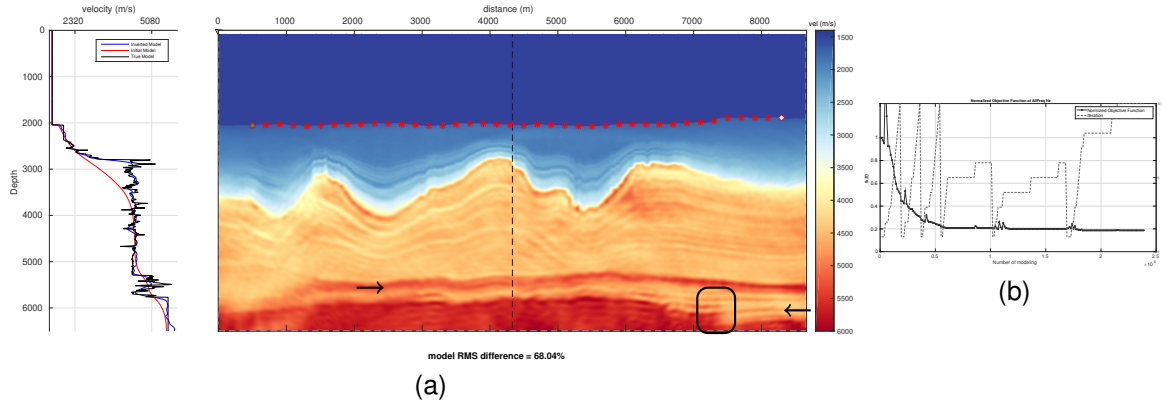
Figura 70 – a) Comparação entre a *wavelet* inicial, *wavelet* fonte e *wavelet* estimada. b) Decomposição da *wavelet* estimada nas faixas de frequências utilizadas na inversão.



Inversão modelo Búzios com a wavelet estimada

A *wavelet* estimada decomposta nas faixas de frequência, mencionada anteriormente, foram utilizadas para estimar o modelo de velocidade através do método de otimização gradiente descendente incluindo a norma L1. A diferença RMS relativa alcançada foi de 68,04%. Iniciar a inversão com frequências mais baixas que o caso com o modelo Marmousi fez o método convergir para uma função objetivo 20% do valor inicial. Além disso, o maior conteúdo de frequência, presente nos dados observados deste teste, permitiu o modelo estimado alcançar estruturas geológicas mais bem definidas que os casos utilizando a *wavelet* Ricker.

Figura 71 – a) Modelo de velocidade estimado pelo método gradiente descendente incluindo a norma L1 e utilizando a wavelet estimada. b) Função objetivo.

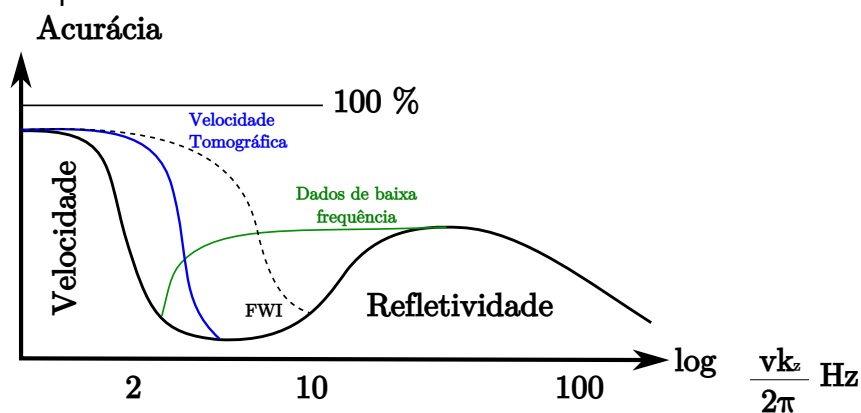


4.6 Os números de onda recuperados pela FWI

Na segunda metade da década de 1980, Claerbout (1985) sugeriu a existência de uma lacuna de informação que está intrinsecamente relacionada ao método sísmico, o que limita a confiabilidade da informação que pode ser recuperada pelo método sísmico, nos comprimentos de onda intermediários (Figura 72). Ainda segundo Claerbout (1985), a velocidade e a refletividade podem ser vistas como duas janelas espectrais que não se sobrepõem. A refletividade compõe as altas frequências (ou número de onda) enquanto a velocidade compõe as baixas frequências (ou número de onda).

Biondi e Almomin (2013) explicam que a identificação do gap de informação entre os baixos números de onda e altos números de onda leva a uma estratégia natural para o imageamento sísmico. As velocidades devem ser estimadas por técnicas tomográficas em primeiro lugar, e em seguida, o modelo estimado é usado como entrada da migração para imagear a refletividade. A Figura 72 indica com a linha azul que as técnicas tomográficas permitem a expansão dos baixos números de ondas, enquanto a linha verde indica que os avanços tecnológicos nas aquisições sísmica permitiram o registro de dados de baixa frequência, como por exemplo aquisições com boa relação sinal ruído até a frequência de 1,5 Hz (BAETEN et al., 2013).

Figura 72 – Confiabilidade da informação obtida a partir de medidas sísmicas na superfície. O imageamento sísmico pode ser visto através de duas bandas de frequências (ou números de ondas) distintas, o modelo de velocidade nas baixas frequências e a imagem migrada (refletividade) nas altas frequências. O método sísmico possui uma lacuna de informação nos comprimentos de ondas intermediários.



Fonte: Modificado de Claerbout (1985) e Biondi e Almomin (2013).

Dentro desse contexto, a FWI utiliza as informações da banda da refletividade,

principalmente as baixas frequências, para a construção do modelo de velocidade (linha pontilhada, Figura 72). O sucesso da FWI na construção de modelos de velocidade de alta resolução foi garantida pelo avanço tecnológico no registro de sinais sísmicos com boa relação sinal ruído mesmo em frequências muito baixas. Além disso, o aumento da capacidade de processamento de dados garantiu que os algoritmos de otimização aplicados a problemas em larga escala se tornassem viáveis.

O imageamento sísmico é pouco sensível aos comprimentos de onda intermediários, principalmente por causa das limitações dos *offsets* registrados e da banda de frequência limitada da fonte sísmica (CLAERBOUT, 1985). A FWI contribui para o preenchimento dessa lacuna reconstruindo os comprimentos de ondas intermediários utilizando métodos tomográficos¹ aos dados registrados na banda de frequência da refletividade.

Portanto, compreender e quantificar os números de onda recuperados pela inversão sísmica pode contribuir para o desenvolvimento de estratégias que mitiguem os problemas de convergência da FWI, como por exemplo o salto de ciclo. Soares et al. (2015) realizaram experimentos numéricos para identificar os possíveis comprimentos de onda que são recuperados pela FWI, nesse estudo, foram introduzidas perturbações bidimensionais de diferentes tamanhos no modelo de velocidade da Bacia de Santos e aplicada a FWI para determinar o limite de resolução.

Uma outra forma de quantificar os números de onda (inverso do comprimento de onda) recuperados pela FWI pode ser alcançada através da transformada de Fourier 1D. Nessa estratégia, é aplicada a transformada de Fourier na direção vertical da diferença entre o modelo estimado e o inicial:

$$\Delta \mathbf{m} = \mathbf{m}^* - \mathbf{m}_0. \quad (4.1)$$

Essa diferença é a perturbação introduzida pela FWI, e a aplicação da transformada de Fourier em $\Delta \mathbf{m}$ fornece os números de onda recuperados pela FWI. O mesmo processo aplicado na diferença entre o modelo verdadeiro e o inicial fornece os números de onda de referência. A diferença com o modelo inicial foi necessária para remover o efeito dos baixos números de onda do modelo estimado ou, de referência, e analisar apenas a refletividade recuperada pela FWI. Mais detalhes sobre a recuperação dos

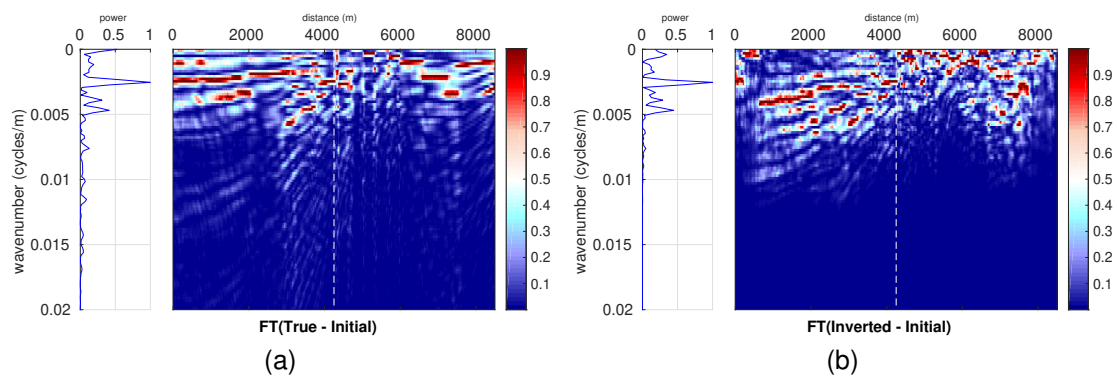
¹ A FWI pode ser entendida como uma tomografia utilizando a equação completa da onda.

números de onda podem ser encontrados no artigo submetido à revista *Journal of Applied Geophysics* no Anexo (COSTA; SANTOS; Soares Filho, 2020).

4.6.1 Modelo Marmousi

Nessa seção serão apresentadas as análises dos números de ondas recuperados pela FWI. Os modelos utilizados para as análises são os modelos obtidos na seção 4.3. Em todos os casos, antes da aplicação da transformada de Fourier, o modelo analisado foi subtraído do modelo inicial. Nesse processo são removidos os baixos números de onda do modelo. Assim, a análise se dá apenas sobre os números de onda recuperados pela FWI. A Figura 74a apresenta os números de onda de referência, obtidos pela diferença entre o modelo inicial (Figura 19) e o modelo de referência (Figura 18) e aplicação da transformada de Fourier na direção vertical.

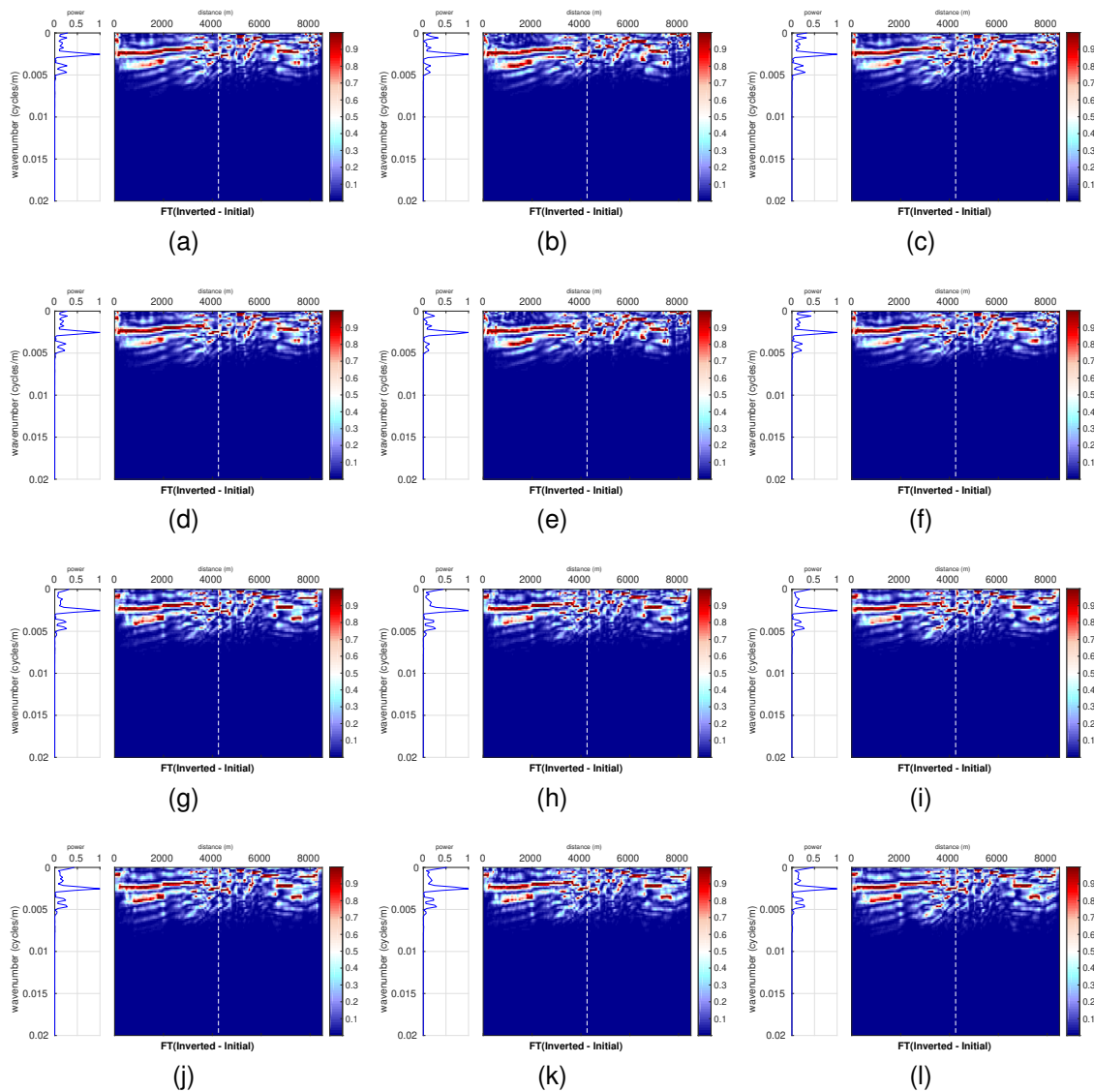
Figura 73 – a) Números de onda de referência. b) Números de onda recuperados pelo método de otimização gradiente descendente combinado com a norma L1 e utilizando a wavelet estimada.



Os números de onda recuperados pelo método gradiente descendente, gradiente conjugado e LBFGS são apresentados na Figura 74 respectivamente, na primeira, segunda e terceira colunas. Na primeira linha estão os resultados para a aplicação da norma L2, na segunda linha os resultados para a aplicação da regularização por *total variation*, a norma L1 para cada método é apresentada na terceira linha, e a combinação entre a norma L1 e regularização por *total variation* está na última linha.

Em geral, os resultados nos quais a norma L1 foi aplicada recuperaram números de ondas mais baixos, o que justifica as menores diferenças RMS relativa para o caso da norma L1. Por outro lado, a aplicação da regularização por *total variation* não contribuiu para mudanças significativas nos números de onda recuperados.

Figura 74 – Números de onda recuperados pela FWI utilizando o método gradiente descendente a) d) g) e j); método gradiente conjugado b),e), h) e k); e método L-BFGS c),f), i) e l). A aplicação da norma L2 é apresentada na primeira linha a),b) e c); a aplicação da regularização por *total variation* na segunda linha d), e) e f); Na terceira linha está a aplicação da norma L1 g),h) e i); e a combinação da norma L1 com a regularização está na última linha j), k) e l) .

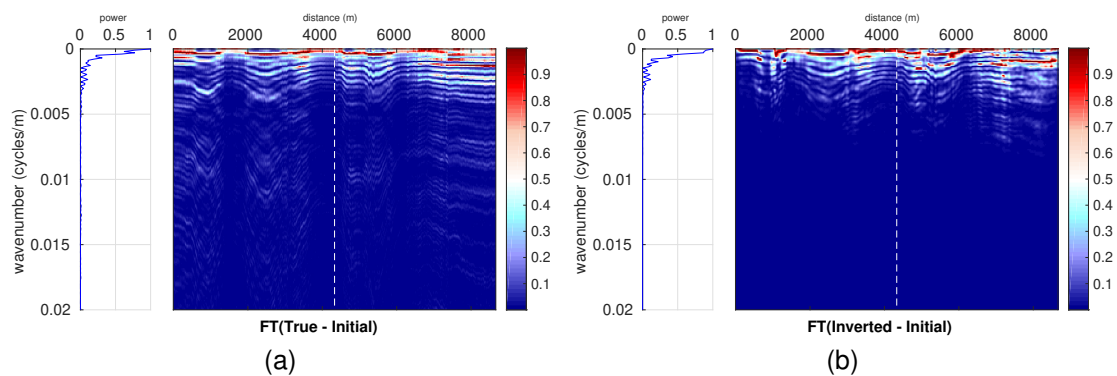


No caso em que foi utilizada a wavelet estimada (Figura 74b) em conjunto com o método gradiente descendente e a norma L1 foram alcançados números de ondas mais altos devido a banda mais larga do dado sísmico. Entretanto, os números de onda mais baixos não foram recuperados corretamente nesse caso, resultando em um modelo com a diferença RMS maior. E além disso, os números de onda mais altos não estão de acordo com os números de onda de referência (Figura 74a)

4.6.2 Modelo Búzios

Nesta seção serão apresentados os números de onda recuperados nos testes de inversão utilizando o modelo Búzios. A Figura 76a apresenta os números de onda de referência, calculados através da transformada de Fourier 1D da diferença entre o modelo de referência e o modelo inicial.

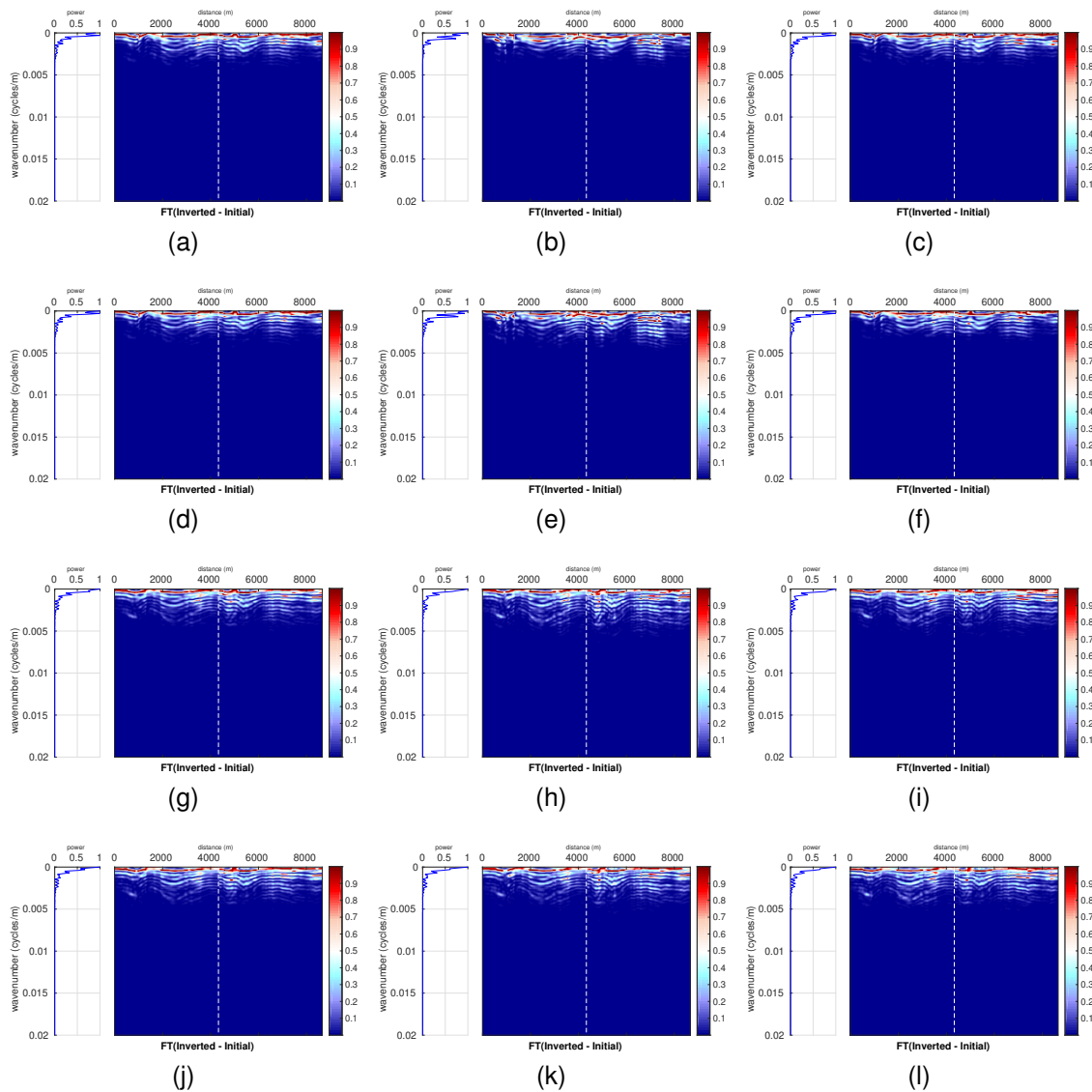
Figura 75 – a) Números de onda de referência. b) Números de onda recuperados utilizando uma *wavelet* fonte estimada e o método de otimização gradiente descendente e a norma L1.



De forma semelhante aos casos com o modelo Marmousi, os números de ondas recuperados utilizando a norma L1 recuperam mais os números de onda mais altos. Como nesse caso foram utilizadas frequências mais baixas durante a abordagem multiescala, todos os testes foram capazes de recuperar os números de onda mais baixos.

O teste utilizando a *wavelet* estimada do dado observado e também o método de otimização gradiente descendente combinada com norma L1 alcançou números de onda ainda maiores que os testes anteriores (Figura 76b). Nesse caso, o espectro de frequência mais amplo dos dados dados e da *wavelet* estimada contribuíram para o aumento da resolução do modelo de velocidade.

Figura 76 – Números de onda recuperados pela FWI utilizando o método gradiente descendente a) d) g) e j); método gradiente conjugado b),e), h) e k); e método L-BFGS c),f), i) e l). A aplicação da norma L2 é apresentada na primeira linha a),b) e c); a aplicação da regularização por *total variation*, na segunda linha d), e) e f); Na terceira linha está a aplicação da norma L1 g),h) e i); e a combinação da norma L1 com a regularização está na última linha j), k) e l) .



4.7 Migração reversa no tempo

A migração sísmica é uma etapa do processamento sísmico que permite a construção da imagem em subsuperfície corrigidas as distorções inerentes ao método sísmico. A seção sísmica obtida através desse método permite imagear as estruturas geológicas em subsuperfície em suas corretas posições em profundidade.

A condição de imagem foi inicialmente formulada por (CLAERBOUT, 1971), onde ele afirma que a um refletor existe onde o campo da fonte e o campo do receptor coincidem em tempo e espaço (Figura 14). Essa condição de imagem pode ser obtida através da correlação cruzada de lag zero entre o campo da fonte e o campo do receptor. O campo da fonte (campo descendente) pode ser calculado através da modelagem direta para cada fonte. Enquanto o campo de receptor (campo ascendente) é calculado através da modelagem reversa no tempo utilizando o sismograma observado como fonte. Esse processo é muito semelhante ao cálculo do gradiente da função objetivo (Equação 2.65).

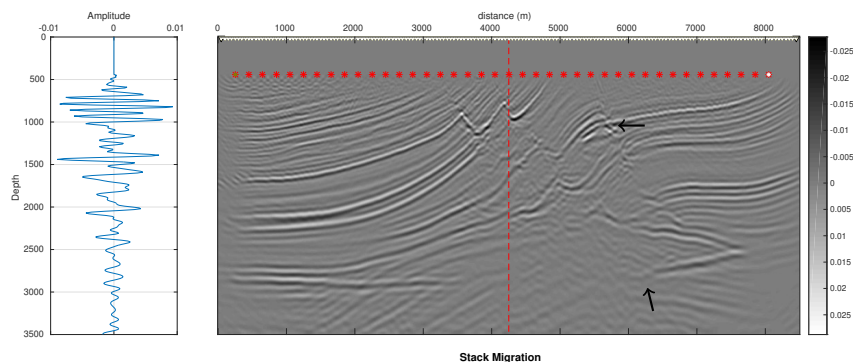
Após a aplicação da condição de imagem é necessário remover os artefatos intrínsecos à migração. As estratégias utilizadas foram o silenciamento dos longos *offsets* e a aplicação do filtro laplaciano. Além disso, a região correspondente a camada de água foi silenciada.

Nessa seção será apresentada a comparação entre as seções sísmicas migradas utilizando os modelos iniciais e estimados para avaliar o impacto do aumento da resolução do modelo de velocidade utilizado na RTM. Os dados observados foram os mesmos utilizados nos testes de inversão que também realizavam a estimativa da *wavelet*. E a geometria de aquisição foi a mesma utilizada nos teste de inversão.

4.7.1 Modelo Marmousi

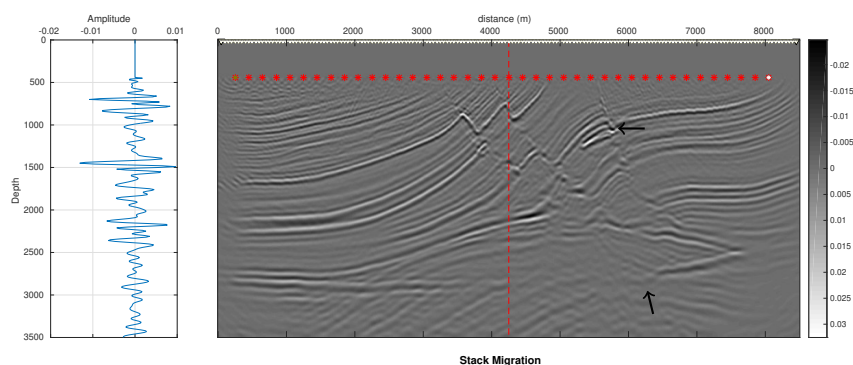
A Figura 77 apresenta seção migrada utilizando o modelo de velocidade inicial (Figura 19). Apesar do modelo de velocidade possuir baixa resolução, e não ser possível identificar as principais estruturas no modelo, a imagem formada foi capaz de identificar diversas camadas e as principais estruturas geológicas. Porém, a imagem da anomalia de baixa velocidade não é bem construída.

Figura 77 – Seção migrada utilizando o modelo de velocidade inicial.



A Figura 78 apresenta a migração utilizando o modelo de velocidade estimado pelo método de otimização gradiente descendente incluindo a norma L1 apresentado na Figura 68a. A melhora na imagem formada devido a melhor qualidade do modelo utilizado na migração pode ser verificada pelo aumento dos detalhes das estruturas geológicas. A anomalia de baixa velocidade e as falhas estão mais bem definidas na nova seção sísmica.

Figura 78 – Seção migrada utilizando o modelo de velocidade estimado pelo método gradiente descendente incluindo a norma L1 no caso em que a *wavelet* foi estimada.

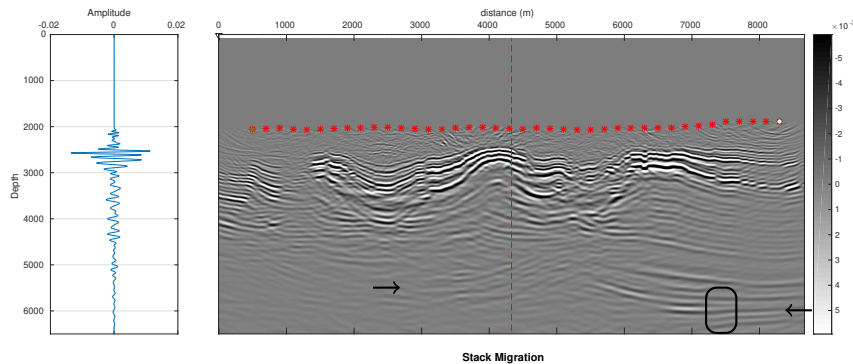


4.7.2 Modelo Búzios

A Figura 79 apresenta a migração utilizando o modelo de velocidade inicial Búzios (Figura 22). Nesse caso, a RTM não foi capaz de fornecer uma imagem com boa qualidade, porém, algumas feições geológicas podem ser reconhecidas, como por

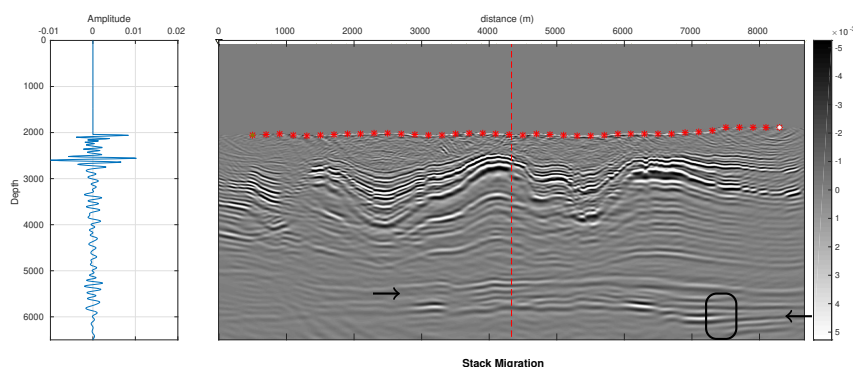
exemplo o topo do sal e os contrastes no pós sal. A dificuldade na construção da imagem nesse caso pode ser associada aos grandes contrastes de velocidade e a baixa resolução do modelo de velocidade inicial.

Figura 79 – Seção migrada utilizando o modelo de velocidade inicial..



Como no caso anterior, a qualidade do modelo de velocidade utilizado na migração tem impacto significativo na qualidade da imagem migrada. A utilização do modelo estimado pelo método gradiente descendente incluindo a norma L1 (Figura 72a) no caso em que a *wavelet* foi estimada permitiu a construção de uma imagem migrada de melhor qualidade (Figura 80). Nesse caso, a base dos corpos salinos foi melhor imageada pela RTM devido a melhora no modelo de velocidade usado na migração. Incluindo a camada do pre-sal que não foi formada com o modelo inicial.

Figura 80 – Seção migrada utilizando o modelo de velocidade estimado pelo método gradiente descendente incluindo a norma L1 no caso em que a *wavelet* foi estimada.



5 Discussão

Nesse trabalho, foram utilizadas diferentes formas de avaliar a qualidade dos modelos estimados. Entre as quais está a diferença RMS entre o modelo verdadeiro e o modelo estimado (Equação 3.1). Quanto menor o valor da diferença RMS, mais próximo o modelo estimado está do modelo de referencia. Os modelos de velocidades utilizados para a avaliação dos diferentes testes de inversão foram o modelo Marmousi (Figura 18) e o modelo Búzios (Figura 20).

A parametrização do cálculo do gradiente utilizando a vagarosidade ao quadrado $s(\mathbf{x}) = \frac{1}{v^2(\mathbf{x})}$ fornece uma melhor direção de atualização que o uso da parametrização na velocidade. Essa parametrização foi utilizada em todos os testes realizados. Por outro lado, o termo de regularização por *total variation*, que é adicionado ao gradiente, indica que as interfaces identificadas durante a inversão podem ser ressaltadas dependendo do peso dado ao parâmetro de regularização (Figuras 36b e 40b). Entretanto, os resultados nos quais foram aplicados a regularização não contribuíram para a redução na diferença RMS relativa. Indicando que a escolha do parâmetro de regularização deve ser mais criteriosa.

Apesar dos gradientes utilizando a norma L1 parecerem contaminados com mais artefatos (Figuras 34 e 38) os resultados dos testes de inversão utilizando a norma L1 alcançaram diferenças RMS menores. Além disso, convergem mais rapidamente, pois utilizam menos modelagens.

Os testes de inversão foram realizados no modelo Marmousi para validar os algoritmos desenvolvidos nesse trabalho. Os diferentes métodos de otimização utilizados no modelo Marmousi indicaram que, em todos os casos, os modelos convergiram para soluções que honram as estruturas do modelo de referência, sendo capazes de reconstruir as principais camadas, falhas e dobras. Após a validação dos algoritmos, o modelo de velocidade baseado nas informações do campo de Búzios foi utilizado para avaliar o comportamento dos algoritmos dentro do contexto do pré sal brasileiro.

A Tabela 8 apresenta o resumo dos resultados dos modelos estimados utilizando

o modelo Marmousi como referência na Seção 4.3. Em todos os casos, as anomalias de baixa velocidades foram mapeadas pelos modelos estimados, indicadas pelas setas pretas em cada modelo. Porém, a anomalia na parte mais profunda nem sempre é visível.

Tabela 8 – Tabela com resumo dos resultados das inversões do modelo Marmousi

Método de otimização	Norma e parâmetro de regularização	nº de modelagens	Diferença relativa
	Modelo Inicial	-	100
Método gradiente descendente	L2 e $\lambda_R = 0,00$	13500	71,39
	L2 e $\lambda_R = 0,01$	25500	72,12
	L1 e $\lambda_R = 0,00$	6200	68,41
	L1 e $\lambda_R = 0,01$	7000	69,21
Método gradiente Conjugado	L2 e $\lambda_R = 0,00$	12800	81,80
	L2 e $\lambda_R = 0,01$	20000	83,28
	L1 e $\lambda_R = 0,00$	11000	68,49
	L1 e $\lambda_R = 0,01$	11000	68,41
Método LBFGS	L2 e $\lambda_R = 0,00$	32200	70,52
	L2 e $\lambda_R = 0,01$	19200	72,63
	L1 e $\lambda_R = 0,00$	10800	62,97
	L1 e $\lambda_R = 0,01$	11700	64,31
Wavelet Estimada (gradiente descendente)	L1 e $\lambda_R = 0,00$	6300	89,37

O teste que alcançou a menor diferença RMS relativa foi utilizando a norma L1 e o método LBFGS. Porém, o teste com o método gradiente descendente, também com a norma L1, obteve um resultado muito similar utilizando menos modelagens.

Em geral, os testes utilizando o gradiente conjugado obtiveram os piores resultados. Mesmo assim, esse método foi capaz de identificar as principais estruturas geológicas. Nesses casos, a norma L1 também se saiu melhor.

Nos testes realizados com o modelo Búzios, foram utilizadas frequências mais baixas durante a abordagem multiescala. Nos testes iniciais com as mesmas frequências que no caso Marmousi, os resultados não convergiram para soluções com significado geológico. Pois eram necessários números de ondas mais baixos para, a partir do modelo inicial, convergir para uma solução próxima ao modelo de referência.

A Tabela 9 resume os resultados obtidos nos diferentes testes de inversão utilizando o modelo Búzios. Em todos os testes, utilizando o modelo Búzios, as principais estruturas foram recuperadas pelos diferentes métodos de otimização. As setas em todos os modelos apontam que a base do sal e as rochas reservatórios,

entre a base do sal e a formação Camboriú, estão mais bem definidas. A falha destacada também é parcialmente recuperada. Além disso, as intercalações na seção pós-sal e no interior do sal foram recuperadas parcialmente pela FWI, indicando que o métodos aplicados são capazes de superar os desafios presentes no contexto da Bacia de Santos.

Tabela 9 – Tabela com resumo dos resultados das inversões do modelo Buzios

Método de otimização	Norma e parâmetro de regularização	nº de modelagens	Diferença relativa
	Modelo Inicial	-	100
Método gradiente descendente	L2 e $\lambda_R = 0,00$	26000	67,52
	L2 e $\lambda_R = 0,01$	37000	67,52
	L1 e $\lambda_R = 0,00$	8900	62,46
	L1 e $\lambda_R = 0,01$	9000	60,86
Método gradiente Conjugado	L2 e $\lambda_R = 0,00$	26000	78,98
	L2 e $\lambda_R = 0,01$	34900	75,60
	L1 e $\lambda_R = 0,00$	13800	62,97
	L1 e $\lambda_R = 0,01$	14800	61,83
Método LBFGS	L2 e $\lambda_R = 0,00$	34900	68,53
	L2 e $\lambda_R = 0,01$	32500	71,58
	L1 e $\lambda_R = 0,00$	15900	56,68
	L1 e $\lambda_R = 0,01$	24800	58,56
Wavelet Estimada (gradiente descendente)	L1 e $\lambda_R = 0,00$	24500	68,04

Novamente, o teste que alcançou a menor diferença RMS relativa foi o método LBFGS em conjunto com a norma L1. Porém, o método gradiente descendente incluindo a norma L1 atingiu um resultado muito similar com menos iterações. Vale ressaltar que em todos os casos a função objetivo foi reduzida para menos de 20% do valor inicial.

Como nos testes com o modelo Marmousi, o método gradiente conjugado não alcançou os valores mais baixos de diferença RMS relativa. Mas mesmo assim os modelos estimados foram capazes de recuperar parte das intercalações no sal e detectar as variações de velocidade no pré-sal.

Os testes de inversão utilizando a *wavelet* fonte estimada a partir dos dados observados utilizou o método gradiente descendente em conjunto com a norma L1, nos modelos Marmousi e Búzios, pois foi o método com melhor performance. No caso Marmousi, claramente houve um problema de convergência como aponta a função objetivo (Figura 68b). Esse problema de convergência pode ser explicado pelas

incertezas adicionada a *wavelet* estimada e também pela falta dos números de onda mais baixos no modelo inicial.

Para contornar esse problema nos testes com o modelo Búzios, foram utilizadas frequências mais baixas na abordagem multiescala. Apesar do modelo inicial não possuir os números de ondas mais baixos, eles podem ser recuperados a partir dos dados nas frequências mais baixas. Dessa maneira, o resultado do teste com o modelo Búzios utilizando a *wavelet* estimada forneceu um resultado tão bom quanto os testes utilizando a *wavelet* Ricker.

A análise dos números de onda recuperados pela FWI, proposto nesse trabalho, forneceu uma forma de avaliar a qualidade dos modelos estimados pelos diferentes métodos de otimização. Esse método permite uma comparação direta entre os modelos estimados, e determinar qual deles alcançou números de onda mais coerentes com o modelo de referência.

Essa análise corrobora que a utilização de um espectro de amplitude largo permite a recuperação de números de onda mais altos, isso significa que, o modelo estimado possui maior resolução. Esse efeito pode ser observado quando comparados os testes utilizando a *wavelet* estimada e a *wavelet* Ricker (Figuras 73 e 74; Figuras 75 e 76). A análise dos números de onda mostra que o teste com a *wavelet* estimada e o modelo Marmousi obteve um pior resultado da diferença RMS relativa pois não foi capaz de recuperar os números de onda mais baixos (Figura 74b) ao contrário dos testes com a *wavelet* Ricker (Figura 74)

Por fim, os modelos estimados pelo método gradiente descendente, combinado com a norma L1 e usando a *wavelet* estimada como fonte (Figuras 68a e 72a) foram utilizados como modelos de velocidade de migração. Em ambos os caso, houve uma melhora significativa nas imagens migradas.

No modelo Marmousi, os refletores ficaram mais bem definidos principalmente nas regiões onde há maior variação lateral de velocidade, assim como as falhas e dobras geológicas do modelo. A anomalia de baixa velocidade também foi melhor construída na imagem migrada com o modelo estimado. A migração aplicada ao modelo inicial de Búzios não construiu uma imagem tão bem definida quanto no caso do modelo Marmousi. Essa dificuldade pode ser atribuída aos grandes contrastes de impedância e as intercalações muito finas presentes no modelo, e também as grande profundidades

dos alvos.

Por outro lado, a utilização do modelo de velocidade estimado pela FWI foi capaz de contribuir para a formação da imagem das camadas do pré-sal, como indicam as setas na Figura 80. Isso mostra que a recuperação dos números de onda pela FWI tem potencial para melhorar as imagens migradas pelo método RTM, principalmente nas regiões abaixo da camada de sal.

6 Conclusão

Os diversos testes realizados nesse trabalho mostraram que os diferentes métodos de inversão são capazes de recuperar modelos próximos dos modelos de referência utilizados. A combinação entre as parametrizações do gradiente no inverso da vagarosidade e a aplicação da norma L1 fizeram o método gradiente descendente ser o mais eficiente. Que alcançou uma diferença RMS tão baixa quanto o método LBFGS, porém, utilizou menos modelagens.

De forma geral, a regularização por *total variation* não forneceu melhoras aos resultados em que foi aplicada. A regularização é muito sensível ao parâmetro de regularização, que neste trabalho foi assumido constante para todos os testes $\lambda_R = 0.01$. Melhores resultados podem ser obtidos escolhendo o parâmetro de regularização de forma que sempre garanta uma redução da função objetivo. Um algoritmo de busca linear pode ajudar na determinação do parâmetro de regularização em cada iteração, porém pode aumentar significativamente o custo computacional, por exigir mais avaliações da função objetivo.

A introdução da estimativa da *wavelet* aos testes de inversão, tornou os experimentos sintéticos mais próximos aos problemas utilizando dados reais. Uma vez que a determinação da *wavelet* utilizada como fonte tem impacto significativo sobre os modelos estimados, e geralmente, os testes sintéticos não levam em consideração seu efeito. Nos experimentos numéricos realizados, a wavelet estimada, também pelo método adjunto, foi capaz de contribuir para a estimativa de modelos de velocidade próximos ao modelo de referência.

A análise dos números de onda recuperados pela FWI (artigo em anexo) permitiu avaliar a qualidade dos modelos estimado, consequentemente, definindo uma forma de determinar qual o método de otimização fornece o melhor resultado. Esse método, confirmou que os testes utilizando uma wavelet de banda larga são capazes de recuperar maiores números de ondas quando comparados a uma *wavelet* Ricker com frequência de corte similar.

Como proposta de trabalhos futuros, estão a natural implementação dos algoritmos

de modelagem e inversão para o caso 3D. Além disso, a utilização de algoritmos de modelagem da equação da onda elástica e viscoelástica, para avaliar os efeitos das conversões de energia (no caso elástico) e da atenuação (viscoelástico) no algoritmo de inversão acústica.

A utilização de diferentes normas na função objetivo, também são tópicos interessantes que podem ser abordados em trabalhos futuros. Alguns exemplos foram citados nesse trabalho, além da norma L2 e L1, como as normas por correlação cruzada, as normas que utilizam a função envelope e fase instantânea, ou ainda, normas que utilizam um fator de amortecimento em função do tempo de registro. Outro tópico de interesse, é a aplicação de diferentes abordagens multiescala para mitigar os problemas de salto de ciclo. Como por exemplo, a aplicação de diferentes tempos de registro (*time windowing*), a aplicação da inversão privilegiando inicialmente os longos afastamentos fonte-receptor e incluir gradualmente os afastamentos mais curtos ou a inversão das camadas mais rasas aumentando a profundidade gradualmente.

Referências

- AKI, K.; RICHARDS, P. G. *Quantitative seismology*. [S.l.: s.n.], 2002. 700 p.
- ALKHALIFAH, T. *Full waveform inversion in an anisotropic world Where are the parameters hiding?* [S.l.]: EAGE, 2014.
- AMERY, G. B. Basics of seismic velocities. *The Leading Edge*, v. 12, n. 11, p. 1087–1091, 1993.
- ANAGAW, A. Y.; SACCHI, M. D. Full Waveform Inversion with Total Variation Regularization. In: *Recovery-CSPG CSEG CWLS Convention*. [S.l.: s.n.], 2011. p. 1–4.
- ANP. *Plano de Desenvolvimento - Campo de Búzios*. [S.l.], 2016.
- ASTER, R.; BORCHERS, B.; THURBER, C. *Parameter identification and inverse problems*. [S.l.: s.n.], 2013. v. 40. 377 p.
- BAETEN, G.; MAAG, J. W. de; PLESSIX, R. É.; KLAASSEN, R.; QURESHI, T.; KLEEMEYER, M.; KROODE, F. T.; RUJIE, Z. The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study. *Geophysical Prospecting*, v. 61, n. 4, p. 701–711, 2013.
- BARTOLO, L. d. *Introdução à Modelagem Sísmica utilizando o MDF*. 2013.
- BIONDI, B.; ALMOMIN, A. Tomographic full-waveform inversion (TFWI) by combining FWI and wave-equation migration velocity analysis. *The Leading Edge*, v. 32, n. 9, p. 1074–1080, 2013.
- BIZZI, F.; GALUZZI, B.; TOGNARELLI, A.; STUCCHI, E.; MAZZOTTI, A. Experience of acoustic FWI on seismic land data. In: *35-esimo Convegno Nazionale GNGTS*. [S.l.: s.n.], 2016. p. 495–498.
- BOHLEN, T. Parallel 3-D viscoelastic finite difference seismic modelling. *Computers and Geosciences*, v. 28, n. 8, p. 887–899, 2002.
- BOYD, S.; VANDENBERGHE, L. *Convex optimization theory*. [S.l.]: Cambridge University Press, 2004. v. 25. 487–487 p.
- BROSSIER, R.; OPERTO, S.; VIRIEUX, J. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, v. 74, n. 6, p. WCC105–WCC118, 2009.
- BULCÃO, A. *Modelagem e migração reversa no tempo empregando operadores elásticos e acústicos*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2004.
- BULCÃO, A.; Soares Filho, D. M.; LOUREIRO, F. P.; FARIAS, F. F.; SANTOS, L. A. Full Waveform Inversion: Comparison between Different Objective Functionals. In: *Thirteenth International Congress of the Brazilian Geophysical Society*. [S.l.: s.n.], 2014. p. 312–316.

- BULHÕES, É. M.; AMORIM, W. N. d. Princípio da SismoCamada Elementar e sua aplicação à Técnica Volume de Amplitudes (tecVA). *9th International Congress of the Brazilian Geophysical Society*, 2005.
- BUNKS, C.; SALECK, F. M.; ZALESKI, S.; CHAVENT, G. Multiscale seismic waveform inversion. *Geophysics*, Society of Exploration Geophysicists, v. 60, n. 5, p. 1457–1473, 1995.
- BYRD, R. H.; LU, P.; NOCEDAL, J.; ZHU, C. *A Limited Memory Algorithm for Bound Constrained Optimization*. [S.l.], 1994. 1–24 p.
- CAO, S.; HAN, D. Wide-band b-spline wavelet with four parameters. *SEG Technical Program Expanded Abstracts 2011*, p. 3840–3844, 2011.
- CARCIONE, J. M.; HERMAN, G. C.; KROODE, A. P. E. ten. Y2K Review Article: Seismic modeling. *Geophysics*, v. 67, n. 4, p. 1304–1325, 2002.
- CARNEIRO, M. d. S. R.; PEREIRA-DIAS, B.; Soares Filho, D. M.; LANDAU, L. On the Scaling of the Update Direction for Multi-parameter Full Waveform Inversion: Applications to 2D Acoustic and Elastic Cases. *Pure and Applied Geophysics*, v. 175, n. 1, p. 217–241, 2018.
- CARNEIRO, M. S. *Estratégias de Inversão de Multiparâmetros Utilizando a Equação Completa da Onda Elástica*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2017.
- CERJAN, C.; KOSLOFF, D.; KOSLOFF, R.; RESHEF, M. A nonreflecting boundary condition for discrete acoustic and elastic wave equations. v. 50, n. 4, p. 705–708, 1985.
- CHEN, G.; CHEN, S.; WU, R.-S. Full Waveform Inversion in time domain using time-damping filters. n. August 2018, p. 1451–1455, 2015.
- CHI, B.; DONG, L.; LIU, Y. Full waveform inversion method using envelope objective function without low frequency data. *Journal of Applied Geophysics*, Elsevier B.V., v. 109, p. 36–46, 2014.
- CHONG, E. K.; ZAK, S. H. *An introduction to optimization*. [S.l.: s.n.], 2008.
- CLAERBOUT, J. F. Toward a unified theory of reflector mapping. *Geophysics*, Society of Exploration Geophysicists, v. 36, n. 3, p. 467–481, 1971.
- CLAERBOUT, J. F. *Imaging the earth's interior*. [S.l.]: Blackwell scientific publications Oxford, 1985. v. 1.
- COSTA, F. T.; KARSOU, A. A.; SILVA, R. C. M. da; MESQUITA, F. C. D.; SANTOS, M. A. C.; MOREIRA, R. M.; Soares Filho, D. M. Inversão do campo de onda completo (FWI) ao modelo conceitual do Campo de Búzios. In: *Sixteenth International Congress of The Brazilian Geophysical Society*. [S.l.: s.n.], 2019.
- COSTA, F. T. da; SANTOS, M. A. C.; Soares Filho, D. M. Wavenumbers illuminated by time-domain acoustic FWI using the L1 and L2 norms. *Journal of Applied Geophysics*, Elsevier B.V., v. 174, 2020.

- DAI, Y. H.; YUAN, Y. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, v. 10, n. 1, p. 177–182, 1999.
- DJIKPESSE, H.; TARANTOLA, A. Multiparameter l_1 norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms. *Geophysics*, v. 64, n. 4, p. 1023–1035, 1999.
- ENGQUIST, B.; MAJDA, A. Absorbing Boundary Conditions for the Numerical Simulation of Waves. *Mathematics of Computation*, v. 31, n. 139, p. 629–651, 1977.
- ESSER, E.; GUASCH, L.; HERRMANN, F. J.; WARNER, M. Constrained waveform inversion for automatic salt flooding. *The Leading Edge*, v. 35, n. 3, p. 235–239, 2016.
- ESSER, E.; LEEUWEN, T. V.; ARAVKIN, A. Y.; HERRMANN, F. J.; SCIENCES, O.; WISKUNDE, C.; IBM, T. J. A scaled gradient projection method for total variation regularized full waveform inversion. In: *SINBAD Spring consortium talks*. [S.l.]: SINBAD, 2014.
- FEYNMAN, R. P.; LEIGHTON; SANDS; LEIGHTON, R. B.; SANDS, M. *The Feynman Lectures on Physics - Volume 1*. [S.l.: s.n.], 1965. v. 33. 750 p.
- FICHTNER, A. *Full Seismic Waveform Modelling and Inversion*. [S.l.: s.n.], 2011. 343 p.
- FICHTNER, A.; KENNETT, B. L.; IGEL, H.; BUNGE, H. P. Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. *Geophysical Journal International*, v. 175, n. 2, p. 665–685, 2008.
- GEISLINGER, A.; OLIVEIRA, L.; PENNA, R.; MATOS, M.; SANSONOWSKI, R.; ROSSETO, J.; ARAÚJO, S. Carbonate and igneous rock characterization through reprocessing, FWI imaging, and elastic inversion of a legacy seismic data set in Brazilian presalt province. *The Leading Edge*, v. 38, n. 1, p. 11–19, 2019.
- HUANG, Y.; SCHUSTER, G. T. Resolution limits for wave equation imaging. *Journal of Applied Geophysics*, Elsevier, v. 107, p. 137–148, 2014.
- IKELLE, L. T.; AMUNDSEN, L. *Introduction to Petroleum Seismology*. [S.l.: s.n.], 2013. v. 53. 1689–1699 p.
- JAEGER, J.; COOK, N.; ZIMMERMAN, R. *Fundamentals of Rock Mechanics*. [S.l.]: Wiley, 2007.
- JIMÉNEZ TEJERO, C. E.; DAGNINO, D.; SALLARÈS, V.; RANERO, C. R. Comparative study of objective functions to overcome noise and bandwidth limitations in full waveform inversion. *Geophysical Journal International*, v. 203, n. 1, p. 632–645, 2015.
- JONES, I. *An Introduction to: Velocity Model Building*. [S.l.]: EAGE, 2014.
- KORMANN, J.; RODR, J. E.; GUTIERREZ, N.; PUENTE, J. D.; HANZICH, M. Using Power-Model Based Preconditioners for 3D Acoustic Full Waveform Inversion. In: *SEG Houston 2013 Annual Meeting*. [S.l.: s.n.], 2013. p. 1105–1109.
- LAILLY, P. The Seismic Inverse Problem as a Sequence of Before Stack Migrations. *Conference on Inverse Scattering, Theory and Applications, Society for Industrial and Applied Mathematics*, p. 206–220, 1983.

- LATIMER, R. B.; DAVIDSON, R.; RIEL, P. van. An interpreter's guide to understanding and working with seismic-derived acoustic impedance data. *The Leading Edge*, v. 19, n. 3, p. 242–256, 2002.
- LEVANDER, A. Finite-difference forward modeling in seismology. *Encyclopedia of Solid Earth Geophysics*, p. 410–431, 01 1989.
- LINES, L. R.; SLAWINSKI, R.; BORDING, R. P. A recipe for stability of finite-difference wave-equation computations. *Geophysics*, v. 64, n. 3, p. 967–969, 1999. ISSN 00168033.
- LOWRIE, W. *Fundamentals of Geophysics*. [S.l.]: Cambridge University Press, 2007.
- MARTIN, G. S.; WILEY, R.; MARFURT, K. J. Marmousi2: An elastic upgrade for marmousi. *The Leading Edge*, v. 25, n. 2, p. 156–166, 2006.
- MATIAS, J. L. d. L. *Análise do Problema de Salto de Ciclo em Inversão de Dados Sísmicos Utilizando a Equação Acústica da Onda e sua Dependência em Relação ao Conteúdo de Frequência*. 2016.
- MENKE, W. *Geophysical data analysis: discrete inverse theory: MATLAB edition*. [S.l.]: Academic press, 2012. v. 45.
- MÉTIVIER, L. *Seiscope optimization toolbox manual*. [S.l.], 2014. 1–46 p.
- MÉTIVIER, L.; BROSSIER, R. The seiscope optimization toolbox: A large-scale nonlinear optimization library based on reverse communicationthe seiscope optimization toolbox. *Geophysics*, GeoScienceWorld, v. 81, n. 2, p. F1–F15, 2016.
- MOCZO, P.; ROBERTSSON, J. O.; EISNER, L. The finite-difference time-domain method for modeling of seismic wave propagation. *Advances in geophysics*, Elsevier, v. 48, p. 421–516, 2007.
- NOCEDAL, J.; WRIGHT, S. *Numerical Optimization*. [S.l.]: Springer New York, 2006.
- PACAL, E. E. *Seismic imaging with ocean-bottom nodes (OBN): new acquisition designs and the atlantis 4c obn survey*. Tese (Doutorado) — University of Houston, 2012.
- PENG, C.; WANG, M.; CHAZALNOEL, N.; GOMES, A. Subsalt imaging improvement possibilities through a combination of FWI and reflection FWI. *The Leading Edge*, v. 37, n. 1, p. 52–57, 2018.
- PLESSIX, R.-E. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, Oxford University Press, v. 167, n. 2, p. 495–503, 2006.
- PRATT, G.; SHIN. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, v. 133, n. 2, p. 341–362, 1998.
- PRATT, R. G. Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, Society of Exploration Geophysicists, v. 64, n. 3, p. 888–901, 1999.

- PUJOL, J. *Elastic wave propagation and generation in seismology*. [S.l.]: Cambridge University Press, 2003.
- PYUN, S.; SON, W.; SHIN, C. Frequency-domain waveform inversion using an l1-norm objective function. *Exploration Geophysics*, n. 1, p. 227–232, 2009.
- QU, S.; VERSCHUUR, D.; CHEN, Y. Full waveform inversion using an automatic directional total variation constraint. In: *79th EAGE Conference and Exhibition 2017*. [S.l.]: EAGE, 2017.
- REYNOLDS, A. C. Boundary Conditions for the Numerical Solution of Wave Propagation Problems. *Geophysics*, v. 43, n. 6, p. 1099–1110, 1978.
- ROSA, A. R. *Análise do Sinal Sísmico*. [S.l.: s.n.], 2010. 668 p.
- RYAN, H. Ricker, Ormsby; Klander, Butterworth -A Choice of wavelets. *CSEG Recorder*, September, p. 24–25, 1994.
- SCALES, J. A.; SMITH, M. L.; TREITEL, S. *Introductory geophysical inverse theory*. [S.l.]: Samizdat Press, 2001. 1–208 p.
- SCHUSTER, G. *Seismic Inversion*. [S.l.]: Society of Exploration Geophysicists, 2017.
- SEARS, T. J.; BARTON, P. J.; SINGH, S. C. Elastic full waveform inversion of multicomponent ocean-bottom cable seismic data: Application to Alba Field, U. K. North Sea. *Geophysics*, v. 75, n. 6, p. R109–R119, 2010.
- SHEARER, P. M. *Introduction to seismology*. [S.l.]: Cambridge University Press, 2009.
- SHERIFF, R. E.; GELDART, L. P. *Exploration Seismology*. [S.l.]: Cambridge University Press, 1995.
- SIRGUE, L. *présentée par Spécialité : Sciences de la Terre Inversion de la forme d ' onde dans le domaine fréquentiel de données sismiques grands offsets*. Tese (Doutorado) — Queen's University, Canada, 2003.
- SLAUGHTER, W. *The Linearized Theory of Elasticity*. [S.l.]: Birkhäuser Boston, 2002.
- SOARES, D. M. F.; SOUZA, A. A. V. B.; BULCÃO, A.; DIAS, B. P.; MARTINS, E. O.; FARIAS, F. F.; LOUREIRO, F. P.; SANTOS, L. A.; SANTOS, M. dos. Full waveform inversion: Investigating resolution on a santos basin model. *14th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, Brazil, 3-6 August 2015*, p. 323–326, 2015.
- SOUZA, R. V. M. *Técnica de Regularização Baseada no IIG para Inversão de Dados Sísmicos*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2018.
- STOCKWELL, J. W. Free software in education: A case study of CWP/SU: Seismic Un*x . *The Leading Edge*, v. 16, n. 7, p. 1045–1050, 2002.
- TARANTOLA. *Inverse Problem Theory and methods for model parameter estimation*. [S.l.: s.n.], 2005.
- TARANTOLA, A. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, v. 49, n. 8, p. 1259–1266, 1984.

- TARANTOLA, A. A strategy for nonlinear elastic inversion of seismic reflection data. *56th Ann. Internat. Mtg.*, v. 51, n. 10, p. Session:S12.4, 1986.
- TEJERO, C. J.; DAGNINO, D.; SALLARES, V.; RANERO, C. Comparison of objective functionals in seismic full waveform inversion. n. June 2014, p. 16–19, 2014.
- VERSTEEG, R. The Marmousi experience: Velocity model determination on a synthetic complex data set. *The Leading Edge*, v. 13, n. 9, p. 927–936, 1994.
- VIRIEUX, J.; BROSSIER, R.; CHALJUB, E.; COUTANT, O.; GARAMBOIS, S.; ETIENNE, V.; MERCERAT, D.; PRIEUX, V.; OPERTO, S.; RIBODETTI, A.; CRUZ-ATIENZA, V.; TAGO, J. *Modelling Seismic Wave Propagation for Geophysical Imaging*. [S.l.: s.n.], 2006. 253–304 p.
- VIRIEUX, J.; OPERTO, S. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, Society of Exploration Geophysicists, v. 74, n. 6, p. WCC1–WCC26, 2009.
- WANG, P.; ZHANG, Z.; MEI, J.; LIN, F.; HUANG, R. Full-waveform inversion for salt : A coming of age. n. March, p. 204–213, 2019.
- WANG, Y. *Seismic Inversion Theory and Applications*. [S.l.]: Wiley, 2016. v. 26.
- WEGLEIN, A. B. A timely and necessary antidote to indirect methods and so-called P-wave FWI. *Leading Edge*, v. 32, n. 10, p. 1192–1204, 2013.
- WELLINGTON, P.; BROSSIER, R.; GARAMBOIS, S.; VIRIEUX, J. 2D Acoustic Time Domain Full Waveform Inversion - A Broadband Application in the Carnarvon Basin , Australia. n. June, p. 1–4, 2015.
- WIRGIN, A. The inverse crime. p. 1–10, 2004.
- WOODWARD, M. J. Wave-equation tomography. *Geophysics*, Society of Exploration Geophysicists, v. 57, n. 1, p. 15–26, 1992.
- WU, R.-S.; LUO, J.; WU, B. Seismic envelope inversion and modulation signal model. *Geophysics*, v. 79, n. 3, p. WA13–WA24, 2014.
- YILMAZ, O. *Seismic Data Analysis*. [S.l.: s.n.], 2000. 1000 p.
- YONG, P.; LIAO, W.; HUANG, J.; LI, Z. Total variation regularization for seismic waveform inversion using an adaptive primal dual hybrid gradient method. *Inverse Problems*, IOP Publishing, v. 34, n. 4, 2018.

Apêndices

APÊNDICE A – A teoria ondulatória

A compreensão dos principais aspectos dos fenômenos ondulatórios é fundamental para a implementação da inversão do campo completo da onda. Neste capítulo será abordada uma revisão sobre a teoria ondulatória. Inicialmente será apresentada a formulação da equação da onda compressional em uma dimensão. Em seguida, é abordada a formulação da equação da onda elástica a partir dos conceitos da mecânica aplicada em meios contínuos. Nesta formulação são utilizadas a equação do movimento e a lei de Hooke. A aproximação para ondas acústicas é apresentada na sequência, ambas as formulações, com densidade variável e com densidade constante. A equação da onda acústica com densidade constante foi utilizada nas implementações numéricas. Por fim, os termos utilizados como fonte sísmica serão apresentados. Foram utilizadas a *wavelet* Ricker e uma *wavelet* conhecida como B-spline.

A.1 A equação da onda 1D

Tendo em vista uma boa compreensão da estratégia FWI para obtenção de um modelo de propriedades mais realístico, o entendimento profundo das leis físicas que regem a propagação das ondas em meios materiais é fundamental. Portanto, nessa seção serão discutidos alguns aspectos da formulação do problema ondulatório.

Inicialmente, leva-se em consideração o problema ondulatório mais simples possível, a propagação da onda em uma dimensão. De acordo com Feynman et al. (1965), são necessárias três suposições para solucionar esse problema:

1. Quando o meio se move a densidade se modifica;
2. A variação da densidade corresponde a uma variação na pressão;
3. As desigualdades na pressão geram movimentos no meio.

Para escrever uma expressão matemática para 2ª suposição, considere a posição $x \in M$, onde M é o meio em que a onda se propaga. É possível assumir que antes da

chegada da onda na posição x a pressão de equilíbrio é dada por p_0 . A densidade ρ_0 na posição x em equilíbrio pode ser relacionada à pressão de equilíbrio por:

$$p_0 = f(\rho_0), \quad (\text{A.1})$$

onde f é uma função contínua e diferenciável. Com a chegada da onda na posição x uma nova pressão p que pode ser relacionada à densidade ρ por:

$$p = f(\rho). \quad (\text{A.2})$$

Tanto a pressão quanto a densidade, durante a perturbação no meio, podem ser escritas em termos da situação de equilíbrio, ou seja:

$$p = p_0 + p_e; \quad \rho = \rho_0 + \rho_e \quad (\text{A.3})$$

onde p_e e ρ_e são, respectivamente, as perturbações na pressão e densidade durante a passagem da frente de onda. Assumindo que as perturbações são muito pequenas quando comparadas a situação de equilíbrio. É possível escrever:

$$p_0 + p_e = f(\rho_0 + \rho_e) = f(\rho_0) + \rho_e f'(\rho_0). \quad (\text{A.4})$$

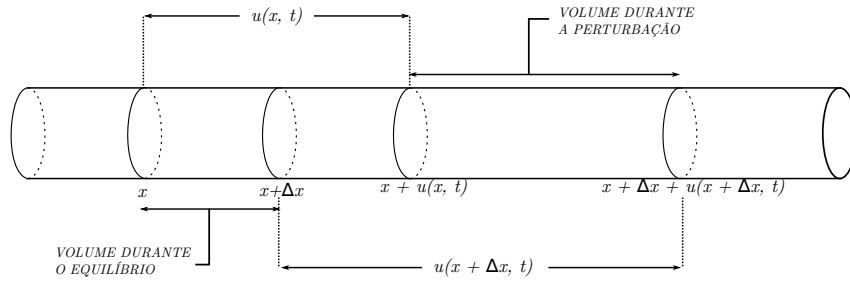
Analisando a Equação A.4, é possível afirmar que a perturbação na pressão p_e é proporcional a perturbação na densidade ρ_e . Logo, é definida uma constante de proporcionalidade κ :

$$p_e = \kappa \rho_e, \quad \text{onde } \kappa = f'(\rho_0) = \left. \frac{dp}{d\rho} \right|_{\rho=\rho_0}; \quad (\text{A.5})$$

que define a expressão matemática para a 2ª suposição.

Avaliando a 1ª suposição, considere a posição x do meio em equilíbrio. Com a chegada da frente de onda essa posição sofre um deslocamento $u(x, t)$, de forma que sua nova posição se encontre em $x + u(x, t)$. Analisando uma posição próxima a x que está afastada de Δx , ou seja, a posição $x + \Delta x$, que antes da chegada da onda está em equilíbrio. Com a chegada da frente de onda, a posição $x + \Delta x$ sofre um deslocamento $u(x + \Delta x, t)$ ficando na posição $x + \Delta x + u(x + \Delta x, t)$. A Figura 81 ilustra a situação.

Figura 81 – Perturbação no meio unidimensional provocada pela chegada da frente de onda.



Fonte: o autor.

Sabendo que a quantidade de matéria tanto na região em equilíbrio Δx , quanto na região durante a perturbação $\Delta x + u(x + \Delta x, t) - u(x, t)$ permanece a mesma, isso implica que a região em equilíbrio possui a mesma massa (m_0) que a região durante a passagem da perturbação (m_e). Logo:

$$\begin{aligned} m_0 &= m_e \\ \rho_0 \Delta x &= \rho [\Delta x + u(x + \Delta x, t) - u(x, t)] \\ \rho_0 &= \rho \left[1 + \frac{u(x + \Delta x, t) - u(x, t)}{\Delta x} \right]. \end{aligned} \quad (\text{A.6})$$

No limite em que as regiões em estudo são infinitesimais, ou seja, $\Delta x \rightarrow 0$ e usando a Equação A.4:

$$\begin{aligned} \rho_0 &= (\rho_0 + \rho_e) + (\rho_0 + \rho_e) \frac{\partial u(x, t)}{\partial x} \\ \rho_e &= -\rho_0 \frac{\partial u(x, t)}{\partial x} - \rho_e \frac{\partial u(x, t)}{\partial x}. \end{aligned} \quad (\text{A.7})$$

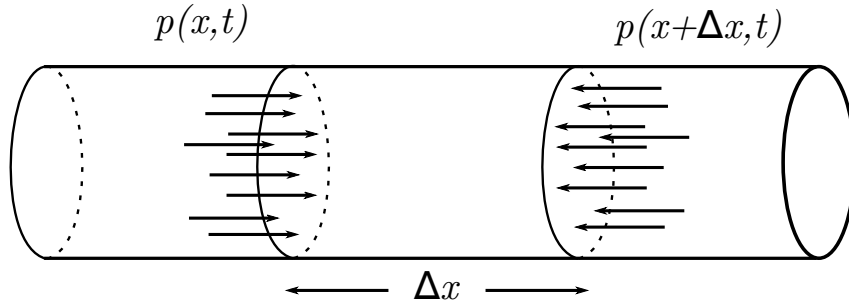
Como $\rho_e \ll \rho_0$ e o deslocamento $u(x, t)$ é muito pequeno, o termo $\rho_e \frac{\partial u(x, t)}{\partial x}$ pode ser desprezado, obtendo:

$$\rho_e = -\rho_0 \frac{\partial u(x, t)}{\partial x}; \quad (\text{A.8})$$

a expressão matemática que traduz a 1ª suposição.

Por fim, para escrever a expressão matemática para a 3ª suposição, considere a posição x do meio que sofre uma diferença de pressão (Figura 82). Esse desequilíbrio na pressão provoca um deslocamento da região Δx . Assim, a força resultante pode ser calculada pela 2ª Lei de Newton, onde a massa é dada por $m_0 = \rho_0 \Delta x$ e a aceleração é dada por $\frac{\partial^2 u(x,t)}{\partial t^2}$. Portanto, temos que:

Figura 82 – A diferença de pressão no meio produz uma força resultante.



Fonte: o autor.

$$\begin{aligned} p(x, t) - p(x + \Delta x, t) &= \rho_0 \Delta x \frac{\partial^2 u(x, t)}{\partial x^2} \\ - \left(\frac{p(x + \Delta x, t) - p(x, t)}{\Delta x} \right) &= \rho_0 \frac{\partial^2 u(x, t)}{\partial t^2}. \end{aligned} \quad (\text{A.9})$$

Fazendo o limite $\Delta \rightarrow 0$:

$$-\frac{\partial p(x, t)}{\partial x} = \rho_0 \frac{\partial^2 u(x, t)}{\partial t^2} \quad (\text{A.10})$$

é obtida a relação matemática que traduz a 3ª suposição.

Como $p_e(x, t)$ é a única parte que varia em $p(x, t)$ (Equação A.3), aplicando o resultado da 2ª suposição (Equação A.5) na expressão anterior (Equação A.10), obtem-se:

$$\begin{aligned} -\frac{\partial p_e(x, t)}{\partial x} &= \rho_0 \frac{\partial^2 u(x, t)}{\partial t^2} \\ -\frac{\partial(\kappa \rho_e)}{\partial x} &= \rho_0 \frac{\partial^2 u(x, t)}{\partial t^2}. \end{aligned} \quad (\text{A.11})$$

e, finalmente, utilizando a 1ª suposição (Equação A.8):

$$\begin{aligned} -\kappa \frac{\partial}{\partial x} \left(-\rho_0 \frac{\partial u(x, t)}{\partial x} \right) &= \rho_0 \frac{\partial^2 u(x, t)}{\partial t^2} \\ \frac{\partial^2 u(x, t)}{\partial t^2} &= \kappa \frac{\partial^2 u(x, t)}{\partial x^2}. \end{aligned} \quad (\text{A.12})$$

Chamando a constante de proporcionalidade κ de c^2 , obtemos a forma mais conhecida da equação da onda em uma dimensão:

$$\frac{\partial^2 u(x, t)}{\partial t^2} = c^2 \frac{\partial^2 u(x, t)}{\partial x^2}. \quad (\text{A.13})$$

A.2 A equação da onda em meios elásticos

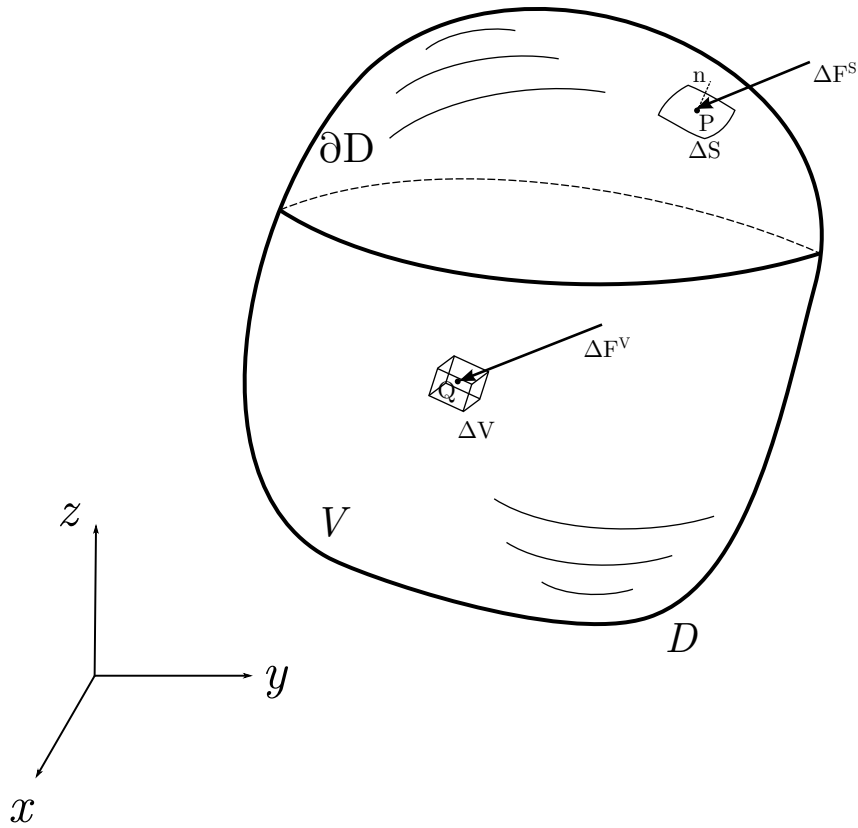
Para generalizar o conceito da propagação de ondas e compreender como a equação da onda elástica surge a partir dos conceitos da mecânica aplicada em meios contínuos, considere um corpo D qualquer com massa m e volume V (AKI; RICHARDS, 2002; LOWRIE, 2007). Negligenciando a atuação da gravidade, no corpo atuam dois tipos de forças: as forças de superfície (ou de contato) e as forças de corpo (Figura 83). Note que a equação da onda encontrada na seção A.1 é um caso particular de meios elásticos.

A.2.1 Forças de superfície

A força de superfície pode ser definida sobre uma região ΔS , na superfície ∂D do corpo D (Figura 83). Considere um ponto $P \in \Delta S$ e o vetor posição $\mathbf{x}$, em um instante de tempo t qualquer. A resultante da distribuição de forças em ΔS é dada por $\Delta \mathbf{F}^S$. No limite em que ΔS tende ao ponto P , ou seja, $\Delta S \rightarrow 0$. Definindo o vetor tração $\mathbf{t}$ como:

$$\mathbf{t} = \lim_{\Delta S \rightarrow 0} \frac{\Delta \mathbf{F}^S}{\Delta S}, \quad (\text{A.14})$$

Figura 83 – Corpo D qualquer, onde são analisadas as forças de superfície e as forças de corpo. Note que o ponto P está na superfície do corpo D e o ponto Q está no interior do corpo D .



Fonte: adaptado de Aki e Richards (2002).

a Equação A.14 representa a força por unidade de área no ponto P sobre a superfície ∂D .

Sabendo que $\mathbf{n}$ é o à superfície ∂D no ponto P , o vetor tração $\mathbf{t}$ pode ser reescrito em termos do tensor de tensões σ , projetado na direção normal $\mathbf{n}$ ao ponto P :

$$\mathbf{t} = \sigma \cdot \mathbf{n}, \quad (\text{A.15})$$

se a projeção for positiva, obtém-se uma tração (ou distensão), caso seja negativa, obtém-se uma compressão.

Portanto, é possível definir que a força resultante na superfície ∂D é dada pela

seguinte integral de superfície:

$$\mathbf{F}^S = \oint_{\partial D} \mathbf{t} dS \quad (\text{A.16a})$$

$$\mathbf{F}^S = \oint_{\partial D} \boldsymbol{\sigma} \cdot \mathbf{n} dS. \quad (\text{A.16b})$$

A.2.2 Forças de corpo

Considere um ponto Q no interior do corpo D , e também, um volume ΔV no interior de D onde $Q \in \Delta V$ (Figura 83). Leve em consideração também o vetor posição $\mathbf{x}$ e um instante de tempo qualquer t , a resultante de forças em ΔV é dada por $\Delta \mathbf{F}^V$ como mostra a Figura 83. Assumindo que ΔV possui uma densidade constante ρ_0 , no limite em que $\Delta V \rightarrow 0$ é possível definir uma densidade de força $\rho \mathbf{f}^1$, onde $\rho(\mathbf{x})$ é a densidade do ponto Q e $\mathbf{f}(\mathbf{x})$ é a força por unidade de massa:

$$\rho \mathbf{f} = \lim_{\Delta V \rightarrow 0} \frac{\Delta \mathbf{F}^V}{\Delta V}. \quad (\text{A.17})$$

A Equação A.17 representa a densidade de força no corpo D , a força que atua sobre todo o corpo D (força de corpo) é dada pela integral de volume:

$$\mathbf{F}^V = \int_V \rho \cdot \mathbf{f} dV. \quad (\text{A.18})$$

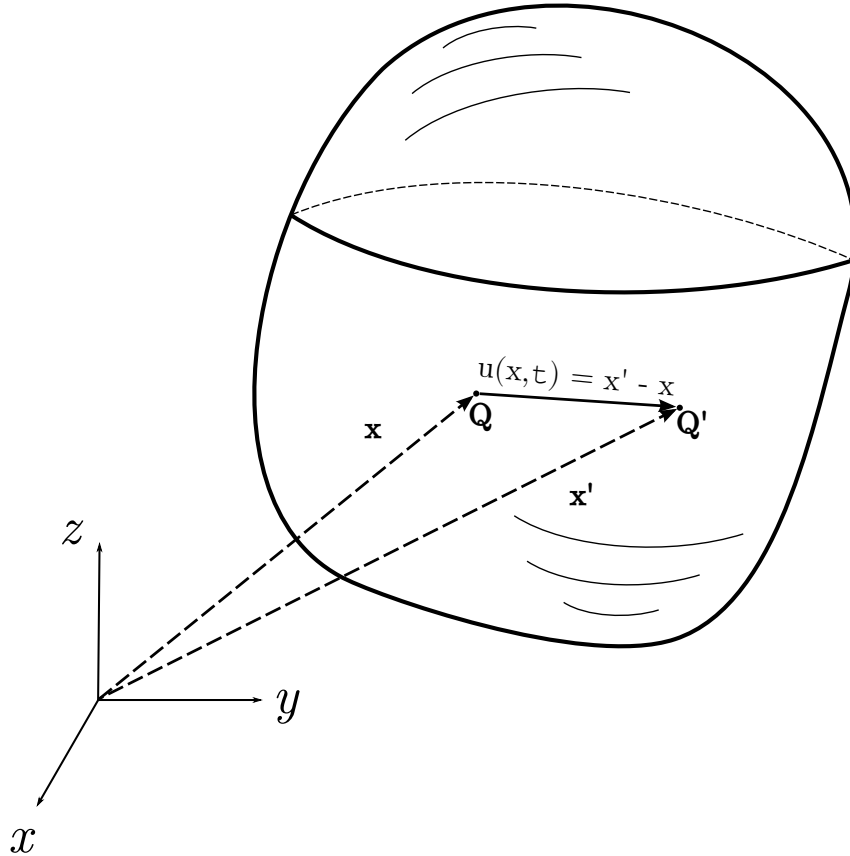
A.2.3 Equação do movimento

Antes de definir a equação do movimento para um meio elástico qualquer, é necessário definir duas grandezas: o deslocamento de partícula e a velocidade de partícula. O vetor deslocamento de partícula $\mathbf{u}(\mathbf{x})$, representa o deslocamento de um ponto

¹ $m = \rho V \rightarrow \overbrace{ma}^F = \overbrace{\rho V a}^F \rightarrow \frac{F}{V} = \frac{\rho \overbrace{ma}^F}{m} \rightarrow \frac{F}{V} = \rho \mathbf{f}$

infinitesimal entre os instantes t e t' dentro do corpo D quando sofre uma deformação (Figura 84).

Figura 84 – Representação do deslocamento de partícula provocado pela deformação do corpo D entre os instantes de tempo t e t' .



Fonte: adaptado de Aki e Richards (2002).

O vetor velocidade de partícula é a taxa de variação com que o vetor deslocamento de partícula varia no tempo:

$$\mathbf{v}(\mathbf{x}, t) = \frac{\partial \mathbf{u}(\mathbf{x}, t)}{\partial t}. \quad (\text{A.19})$$

Vale a pena ressaltar que tanto o deslocamento de partícula $\mathbf{u}(\mathbf{x}, t)$, como a velocidade de partícula $\mathbf{v}(\mathbf{x}, t)$, são independentes do sistema de coordenadas (PUJOL, 2003).

Utilizando o princípio da conservação do momento linear generalizado, que postula que a resultante das forças aplicadas ao corpo D deve ser igual à taxa de variação do momento linear do corpo D , é possível escrever:

$$\begin{aligned}
\frac{\partial}{\partial t} \int_V \rho \mathbf{v} dV &= \mathbf{F}^S + \mathbf{F}^V \\
\frac{\partial}{\partial t} \int_V \rho \mathbf{v} dV &= \oint_{\partial D} \boldsymbol{\sigma} \cdot \mathbf{n} dS + \int_V \rho \mathbf{f} dV.
\end{aligned} \tag{A.20}$$

Utilizando o teorema de Gauss, a integral de superfície da Equação A.20 pode ser transformada em uma integral de volume como mostra a equação a seguir:

$$\oint_{\partial D} \boldsymbol{\sigma} \cdot \mathbf{n} dS = \int_V \nabla \cdot \boldsymbol{\sigma} dV. \tag{A.21}$$

Substituindo esse resultado na Equação A.20 encontra-se:

$$\frac{\partial}{\partial t} \int_V \rho \cdot \mathbf{v} dV = \int_V \nabla \cdot \boldsymbol{\sigma} dV + \int_V \rho \mathbf{f} dV. \tag{A.22}$$

Como a densidade $\rho(\mathbf{x})$ não depende do tempo e todos os termos da Equação A.22 são integrados em todo o volume, a equação diferencial do movimento fica:

$$\underbrace{\nabla \cdot \boldsymbol{\sigma}(\mathbf{x}, t)}_{\text{Força de Superfície} / V} - \underbrace{\rho(\mathbf{x}) \cdot \frac{\partial}{\partial t} \mathbf{v}(\mathbf{x}, t)}_{\text{Conservação do Momento linear}} = \underbrace{\rho(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}, t)}_{\text{Força de Corpo} / V} \tag{A.23}$$

onde o primeiro termo está relacionado com a definição da força de superfície, o segundo surge a partir da aplicação da conservação do momento linear e o último está relacionado com as forças de corpo.

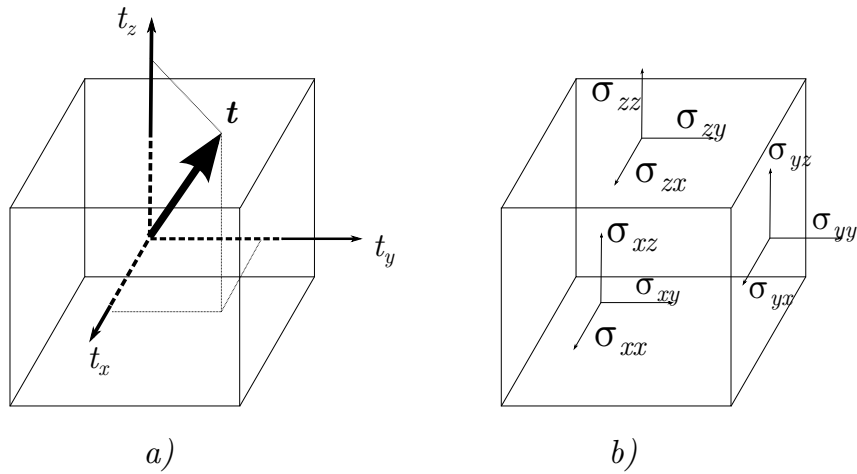
A.2.4 O tensor de tensões

O tensor de tensões $\boldsymbol{\sigma}$ (*stress tensor*) descreve o estado de tensões na vizinhança de um ponto material em deformação. Para defini-lo, é analisado o vetor tração (Figura 85a) apresentado na Equação A.14, onde sua magnitude e direção variam em função da orientação do plano infinitesimal do ponto em deformação (Figura 85b). Assim, para descrever completamente as forças internas do meio é necessário determinar a

tração $\mathbf{t}$, em função do vetor normal $\mathbf{n}$ ao plano infinitesimal, que sofre a deformação (SHEARER, 2009). Essa relação é linear e dada pelo tensor de tensões σ (Figura 85b):

$$\mathbf{t} = \sigma \cdot \mathbf{n} \quad \rightarrow \quad \begin{cases} t_x = \sigma_{xx}\hat{x} + \sigma_{xy}\hat{y} + \sigma_{xz}\hat{z} \\ t_y = \sigma_{yx}\hat{x} + \sigma_{yy}\hat{y} + \sigma_{yz}\hat{z} \\ t_z = \sigma_{zx}\hat{x} + \sigma_{zy}\hat{y} + \sigma_{zz}\hat{z}. \end{cases} \quad (\text{A.24})$$

Figura 85 – a) Representação do vetor tração. b) Representação do tensor de tensões.



Fonte: o autor.

Ao lidar com tensores a notação vetorial nem sempre é indicada, pois nos obriga a pensar sempre em uma perspectiva geométrica, o que de certa forma dificulta a manipulação algébrica das equações. Na Física-Matemática é comum utilizar a notação tensorial em conjunto com a convenção de Einstein, para simplificar as manipulações algébricas. Portanto, a Equação A.24 utilizando a notação de Einstein, fica:

$$\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \rightarrow \quad t_i = \sigma_{ij}n_j \quad (\text{A.25})$$

onde os índices i e j representam as coordenadas espaciais, e variam de 1 a 3. Além disso, é importante perceber que no tensor de tensões o índice i representa a orientação da superfície de aplicação da tensão (normal ao eixo x_i), e o índice j representa a orientação da tensão (paralela ao eixo x_j). Portanto, quando $i = j$ temos as tensões normais (ou principais), e quando $i \neq j$ temos as tensões cisalhantes.

O tensor de tensões é simétrico, ou seja, $\sigma_{ij} = \sigma_{ji}$, isso pode ser demonstrado através da aplicação do princípio de conservação do momento angular (PUJOL, 2003). Assim, a equação do movimento (Equação A.23) pode ser escrita utilizando a notação de Einstein:

$$\partial_j \sigma_{ij} - \rho \partial_t v_i = \rho f_i. \quad (\text{A.26})$$

A Equação A.26 pode ser reescrita em termos do vetor deslocamento, utilizando a Equação A.19:

$$\partial_j \sigma_{ij} - \rho \partial_t^2 u_i = \rho f_i. \quad (\text{A.27})$$

A.2.5 O tensor de deformação

A relação entre a tensão σ e o vetor deslocamento de partícula u é alcançada através da lei de Hooke (SHEARER, 2009). Será possível remover a indeterminação da equação do movimento (Equação A.23) e reescrevê-la em termos do deslocamento de partícula, obtendo assim a equação da onda para meios elásticos. Dessa forma, considere a relação tensão-deformação dada por:

$$\sigma_{ij} = c_{ijkl} \varepsilon_{kl} \quad (\text{A.28})$$

onde c_{ijkl} é um tensor de 4ª ordem, conhecido como tensor de rigidez (ou tensor elástico) e ε_{kl} é o tensor de deformações. A Equação A.28 também é conhecida como relação constitutiva.

A lei de Hooke definirá em que tipo de meio a perturbação se propaga. Para meios elásticos, anisotrópicos e heterogêneos, a relação anterior é linear, isso significa que o operador c_{ijkl} que relaciona σ_{ij} com ε_{kl} é independente da deformação ε_{kl} e invariante no tempo (SLAUGHTER, 2002).

Ainda de acordo com Shearer (2009), a deformação é uma medida local das variações relativas no campo de deslocamento de partícula. Ela relaciona as variações no volume e na forma, além das variações na posição do meio. Shearer (2009) mostra que o tensor de deformações pode ser escrito em termos do deslocamento de partícula, através de:

$$\varepsilon_{ij} = \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{1}{2} \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right) & \frac{1}{2} \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \right) \\ \frac{1}{2} \left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2} \right) & \frac{\partial u_2}{\partial x_2} & \frac{1}{2} \left(\frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2} \right) \\ \frac{1}{2} \left(\frac{\partial u_3}{\partial x_1} + \frac{\partial u_1}{\partial x_3} \right) & \frac{1}{2} \left(\frac{\partial u_3}{\partial x_2} + \frac{\partial u_2}{\partial x_3} \right) & \frac{\partial u_3}{\partial x_3} \end{pmatrix} \quad (\text{A.29})$$

ou utilizando a notação tensorial:

$$\varepsilon_{ij} = \frac{1}{2} (\partial_i u_j + \partial_j u_i) \quad (\text{A.30})$$

A análise da Equação A.29 permite dizer que o tensor de deformação também é simétrico. Portanto, $\varepsilon_{ij} = \varepsilon_{ji}$.

A.2.6 O tensor elástico e a relação tensão-deformação

O tensor c_{ijkl} é um tensor de 4ª ordem com 81 componentes ($3 \times 3 \times 3 \times 3$). Mas devido à simetria dos tensores de tensões σ_{ij} e deformações ε_{ij} , e além disso, à consideração termodinâmica de que a perturbação do meio elástico é um processo adiabático e isotérmico, somente 21 componentes são linearmente independentes, apesar do tensor c_{ijkl} possuir 36 componentes. As 21 componentes linearmente independentes do tensor elástico descrevem um meio elástico, anisotrópico e heterogêneo. Dessa forma, a simetria do tensor de rigidez é descrita por:

$$c_{ijkl} = c_{jikl} = c_{ijlk} = c_{klij} \quad (\text{A.31})$$

Devido à simetria do tensor de rigidez é possível reescrever a relação tensão-deformação, dada pela Equação A.28, em uma forma mais concisa, utilizando a notação de Voigt (JAEGER; COOK; ZIMMERMAN, 2007). Nessa notação os tensores de segunda ordem σ e ε são escritos como vetores, com 6 componentes cada, e o tensor elástico de quarta ordem é escrito como uma matriz 6×6 . Apesar dos tensores de deformação e tensão possuírem 9 componentes apenas 6 delas são linearmente independentes, por causa de suas respectivas simetrias. Assim, a notação de Voigt indexa os índices da seguinte forma:

$$\begin{cases} ij \rightarrow m \\ kl \rightarrow n \end{cases} \rightarrow \begin{cases} 11 \rightarrow 1 & 23 \rightarrow 4 \\ 22 \rightarrow 2 & 13 \rightarrow 5 \\ 33 \rightarrow 3 & 12 \rightarrow 6 \end{cases}$$

e, a lei de Hooke fica:

$$\sigma_{ij} = c_{ijkl} \varepsilon_{kl}$$

$$\sigma_m = c_{mn} \varepsilon_n$$

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ & & c_{33} & c_{34} & c_{35} & c_{36} \\ & & & c_{44} & c_{45} & c_{46} \\ & & & & c_{55} & c_{56} \\ & & & & & c_{66} \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ 2\varepsilon_4 \\ 2\varepsilon_5 \\ 2\varepsilon_6 \end{pmatrix} \quad (\text{A.32})$$

Um meio anisotrópico é aquele em que suas propriedades, além de variarem com a posição, também variam de acordo com a direção de análise. A aproximação para meios isotrópicos, ou seja, meios em que a propriedade não varia de acordo com a direção de propagação da onda são bastante utilizados na descrição da propagação de ondas sísmicas. Uma das principais vantagens de se utilizar a aproximação para meios isotrópicos é que o tensor elástico que descreve o meio possui apenas 2 componentes linearmente independentes e 6 componentes diferentes de zero, o que simplifica a descrição da propagação de ondas em tais meios (SLAUGHTER, 2002).

Portanto, em meios elásticos, isotrópicos e heterogêneos, o tensor de rigidez é dado por:

$$c_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{il} \delta_{jk} + \delta_{ik} \delta_{jl}), \quad (\text{A.33})$$

onde λ é o parâmetro de Lamé, μ é o módulo de cisalhamento e δ_{ij} é o delta de Kronecker, definido por:

$$\delta_{ij} = \begin{cases} 1 & \text{para } i = j \\ 0 & \text{para } i \neq j. \end{cases} \quad (\text{A.34})$$

Utilizando a Equação A.34 na Equação A.33, é encontrada a relação tensão-deformação em termos dos parâmetros de Lamé λ e μ :

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} = \begin{pmatrix} \lambda + 2\mu & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda + 2\mu & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda + 2\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ 2\varepsilon_4 \\ 2\varepsilon_5 \\ 2\varepsilon_6 \end{pmatrix}. \quad (\text{A.35})$$

A.2.7 A equação da onda para meios elásticos

A equação do movimento A.27, em conjunto com a lei de Hooke (Equação A.28) e a relação entre deformação e deslocamento de partícula, dada pela Equação A.30 permitem escrever uma expressão que descreve o movimento ondulatório em um meio elástico, anisotrópico e heterogêneo, dado por:

$$\begin{cases} \nabla \cdot \boldsymbol{\sigma}(\mathbf{x}, t) - \rho(\mathbf{x}) \frac{\partial^2}{\partial t^2} \mathbf{u}(\mathbf{x}, t) = \rho(\mathbf{x}) \mathbf{f}(\mathbf{x}, t) \\ \boldsymbol{\sigma}(\mathbf{x}, t) = \mathbf{c}(\mathbf{x}) : \frac{1}{2} [\nabla \mathbf{u}(\mathbf{x}, t) + (\nabla \mathbf{u}(\mathbf{x}, t))^T] \end{cases} \quad (\text{A.36a})$$

ou na notação de Einstein:

$$\begin{cases} \partial_j \sigma_{ij} - \rho \partial_t^2 u_i = -\rho f_i \\ \sigma_{ij} = c_{ijkl} \frac{1}{2} (\partial_k u_l + \partial_l u_k). \end{cases} \quad (\text{A.36b})$$

Utilizando o tensor elástico para meios isotrópicos (Equação A.33) na equação A.36b, é obtida a equação da onda elástica para meios isotrópicos e heterogêneos em sua formulação tensão-deslocamento:

$$\begin{cases} \partial_j \sigma_{ij} - \rho \partial_t \partial_t u_i = -\rho f_i \\ \sigma_{ij} = \lambda \delta_{ij} \partial_k u_k + \mu (\partial_i u_j + \partial_j u_i). \end{cases} \quad (\text{A.37})$$

A equação da onda elástica pode ser reescrita em sua formulação tensão-velocidade. A vantagem de se utilizar essa formulação é que a equação que descreve o comportamento ondulatório é um sistema de 1ª ordem tanto no tempo quanto no espaço. Em contraste, a formulação tensão-deslocamento é dada por um sistema de 2ª ordem no tempo e no espaço.

Logo, a formulação tensão-velocidade da equação da onda elástica, em meios anisotrópicos e heterogêneos, é alcançada utilizando a Equação A.19, e derivando com relação ao tempo a lei de Hooke na Equação A.36a ou A.36b:

$$\begin{cases} \nabla \cdot \boldsymbol{\sigma}(\mathbf{x}, t) - \rho(\mathbf{x}) \cdot \frac{\partial}{\partial t} \mathbf{v}(\mathbf{x}, t) = -\rho(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}, t) \\ \frac{\partial}{\partial t} \boldsymbol{\sigma}(\mathbf{x}, t) = \mathbf{c}(\mathbf{x}) : \frac{1}{2} [\nabla \mathbf{v}(\mathbf{x}, t) + (\nabla \mathbf{v}(\mathbf{x}, t))^T] \end{cases} \quad (\text{A.38a})$$

ou:

$$\begin{cases} \partial_j \sigma_{ij} - \rho \cdot \partial_t v_i = -\rho f_i \\ \partial_t \sigma_{ij} = c_{ijkl} \frac{1}{2} (\partial_k v_l + \partial_l v_k). \end{cases} \quad (\text{A.38b})$$

A.2.8 A equação da onda elástica em meios homogêneos e isotrópicos

A análise da aproximação da equação da onda para meios homogêneos e isotrópicos na ausência de fonte ($\mathbf{f} = 0$), apesar de ser um problema pouco realístico, nos permite obter alguns resultados importantes como será apresentado a seguir. Para isso, será substituída a relação tensão-deformação (2ª Equação do sistema A.37) na equação do movimento (1ª Equação do sistema A.37), assim, obtêm-se a seguinte expressão:

$$\begin{aligned}
\rho \partial_t \partial_t u_i - \partial_j [\lambda \delta_{ij} \partial_k u_k + \mu (\partial_i u_j + \partial_j u_i)] &= 0 \\
\rho \partial_t \partial_t u_i - \partial_j [\lambda \delta_{ij} \partial_k u_k] + \partial_j [\mu (\partial_i u_j + \partial_j u_i)] &= 0 \\
\rho \partial_t \partial_t u_i - \underbrace{\delta_{ij} \partial_j}_{\partial_i} \lambda \partial_k u_k + \lambda \underbrace{\delta_{ij} \partial_j}_{\partial_i} \partial_k u_k + \partial_j \mu (\partial_i u_j + \partial_j u_i) + \mu (\partial_j \partial_i u_j + \partial_j \partial_j u_i) &= 0.
\end{aligned}
\tag{A.39}$$

Em meios homogêneos, os termos que contêm as derivadas em relação ao espaço dos parâmetros de Lamé (λ e μ) são iguais a zero, tornando a equação anterior mais simples, como pode ser visto a seguir:

$$\rho \partial_t \partial_t u_i - \lambda \partial_j \partial_k u_k + \mu \partial_j \partial_i u_j + \mu \partial_j \partial_j u_i = 0, \tag{A.40a}$$

utilizando a notação vetorial:

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} - \lambda \nabla \nabla \cdot \mathbf{u} + \mu \nabla \nabla \cdot \mathbf{u} + \mu \nabla^2 \mathbf{u} = 0. \tag{A.40b}$$

Para remover o termo $\nabla^2 \mathbf{u}$, é utilizada a seguinte identidade vetorial:

$$\begin{aligned}
\nabla \times \nabla \times \mathbf{u} &= \nabla \nabla \cdot \mathbf{u} - \nabla^2 \mathbf{u} \\
\nabla^2 \mathbf{u} &= \nabla \nabla \cdot \mathbf{u} - \nabla \times \nabla \times \mathbf{u}
\end{aligned}
\tag{A.41}$$

e substituindo em A.40b:

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} - (\lambda + 2\mu) \nabla \nabla \cdot \mathbf{u} + \mu \nabla \times \nabla \times \mathbf{u} = 0. \tag{A.42}$$

A equação A.42 é conhecida como equação de Navier-Stokes. Essa equação descreve a propagação de dois campos, um campo escalar ($\nabla \cdot \mathbf{u}$) e um campo vetorial ($\nabla \times \mathbf{u}$). Para desacoplar o campo escalar do vetorial na Equação A.42, é necessário aplicar o operador divergente a todos os termos da equação anterior:

$$\rho \frac{\partial^2(\nabla \cdot \mathbf{u})}{\partial t^2} - (\lambda + 2\mu) \nabla^2(\nabla \cdot \mathbf{u}) + \mu \overbrace{\nabla \cdot (\nabla \times \nabla \times \mathbf{u})}^0 = 0$$

$$\underbrace{\frac{\partial^2(\nabla \cdot \mathbf{u})}{\partial t^2}}_{\text{derivada temporal}} = \underbrace{\frac{(\lambda + 2\mu)}{\rho}}_{\text{velocidade}} \underbrace{\nabla^2(\nabla \cdot \mathbf{u})}_{\text{derivada espacial}} \quad (\text{A.43})$$

e agora, aplicando o operador rotacional novamente na Equação A.42:

$$\rho \frac{\partial^2(\nabla \times \mathbf{u})}{\partial t^2} - (\lambda + 2\mu) \overbrace{\nabla \times \nabla \nabla \cdot \mathbf{u}}^0 + \mu \nabla \times \nabla \times \nabla \times \mathbf{u} = 0$$

e usando novamente a identidade vetorial A.41: $\nabla \times \nabla \times (\nabla \times \mathbf{u}) = \nabla \nabla \cdot (\nabla \times \mathbf{u}) - \nabla^2(\nabla \times \mathbf{u})$. É possível chegar em:

$$\rho \frac{\partial^2(\nabla \times \mathbf{u})}{\partial t^2} + \mu \overbrace{\nabla \nabla \cdot (\nabla \times \mathbf{u})}^0 - \mu \nabla^2(\nabla \times \mathbf{u}) = 0$$

$$\underbrace{\frac{\partial^2(\nabla \times \mathbf{u})}{\partial t^2}}_{\text{derivada temporal}} = \underbrace{\frac{\mu}{\rho}}_{\text{velocidade}} \underbrace{\nabla^2(\nabla \times \mathbf{u})}_{\text{derivada espacial}} \quad (\text{A.44})$$

Se compararmos as Equações A.43 e A.44 com a Equação A.13 (equação da onda em uma dimensão), é possível perceber que elas possuem a mesma forma. Portanto, a Equação A.43 descreve a propagação do campo compressional (ou onda P) com velocidade de propagação dada por:

$$v_p^2 = \frac{\lambda + 2\mu}{\rho} \quad (\text{A.45})$$

sendo o campo de pressão p proporcional a variação de volume definido por $\nabla \cdot \mathbf{u}$. E a Equação A.44 descreve a propagação do campo cisalhante (ou onda S) com velocidade de propagação dada por:

$$v_s^2 = \frac{\mu}{\rho} \quad (\text{A.46})$$

onde o campo que representa as pequenas rotações do deslocamento de partícula $\mathbf{u}$ é proporcional a $\nabla \times \mathbf{u}$.

A.3 A equação da onda em meios acústicos

Como apresentado na seção anterior, a equação da onda para meios elásticos, isotrópicos e heterogêneos pode ser descrita em termos dos parâmetros de Lamé λ e μ , que caracterizam o meio onde a onda se propaga. Em meios elásticos, o parâmetro μ representa a resistência ao cisalhamento do meio. No caso acústico, a resistência ao cisalhamento é nula ($\mu = 0$), o que torna o tensor de rigidez A.33 muito mais simples:

$$c_{ijkl} = \lambda \delta_{ij} \delta_{kl} = \begin{pmatrix} \lambda & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (\text{A.47})$$

Portanto, aplicando o tensor de rigidez para o caso acústico (Equação A.47) na Equação A.38b, é possível descrever a propagação da onda para meios acústicos, isotrópicos e heterogêneos na formulação pressão-velocidade:

$$\begin{cases} \left[\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + \frac{\partial \sigma_{xz}}{\partial z} \right] - \rho(\mathbf{x}) \frac{\partial v_x}{\partial t} = \rho f_x \\ \left[\frac{\partial \sigma_{yx}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \sigma_{yz}}{\partial z} \right] - \rho(\mathbf{x}) \frac{\partial v_y}{\partial t} = \rho f_y \\ \left[\frac{\partial \sigma_{zx}}{\partial x} + \frac{\partial \sigma_{zy}}{\partial y} + \frac{\partial \sigma_{zz}}{\partial z} \right] - \rho(\mathbf{x}) \frac{\partial v_z}{\partial t} = \rho f_z \\ \frac{\partial p}{\partial t} - \kappa \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right) = 0. \end{cases} \quad (\text{A.48})$$

Note que, na ausência de tensões cisalhantes ($\sigma_{xy} = \sigma_{xz} = \sigma_{yz} = 0$), e por se tratar também de um meio isotrópico, as tensões normais são iguais ($\sigma_{xx} = \sigma_{yy} = \sigma_{zz}$), o que nos permite escrevê-la em termos da pressão p . Em meios acústicos, por não ocorrer tensões cisalhantes, o meio apenas sofre dilatação (ou compressão), logo é possível escrever que:

$$p = \kappa \nabla \cdot \mathbf{u}, \quad (\text{A.49})$$

ou seja, p é proporcional a variação do meio (LOWRIE, 2007), onde κ é a constante de proporcionalidade e descreve a incompressibilidade do meio. Como é possível

perceber, o parâmetro de Lamé λ em meios acústicos é igual a incompressibilidade do meio.

A Equação A.48 escrita em notação vetorial, fica:

$$\begin{cases} \nabla p(\mathbf{x}, t) - \rho(\mathbf{x}) \frac{\partial}{\partial t} \mathbf{v}(\mathbf{x}, t) = \rho(\mathbf{x}) \mathbf{f} \\ \frac{\partial}{\partial t} p(\mathbf{x}, t) - \kappa(\mathbf{x}) \nabla \cdot \mathbf{v}(\mathbf{x}, t) = 0 \end{cases}, \quad (\text{A.50a})$$

e, em notação tensorial, fica:

$$\begin{cases} \partial_j p - \rho \partial_t v_i = \rho f_i \\ \partial_t p - \kappa \partial_i v_i = 0 \end{cases}, \quad (\text{A.50b})$$

essa formulação também é conhecida como equação da onda acústica com densidade variável.

A equação da onda acústica também pode ser escrita somente em termos do campo de pressão. Para isso, precisamos tomar o divergente da primeira Equação A.50a, e a derivada temporal da segunda equação:

$$\begin{cases} \nabla \cdot \left(\frac{1}{\rho(\mathbf{x})} \nabla p(\mathbf{x}, t) \right) - \nabla \cdot \left(\frac{\partial}{\partial t} \mathbf{v}(\mathbf{x}, t) \right) = \nabla \cdot \mathbf{f}(\mathbf{x}, t) \\ \frac{1}{\kappa(\mathbf{x})} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) - \frac{\partial}{\partial t} \nabla \cdot \mathbf{v}(\mathbf{x}, t) = 0. \end{cases} \quad (\text{A.51})$$

E subtraindo as duas Equações A.51 encontra-se:

$$\nabla \cdot \left(\frac{1}{\rho(\mathbf{x})} \nabla p(\mathbf{x}, t) \right) - \frac{1}{\kappa(\mathbf{x})} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) = \nabla \cdot \mathbf{f}(\mathbf{x}, t). \quad (\text{A.52})$$

Assumindo que o gradiente do inverso da densidade é desprezível na Equação A.51, ou seja, que a densidade é aproximadamente constante, e multiplicando todos os termos da equação por $\rho(\mathbf{x})$:

$$\frac{\rho(\mathbf{x})}{\rho(\mathbf{x})} \nabla \cdot (\nabla p(\mathbf{x}, t)) - \frac{\rho(\mathbf{x})}{\kappa(\mathbf{x})} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) = \rho(\mathbf{x}) \nabla \cdot \mathbf{f}(\mathbf{x}, t) \quad (\text{A.53})$$

a velocidade de propagação $v_p(\mathbf{x})$ em meios acústicos é definida como:

$$\frac{\rho(\mathbf{x})}{\kappa(\mathbf{x})} = \frac{1}{v_p^2(\mathbf{x})} \quad \text{ou} \quad v_p(\mathbf{x}) = \sqrt{\frac{\kappa(\mathbf{x})}{\rho(\mathbf{x})}}. \quad (\text{A.54})$$

E usando o termo fonte como:

$$w(\mathbf{x}, t) = \rho(\mathbf{x}) \nabla \mathbf{f}(\mathbf{x}, t), \quad (\text{A.55})$$

a equação da onda acústica fica:

$$\nabla^2 p(\mathbf{x}, t) - \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) = w(\mathbf{x}, t). \quad (\text{A.56})$$

A.4 O termo fonte

Existem diversas formas de adicionar uma perturbação à equação da onda para simular a fonte sísmica (ou *wavelet*) como, por exemplo, funções senos e cossenos. Entretanto, essas funções não são adequadas para simular as fontes sísmicas reais. A fonte sísmica tem papel fundamental na modelagem sísmica, consequentemente, na aplicação da FWI. A escolha de uma *wavelet* inicial apropriada é crucial para uma aplicação da FWI bem sucedida.

A *wavelet* Ricker é um exemplo de fonte sísmica largamente utilizada em modelagens sísmicas devido às suas propriedades. Essa *wavelet* possui uma expressão analítica onde é possível definir sua frequência de corte com facilidade.

Um outro exemplo de fonte sísmica é a *wavelet wide-band B-spline*, que também possui propriedades muito úteis para aplicação de testes numéricos. A equação que descreve essa *wavelet* permite definir a frequência inicial e a frequência de corte da fonte. Essa propriedade permite a construção de *wavelets* que simulam aquisições sísmicas que utilizam fontes com banda de frequência larga (*Broadband*).

A.4.1 A *wavelet* Ricker

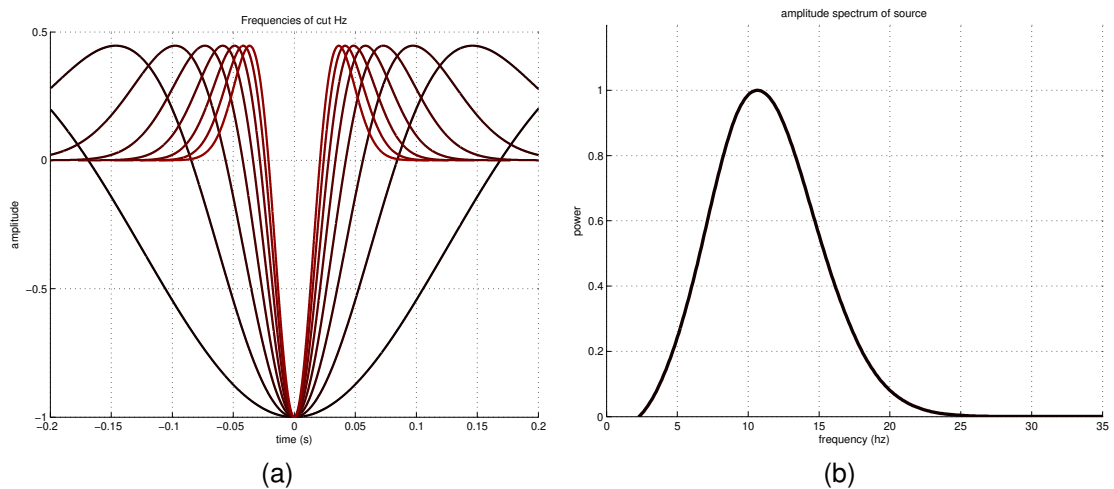
A *wavelet* Ricker possui um pico central e dois lobos menores, possui fase zero e pode ser definida através de um único parâmetro, f_{corte} . A Ricker pode ser obtida através do cálculo da segunda derivada de uma função gaussiana (RYAN, 1994). Portanto, o termo fonte pode ser descrito pela seguinte expressão:

$$w(\mathbf{x}, t) = \rho(\mathbf{x}) \nabla \cdot \mathbf{f}(\mathbf{x}, t) = A_{max} [2\pi(\pi f_c t_d)^2 - 1] e^{-\pi(\pi f_c t_d)^2} \delta(\mathbf{x} - \mathbf{x}_f) \quad (\text{A.57})$$

onde f_c é a frequência central, que pode ser escrita em termos da frequência de corte ($f_c = \frac{f_{corte}}{3\sqrt{\pi}}$); t_d é o tempo deslocado para que a *wavelet* seja centrada em zero. A_{max} é a amplitude máxima e $\mathbf{x}_f$ a posição da fonte.

A Figura 87a apresenta diversos exemplos com *wavelets* Ricker, com frequências de corte de 4 até 32 Hz, e a Figura 87b apresenta o espectro de amplitude normalizado para a Ricker com frequência de corte de 32 Hz.

Figura 86 – a) Exemplos com diversas *wavelets* Ricker com frequências de corte de 4 até 32 Hz. Os tons de vermelho indicam as frequências mais altas. b) Espectro de amplitude da *wavelet* Ricker com frequência de corte de 32 Hz



Fonte: o autor.

A.4.2 A wavelet B-spline

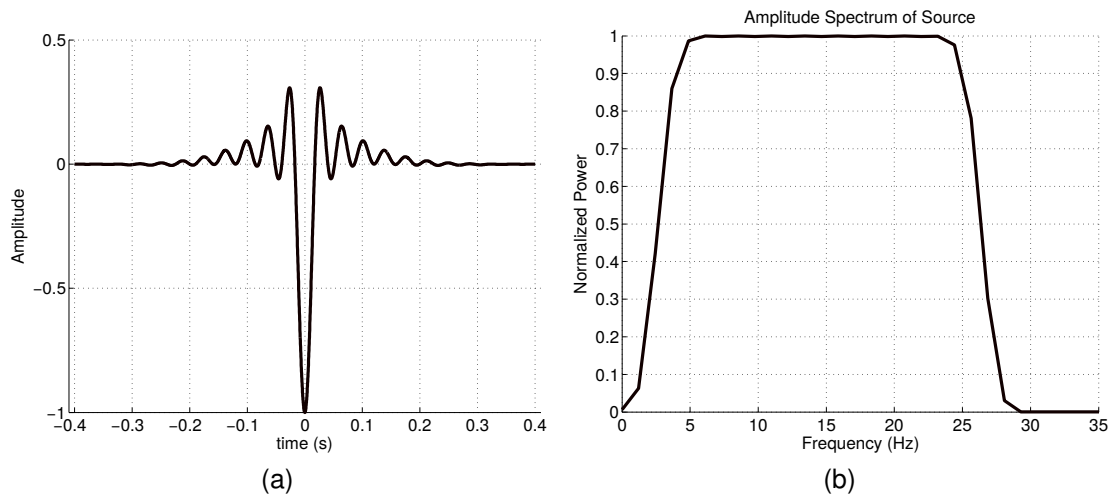
A *wavelet Wide-Band B-spline* foi proposta por Cao e Han (2011) e possui a seguinte expressão:

$$w(\mathbf{x}, t) = \frac{1}{q-p} \sqrt{f_b} \left(\text{sinc} \left(\frac{f_b \cdot t}{m} \right) \right)^m (q \text{sinc}(2qt) - p \text{sinc}(2pt)) \delta(\mathbf{x} - \mathbf{x}_f). \quad (\text{A.58})$$

A Equação A.58 foi baseada na teoria da decomposição no domínio *wavelet* e sua respectiva reconstrução no domínio do tempo. Essa *wavelet* foi formulada em função de 4 parâmetros: p (frequência inicial), q (frequência final), m (ordem) e f_b (largura da banda), que definem seu conteúdo de frequência com precisão.

Com essa expressão é possível criar *wavelets* com uma banda de frequência larga, como apresentado na Figura 87. Nesse caso foram utilizados os seguintes parâmetros: $p = 2$ Hz, $q = 27$ Hz, $m = 40$ e $f_b = 25$ Hz.

Figura 87 – a) *Wavelet Wide-Band B-spline*. b) Espectro de amplitude da *wavelet Wide-Band B-spline*. Foram utilizados os seguintes parâmetros: $p = 2$ Hz, $q = 27$ Hz, $m = 40$ e $f_b = 25$ Hz.



Fonte: o autor.

Também é possível criar *wavelets* com o espectro de amplitude estreito, como na Figura 88. Nesse caso foram utilizados os parâmetros: $p = 4$ Hz, $q = 6$ Hz, $m = 1$ e $f_b = 1$ Hz.

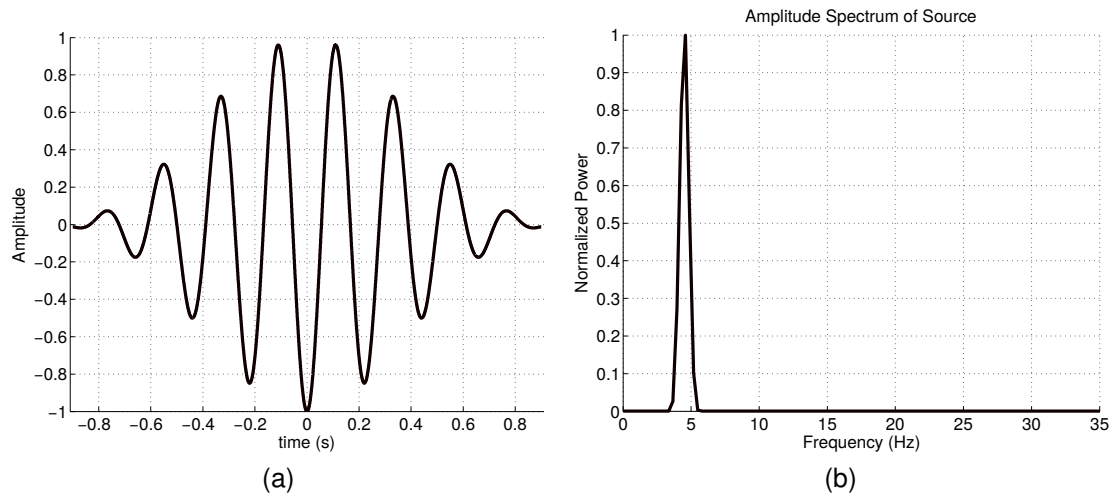


Figura 88 – a) *Wavelet Wide-Band B-spline*. b) Espectro de frequência da *wavelet Wide-Band B-spline*. Foram utilizados os parâmetros: $p = 4$ Hz, $q = 6$ Hz, $m = 1$ e $f_b = 1$ Hz.

APÊNDICE B – Implementação computacional

Nesse capítulo serão apresentados alguns aspectos sobre a implementação da equação da onda acústica. Será apresentada a discretização do modelo de propriedades, e também, como solucionar a equação da onda utilizando o método das diferenças finitas. Sabendo que a discretização da equação da onda pode ser realizada no domínio do tempo, a solução é obtida através dos operadores de marcha no tempo. Vale ressaltar que a discretização também pode ser feita no domínio da frequência, entretanto não será abordada nesse estudo. Em seguida, serão discutidos também os critérios de não-dispersão e estabilidade que garantem o correto funcionamento do algoritmo de modelagem sísmica.

Devido à definição do domínio computacional finito para a solução da equação da onda, são introduzidas bordas artificiais no modelo de propriedades. Essas bordas produzem reflexões espúrias que não condizem com o experimento físico. Para atenuar esses artefatos são introduzidas modificações na equação da onda na região das bordas artificiais, ou seja, são modificadas as condições de contorno. Nesse capítulo são apresentadas duas condições de borda, camada não-reflexiva e camadas absorcivas.

B.1 O método das diferenças finitas

O método das diferenças finitas aplicado à problemas sísmicos começou a ter uma significativa relevância a partir da década de 1960, devido ao aumento crescente do poder de processamento e memória dos computadores. No contexto da propagação de ondas, uma das principais vantagens do método das diferenças finitas sobre os métodos assintóticos, como por exemplo o traçamento de raios e a equação eikonal, é que é possível descrever o comportamento ondulatório de um meio com qualquer variação espacial das propriedades elásticas. Entretanto, o método de diferenças finitas possui um alto custo computacional (LEVANDER, 1989).

No método de diferenças finitas, o domínio computacional é representado por uma malha espacial e temporal, onde os pontos que constituem a malha recebem os valores do campo de onda, ou das propriedades do meio (MOCZO; ROBERTSSON; EISNER, 2007). Usualmente a malha estruturada (*structured grid*) possui espaçamento uniforme $\Delta x = \Delta y = \Delta z = h$.

B.1.1 Aproximação de derivadas por diferenças finitas

Basicamente, o operador de diferenças finitas aproxima uma derivada em uma diferença entre dois pontos adjacentes da função $f(x)$:

$$\frac{df(x)}{dx} \approx \frac{f(x_{i+1}) - f(x_i)}{\Delta x} \quad (\text{B.1})$$

onde Δx é a distância entre dois pontos da malha.

Ainda segundo Moczo, Robertsson e Eisner (2007), a aproximação das derivadas pode ser progressiva:

$$\frac{df}{dx}(x_i) = \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x} + O(\Delta x^1), \quad (\text{B.2a})$$

regressiva:

$$\frac{df}{dx}(x_i) = \frac{f(x_i) - f(x_i - \Delta x)}{\Delta x} + O(\Delta x^1), \quad (\text{B.2b})$$

ou central:

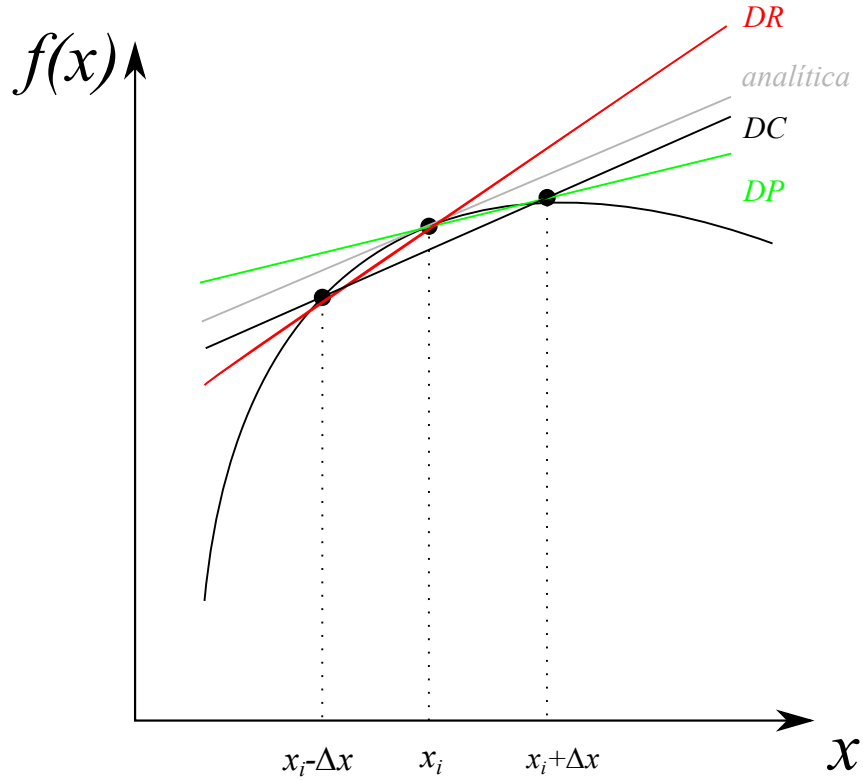
$$\frac{df}{dx}(x_i) = \frac{f(x_i + \Delta x) - f(x_i - \Delta x)}{2\Delta x} + O(\Delta x^2). \quad (\text{B.2c})$$

O termo O representa o erro da aproximação. A Figura 89 ilustra as aproximações por diferenças finitas geometricamente.

De forma geral, o operador de diferenças finitas pode ser obtido a partir da expansão da série de Taylor de uma função $f(x)$, em torno do ponto x_0 :

$$f(x_0) = \sum_{n=0}^{\infty} \frac{1}{n!} \left. \frac{d^n}{dx^n} f(x) \right|_{x=x_0} (x - x_0)^n, \quad (\text{B.3})$$

Figura 89 – Representação geométrica das diferenças finitas. As retas DP, DC e DR são, respectivamente, as diferenças progressivas, centrais e regressivas.



assumindo que $f(x)$ é discretizada em uma malha regular, é possível escrever a expansão em série de Taylor da seguinte forma:

$$\begin{aligned}
 f_{i\pm l} &= f(x_i \pm l\Delta x) \\
 &= \sum_{n=0}^{\infty} \frac{1}{n!} \left. \frac{d^n}{dx^n} f(x) \right|_{x=x_i} (\pm l\Delta x)^n \\
 &= f(x_i) + \frac{df(x_i)}{dx} (\pm l\Delta x) + \frac{1}{2} \frac{d^2 f(x_i)}{dx^2} (\pm l\Delta x)^2 + \frac{1}{6} \frac{d^3 f(x_i)}{dx^3} (\pm l\Delta x)^3 + \dots
 \end{aligned} \tag{B.4}$$

Para obter a expressão de diferença finita central é necessário calcular a aproximação da derivada usando os pontos adjacentes ao ponto x_i , fazendo $l = 1$ na Equação B.4:

$$\begin{aligned}
 f_{i+1} &= f_i + \frac{df_i}{dx} \Delta x + \frac{1}{2} \frac{d^2 f_i}{dx^2} \Delta x^2 + \frac{1}{6} \frac{d^3 f_i}{dx^3} \Delta x^3 + \frac{1}{24} \frac{d^4 f_i}{dx^4} \Delta x^4 + \dots \\
 f_{i-1} &= f_i - \frac{df_i}{dx} \Delta x + \frac{1}{2} \frac{d^2 f_i}{dx^2} \Delta x^2 - \frac{1}{6} \frac{d^3 f_i}{dx^3} \Delta x^3 + \frac{1}{24} \frac{d^4 f_i}{dx^4} \Delta x^4 + \dots
 \end{aligned} \tag{B.5}$$

subtraindo as equações anteriores:

$$\begin{aligned}
 f_{i+1} - f_{i-1} &= 2\frac{df_i}{dx}\Delta x + 2\frac{1}{6}\frac{d^3f_i}{dx^3}\Delta x^3 + 2\frac{1}{120}\frac{d^5f_i}{dx^5}\Delta x^5 \dots \\
 \frac{df_i}{dx} &= \frac{f_{i+1} - f_{i-1}}{2\Delta x} - \underbrace{\frac{1}{2\Delta x} \left(2\frac{1}{6}\frac{d^3f_i}{dx^3}\Delta x^3 + 2\frac{1}{120}\frac{d^5f_i}{dx^5}\Delta x^5 \dots \right)}_{O(\Delta x^2)} \\
 \frac{df_i}{dx} &= \frac{f_{i+1} - f_{i-1}}{2\Delta x} + O(\Delta x^2)
 \end{aligned} \tag{B.6}$$

Como Δx é pequeno, os termos que possuem expoente maior que 2 são negligenciados e assumidos como erro da aproximação $O(\Delta x^2)$. Desprezar os termos a partir de Δx^2 leva a uma aproximação cujo o erro da aproximação é de segunda ordem. Se aumentar a ordem do erro de aproximação manipulando a expansão da série de Taylor, a aproximação por diferenças finitas se torna mais precisa. Em contrapartida, a aproximação se torna mais custosa computacionalmente, por utilizar mais pontos da malha discretizada.

B.1.2 Aproximação da 2ª derivada por diferenças finitas

De maneira análoga, para a obtenção da aproximação por diferenças finitas da 2ª derivada de uma função basta realizar a soma das Equações B.5:

$$\begin{aligned}
 f_{i+1} + f_{i-1} &= 2f_i + 2\frac{1}{2}\frac{d^2f_i}{dx^2}\Delta x^2 + 2\frac{1}{24}\frac{d^4f_i}{dx^4}\Delta x^4 + 2\frac{1}{720}\frac{d^6f_i}{dx^6}\Delta x^6 + \dots \\
 \frac{d^2f_i}{dx^2} &= \frac{f_{i+1} - 2f_i + f_{i-1}}{\Delta x^2} - \underbrace{\frac{1}{\Delta x^2} \left(2\frac{1}{24}\frac{d^4f_i}{dx^4}\Delta x^4 + 2\frac{1}{720}\frac{d^6f_i}{dx^6}\Delta x^6 + \dots \right)}_{O(\Delta x^2)} \\
 \frac{d^2f_i}{dx^2} &= \frac{f_{i+1} - 2f_i + f_{i-1}}{\Delta x^2} + O(\Delta x^2),
 \end{aligned} \tag{B.7}$$

e assim obter a aproximação de 2ª ordem.

Para aproximações de ordens superiores são necessários mais pontos da malha discretizada. Por exemplo, para a aproximação por diferença finita da 2ª derivada de uma função qualquer com erro de 4ª ordem, serão necessários mais dois pontos, além

daqueles utilizados na aproximação de 2ª ordem. Para isso, é assumido $l = 2$ na Equação B.4:

$$\begin{aligned} f_{i+2} &= f_i + \frac{df_i}{dx} 2\Delta x + \frac{1}{2} \frac{d^2 f_i}{dx^2} 2^2 \Delta x^2 + \frac{1}{6} \frac{d^3 f_i}{dx^3} 2^3 \Delta x^3 + \frac{1}{24} \frac{d^4 f_i}{dx^4} 2^4 \Delta x^4 + \dots \\ f_{i-2} &= f_i - \frac{df_i}{dx} 2\Delta x + \frac{1}{2} \frac{d^2 f_i}{dx^2} 2^2 \Delta x^2 - \frac{1}{6} \frac{d^3 f_i}{dx^3} 2^3 \Delta x^3 + \frac{1}{24} \frac{d^4 f_i}{dx^4} 2^4 \Delta x^4 + \dots \end{aligned} \quad (\text{B.8})$$

e como anteriormente, somando as duas equações:

$$\begin{aligned} f_{i+2} + f_{i-2} &= 2f_i + 2 \frac{1}{2} \frac{d^2 f_i}{dx^2} 2^2 \Delta x^2 + 2 \frac{1}{24} \frac{d^4 f_i}{dx^4} 2^4 \Delta x^4 + \dots \\ \frac{d^2 f_i}{dx^2} &= \frac{f_{i+2} - 2f_i + f_{i-2}}{4\Delta x^2} - \underbrace{\frac{1}{4\Delta x^2} \left(2 \frac{1}{24} \frac{d^4 f_i}{dx^4} 2^4 \Delta x^4 + 2 \frac{1}{720} \frac{d^6 f_i}{dx^6} 2^6 \Delta x^6 \right)}_{O(\Delta x^2)}. \end{aligned} \quad (\text{B.9})$$

Analisando os termos correspondentes aos erros nas Equações B.7 e B.9, é possível perceber que o termo que contém Δx^4 pode ser eliminado, multiplicando a Equação B.7 por 4 e subtraindo-a da Equação B.9. O que resulta em:

$$\frac{d^2 f_i}{dx^2} = \frac{4}{3} \frac{f_{i+1} - 2f_i + f_{i-1}}{\Delta x^2} - \frac{1}{3} \frac{f_{i+2} - 2f_i + f_{i-2}}{4\Delta x^2} + O(\Delta x^4), \quad (\text{B.10})$$

onde o termo $O(\Delta x^4)$ indica a ordem da aproximação. Por fim, simplificando a equação anterior obtemos a aproximação da 2ª derivada, com erro de 4ª ordem, dada por:

$$\frac{d^2 f_i}{dx^2} = \frac{-f_{i+2} + 16f_{i+1} - 30f_i + 16f_{i-1} - f_{i-2}}{12\Delta x^2} + O(\Delta x^4). \quad (\text{B.11})$$

As aproximações de ordens superiores podem ser calculadas através do mesmo procedimento. A tabela, a seguir, apresenta os coeficientes para a obtenção dos operadores de diferenças finitas até a oitava ordem, para as primeiras e segundas derivadas.

Tabela 10 – Coeficientes dos operadores de diferença central. Modificado de Bartolo (2013)

Derivada	Precisão	-4	-3	-2	-1	0	1	2	3	4
1	2				-1/2	0	1/2			
	4			1/12	-2/3	0	2/3	-1/12		
	6		-1/60	3/20	-3/4	0	3/4	-3/20	1/60	
	8	1/280	-4/105	1/5	-4/5	0	4/5	-1/5	4/105	-1/280
2	2				1	-2	1			
	4			-1/12	4/3	-5/2	4/3	-1/12		
	6		1/90	-3/20	3/2	-49/18	3/2	-3/20	1/90	
	8	-1/560	8/315	-1/5	8/5	205/72	8/5	-1/5	8/315	-1/560

B.2 A discretização da equação da onda 2D

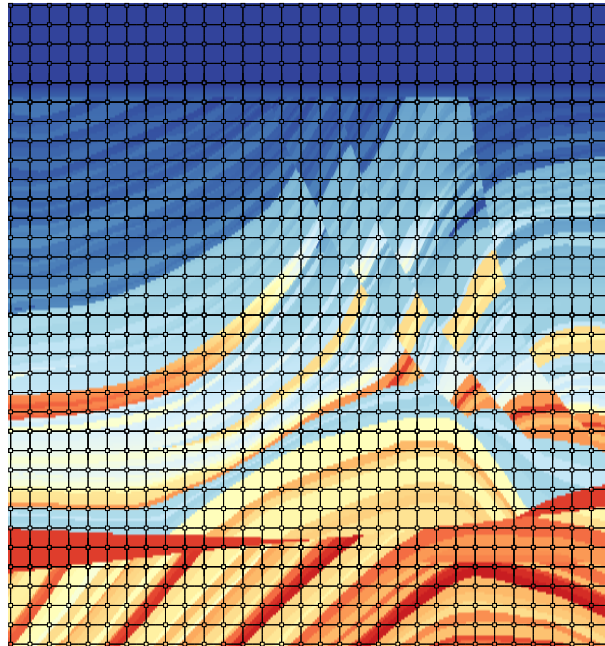
A solução da equação da onda utilizando um método direto, como a aproximação por diferença finitas, exige que o modelo geológico seja discretizado em uma malha regular (Figura 90), com um número finito de pontos (CARCIONE; HERMAN; KROODE, 2002). Cada nó da malha contém o valor da propriedade do modelo e está relacionado com sua respectiva coordenada espacial, no caso 2D, $\mathbf{x}_i = (x_i, z_j) = (i\Delta x, j\Delta z)$. Em geral, a malha estruturada é regular, ou seja, o espaçamento entre os nós da malha é fixo $\Delta x = \Delta z = h$.

No caso acústico, os nós da malha também estão relacionados aos campos de pressão, assim como o campo de velocidade $v_p(x, z)$. A convenção de índices adotada para identificar uma função qualquer, por exemplo o campo de pressão, discretizada no tempo e no espaço será: $p(x, z, t) \rightarrow u(x_i, z_j, t_m) = u_{ij}^m$.

B.2.1 A derivada temporal

Para obter a solução da equação da onda acústica, pelo método das diferenças finitas, é necessário substituir o termo que contém a segunda derivada temporal pela aproximação por diferenças finitas de segunda ordem (Equação B.7). No caso da simulação do experimento sísmico, a aplicação da aproximação de segunda ordem possui acurácia suficiente, como aponta Virieux et al. (2006). O aumento da ordem do operador de diferenças finitas temporal implica no aumento dos campos que precisam ser armazenados na memória para a aplicação da marcha no tempo.

Figura 90 – Representação da discretização de um modelo geológico (Marmousi). Em cada nó do *grid* são armazenados os valores do modelo de propriedade e os campos do fenômeno ondulatório.



Para a aplicação da aproximação de segunda ordem da derivada temporal é necessário o armazenamento de 3 campos de pressão durante a solução da equação da onda¹: o passado p_{ij}^{m-1} , o presente p_{ij}^m e o futuro p_{ij}^{m+1} , em todos os pontos do modelo. Assim, a aproximação da derivada temporal fica:

$$\frac{\partial^2 p(x, z, t)}{\partial t^2} = \frac{p_{i,k}^{m+1} - 2p_{i,k}^m + p_{i,k}^{m-1}}{\Delta t^2} + O(\Delta t^2). \quad (\text{B.12})$$

B.2.2 A derivada espacial

As aproximações das derivadas espaciais (Laplaciano), utilizadas nesse trabalho, foram as de quarta e oitava ordem. Em 2D, as aproximações de 4^a e 8^a ordem das 2^a derivadas espaciais do campo de pressão necessitam acessar 9 e 17 posições do *grid* (*stencil*), respectivamente, para calcular a estimativa da derivada no ponto (i,k) do *grid*.

¹ Na implementação pode-se empregar área na memória para apenas 2 campos.

No caso da aproximação de 4ª:

$$\frac{\partial^2 p(x, z, t)}{\partial x^2} = \frac{-p_{i+2,k}^m + 16p_{i+1,k}^m - 30p_{i,k}^m + 16p_{i-1,k}^m - p_{i-2,k}^m}{12h^2} + O(\Delta x^4)$$

$$\frac{\partial^2 p(x, z, t)}{\partial z^2} = \frac{-p_{i,k+2}^m + 16p_{i,k+1}^m - 30p_{i,k}^m + 16p_{i,k-1}^m - p_{i,k-2}^m}{12h^2} + O(\Delta z^4),$$
(B.13)

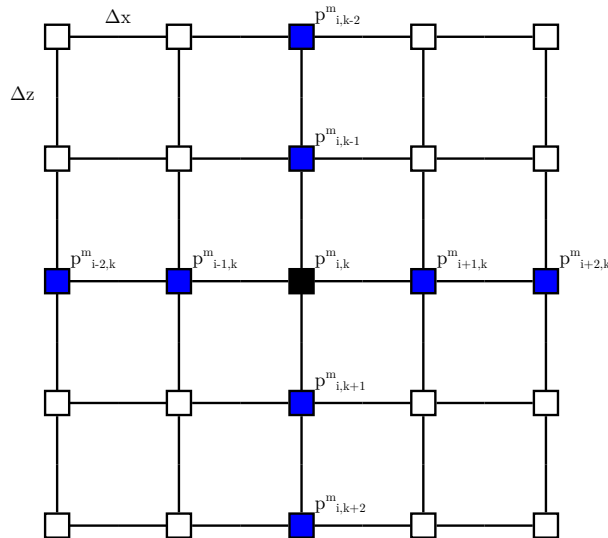
e no caso de 8ª ordem:

$$\frac{\partial^2 p(x, z, t)}{\partial x^2} = \frac{-9p_{i+4,k}^m + 128p_{i+3,k}^m - 1008p_{i+2,k}^m + 8064p_{i+1,k}^m - 14350p_{i,k}^m + 8064p_{i-1,k}^m - 1008p_{i-2,k}^m + 128p_{i-3,k}^m - 9p_{i-4,k}^m}{5040h^2} + O(\Delta x^8).$$

$$\frac{\partial^2 p(x, z, t)}{\partial z^2} = \frac{-9p_{i,k+4}^m + 128p_{i,k+3}^m - 1008p_{i,k+2}^m + 8064p_{i,k+1}^m - 14350p_{i,k}^m + 8064p_{i,k-1}^m - 1008p_{i,k-2}^m + 128p_{i,k-3}^m - 9p_{i,k-4}^m}{5040h^2} + O(\Delta z^8).$$
(B.14)

A Figura 91 mostra as posições do *grid* utilizadas para o cálculo do operados de diferenças finitas espaciais de 4ª ordem em 2D.

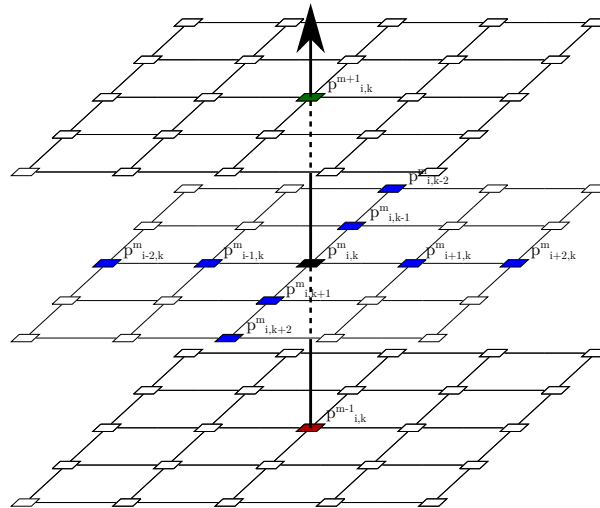
Figura 91 – Representação do *stencil* para a aproximação de 4ª ordem da 2ª derivada espacial.



B.2.3 O esquema de marcha no tempo

A substituição dos operadores temporais (Equação B.12) e espaciais (Equação B.13), na equação da onda acústica, fornece o esquema de marcha temporal. Esse esquema calcula o campo de pressão para todos os instantes de tempo para todas as posições do modelo. O operador de marcha no tempo também é utilizado para o cálculo do campo adjunto $p^\dagger$, que faz parte da expressão que fornece o gradiente da função objetivo (Equação 2.85).

Figura 92 – Representação de uma malha espacial e temporal no domínio computacional que discretiza os campos contínuos. Nesse exemplo o campo contínuo é discretizado usando duas dimensões espaciais e uma dimensão temporal.



Utilizando a equação da onda acústica (Equação A.56) em duas dimensões na ausência de fonte:

$$\left[\frac{\partial^2 p}{\partial x^2}(x, z, t) + \frac{\partial^2 p}{\partial z^2}(x, z, t) \right] - \frac{1}{v_p^2(x, z)} \frac{\partial^2 p}{\partial t^2}(x, z, t) = 0, \quad (\text{B.15})$$

e substituindo as aproximações das derivadas temporais e espaciais (Equações B.12 e

B.13), a discretização da equação da onda fica:

$$\left[\frac{-p_{i+2,k}^m + 16p_{i+1,k}^m - 30p_{i,k}^m + 16p_{i-1,k}^m - p_{i-2,k}^m}{12h^2} + \frac{-p_{i,k+2}^m + 16p_{i,k+1}^m - 30p_{i,k}^m + 16p_{i,k-1}^m - p_{i,k-2}^m}{12h^2} \right] - \frac{1}{v_{p,i,k}^2} \frac{p_{i,k}^{m+1} - 2p_{i,k}^m + p_{i,k}^{m-1}}{\Delta t^2} = 0 \quad (\text{B.16})$$

Reescrevendo a equação anterior B.16, e isolando o termo correspondente ao campo futuro $p_{i,k}^{m+1}$, obtém-se o operador de marcha no tempo para modelagem da equação da acústica, também representado na figura 92:

$$p_{i,k}^{m+1} = 2p_{i,k}^m - p_{i,k}^{m-1} + \frac{1}{12} \frac{\Delta t^2 v_{p,i,k}^2}{h^2} [16(p_{i+1,k}^m + p_{i-1,k}^m + p_{i,k+1}^m + p_{i,k-1}^m) - (p_{i+2,k}^m + p_{i-2,k}^m + p_{i,k+2}^m + p_{i,k-2}^m) - 60p_{i,k}^m]. \quad (\text{B.17})$$

B.3 Critérios de dispersão e estabilidade

Para garantir que a solução numérica da equação da onda discretizada por diferenças finitas não se degrade devido a aproximação dos operadores de derivada espacial, é necessário impor uma condição que impeça os efeitos de dispersão numérica. Essa condição é dada por (LINES; SLAWINSKI; BORDING, 1999):

$$h \leq \frac{v_{pmin}}{\alpha f_{cut}} \quad (\text{B.18})$$

onde h é o espaçamento da malha, que garante a não-dispersão numérica da modelagem, f_{cut} é a frequência de corte da fonte sísmica usada para simulação, v_{pmin} é a velocidade mínima e α é um parâmetro que varia de acordo com a ordem da aproximação por diferenças finitas utilizada. Para aproximações de segunda ordem $\alpha = 10$, e para aproximações de quarta ordem $\alpha = 5$. O aumento da ordem do operador de derivada espacial permite a utilização de espaçamentos h maiores, melhorando a performance computacional.

O critério de estabilidade numérica garante que os erros de aproximação não se amplifiquem ao longo do tempo. Esse critério depende da escolha do espaçamento do *grid*, sendo limitado pelo critério de dispersão B.18, e definido como:

$$\Delta t \leq \frac{h}{\beta v_{pmax}}, \quad (\text{B.19})$$

onde Δt é o intervalo de tempo, h é o espaçamento da malha, v_{pmax} é a velocidade máxima e β é um parâmetro intrínseco. A escolha de $\beta = 4$ garante estabilidade para o operador de derivada temporal.

B.4 Simulando um meio infinito

De acordo com Engquist e Majda (1977), a introdução de bordas artificiais é essencial para resolver a equação da onda numericamente de forma correta. A solução numérica das equações diferenciais exige o estabelecimento de um domínio computacional finito, que implica em uma barreira artificial, definida pela limitação dos recursos computacionais disponíveis. Essas barreiras artificiais geram reflexões indesejadas nos sismogramas sintéticos. Para atenuar os efeitos das reflexões artificiais são impostas condições de fronteira que simulam um meio infinito.

B.4.1 Bordas não-reflexivas

As condições de contorno (ou de borda) de Dirichlet assumem que o campo de pressão é igual a zero nas bordas do domínio computacional:

$$p(x = x_L, z = z_L, t) = 0. \quad (\text{B.20})$$

Essa condição de contorno da equação diferencial insere barreiras artificiais nas bordas do modelo computacional, implicando em reflexões indesejadas durante a etapa da modelagem da equação da onda. Na Equação B.20, x_L e z_L representam as coordenadas espaciais da borda do domínio computacional.

Uma das formas de contornar as reflexões espúrias, provocadas pela fronteira artificial, é utilizar a equação da onda *one-way* para modificar as condições de contorno do problema. Essa abordagem foi proposta inicialmente por Reynolds (1978), assim a equação da onda *one-way* é obtida através da fatoração do operador diferencial da equação da onda.

Sabendo que a equação da onda acústica (Equação A.56), também conhecida como equação da onda *two-way*, pode ser reescrita como:

$$\left[\frac{1}{v_p^2(x, z)} \frac{\partial^2}{\partial t^2} - \nabla^2 \right] p(x, z, t) = 0,$$

a fatoração do operador diferencial leva a:

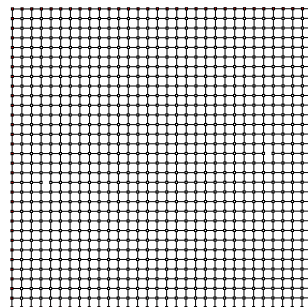
$$\overbrace{\left[\frac{1}{v_p(x, z)} \frac{\partial}{\partial t} - \nabla \right]}^{\text{Operador One Way}} \overbrace{\left[\frac{1}{v_p(x, z)} \frac{\partial}{\partial t} + \nabla \right]}^{\text{Operador One Way}} p(x, z, t) = 0. \quad (\text{B.21})$$

Assim, a equação da onda acústica *one-way* pode ser escrita como:

$$\left[\frac{1}{v_p(x, z)} \frac{\partial}{\partial t} \pm \nabla \right] p(x, z, t) = 0. \quad (\text{B.22})$$

Em duas dimensões, as condições de contorno são assumidas como ondas planas viajando em apenas uma direção. Dessa forma, cada borda (superior, inferior, esquerda e direita) possui uma condição de contorno que simula essa situação. A Figura 93 indica a posição onde a equação da onda é modificada para simular o meio infinito.

Figura 93 – Representação da malha de modelagem indicando em vermelho as posições da malha onde a equação da onda será modificada para simular um meio infinito (bordas não reflexivas).



A equação da onda aplicada na borda esquerda ($x = 1$ e $1 \leq z \leq N_z$), fica:

$$\left[\frac{1}{v_p(x, z)} \frac{\partial}{\partial t} - \frac{\partial}{\partial x} \right] p(x, z, t) = 0, \quad (\text{B.23})$$

a equação da borda direita ($x = N_x$ e $1 \leq z \leq N_z$):

$$\left[\frac{1}{v_p(x, z)} \frac{\partial}{\partial t} + \frac{\partial}{\partial x} \right] p(x, z, t) = 0, \quad (\text{B.24})$$

a equação da borda superior ($1 \leq x \leq N_x$ e $z = N_z$):

$$\left[\frac{1}{v_p(x, z)} \frac{\partial}{\partial t} - \frac{\partial}{\partial x} \right] p(x, z, t) = 0, \quad (\text{B.25})$$

a equação da borda inferior ($1 \leq x \leq N_x$ e $z = N_z$):

$$\left[\frac{1}{v_p(x, z)} \frac{\partial}{\partial t} + \frac{\partial}{\partial x} \right] p(x, z, t) = 0. \quad (\text{B.26})$$

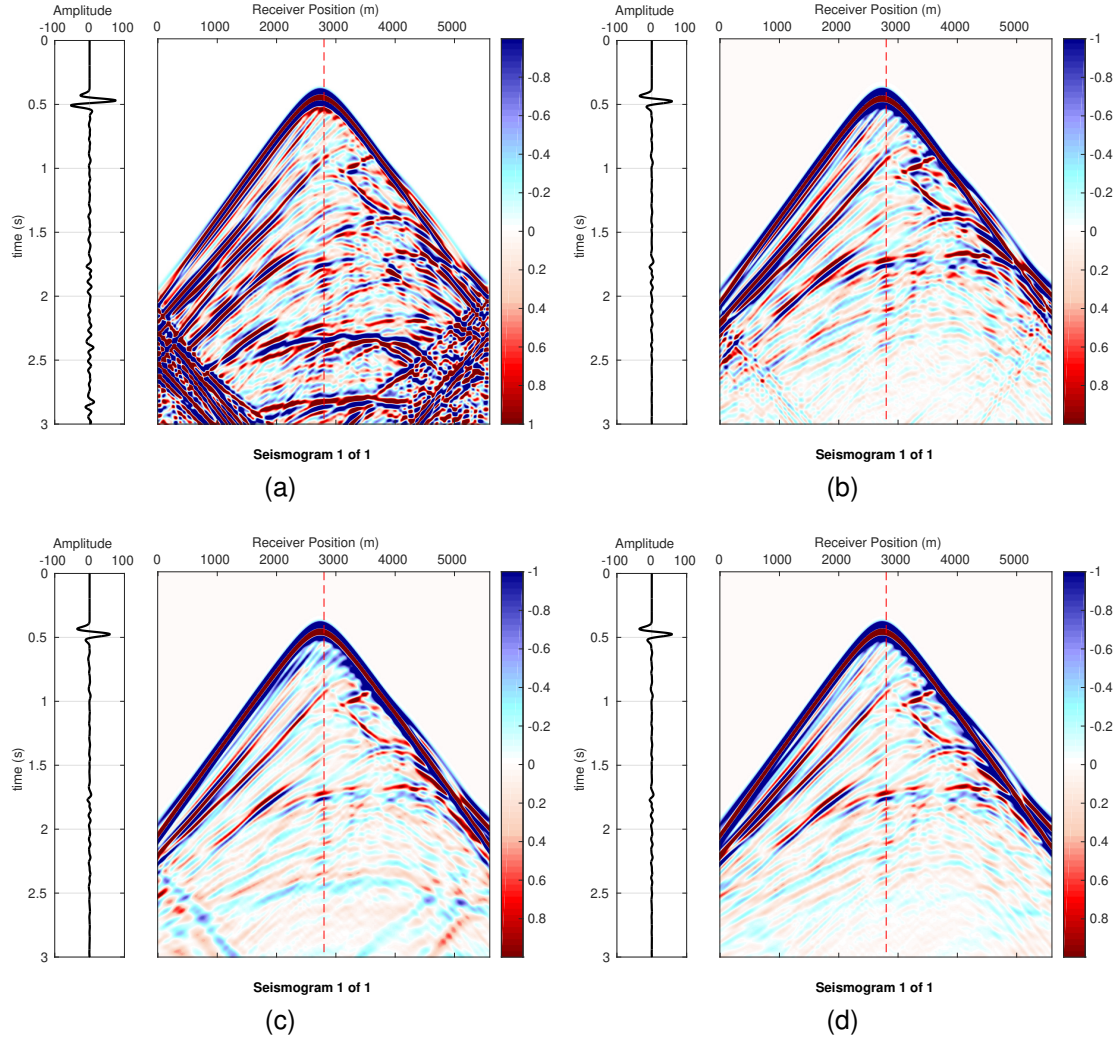
A discretização por diferenças finitas das condições de contorno pode ser feita utilizando o operador de diferenças progressivas ou regressivas de acordo com a borda. As Figuras 95a e 95b apresentam os sismogramas sintéticos de um único tiro sem a aplicação de borda absorciva e com a aplicação de bordas absorcivas, respectivamente.

B.4.2 Camada de absorção Cerjan

As bordas não-reflexivas não atenuam satisfatoriamente as reflexões provocadas pelo limite do modelo computacional, como é possível observar na Figura 95b. Para contornar esse problema, Cerjan et al. (1985) propôs uma camada absorciva que reduz gradativamente as amplitudes do campo de pressão, até chegar na borda do modelo, de forma que as reflexões espúrias provocadas pelas bordas são atenuadas significativamente.

Esse método é simples e robusto, porém sua eficácia depende do tamanho da camada de absorção. Tipicamente, para a aplicação desse método são adicionadas camadas ao redor do domínio válido (Figura 96a) para não atenuar ondas dentro do domínio computacional de interesse. O aumento da camada absorciva implica

Figura 94 – Exemplos de sismogramas sintéticos. a) Sem aplicação de bordas absorcivas. b) Aplicação de bordas absorcivas. c) Aplicação de camada absorciva do tipo Cerjan com 30 pontos. d) Aplicação de camada absorciva do tipo Cerjan com 100 pontos.



em uma melhor atenuação das reflexões indesejadas, porém o aumento do domínio computacional prejudica a performance da modelagem da equação da onda.

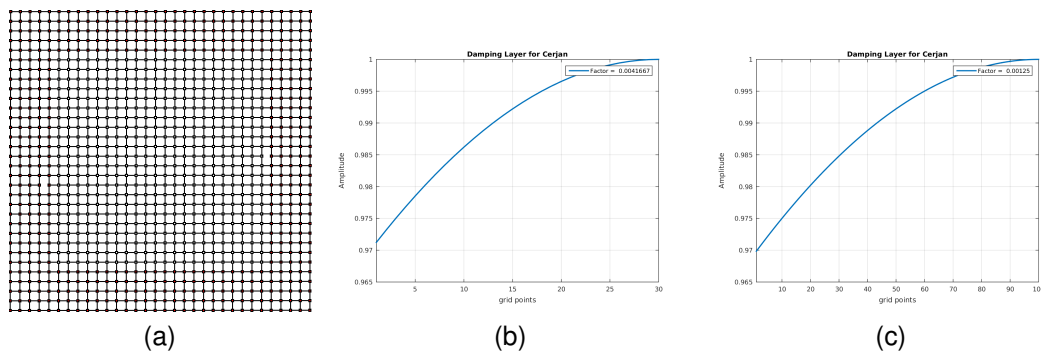
Ainda de acordo com Cerjan et al. (1985), a absorção do campo de pressão é realizada através da aplicação de uma função de decaimento, dentro da camada de absorção. A função de decaimento $d(x)$ deve ser suave e possui a seguinte expressão:

$$d(x) = e^{-fat(N_{amort}-x)^2} \quad (B.27)$$

onde fat é o fator de amortecimento, N_{amort} é o número de pontos da camada de

amortecimento utilizada e x é a posição dentro da camada de amortecimento. A Figura 95 apresenta o comportamento da função de decaimento para 30 e 100 pontos de camada de amortecimento.

Figura 95 – a) Representação do domínio computacional onde será resolvida a equação da onda. A região em vermelho indica onde a equação da onda será modificada para atenuar os efeitos das bordas. Essa camada é adicionada externamente ao domínio de interesse. Função de amortecimento utilizando uma camada de amortecimento de a) 30 pontos e b) 100 pontos.



Assim, dentro da camada de absorção, o campo de pressão é multiplicado pela função de decaimento de acordo com sua posição na camada de amortecimento. As Figuras 95c e 95d apresentam os exemplos utilizando 30 e 100 pontos de camada de amortecimento. Como é possível observar nesses exemplos, 30 pontos não foram suficientes para atenuar as reflexões espúrias. Entretanto, 100 pontos de camada atenuaram os artefatos provocados pelas bordas artificiais satisfatoriamente.

APÊNDICE C – Fluxogramas

Nessa seção são apresentados os fluxogramas que representam as implementações dos algoritmos desenvolvidos nessa tese.

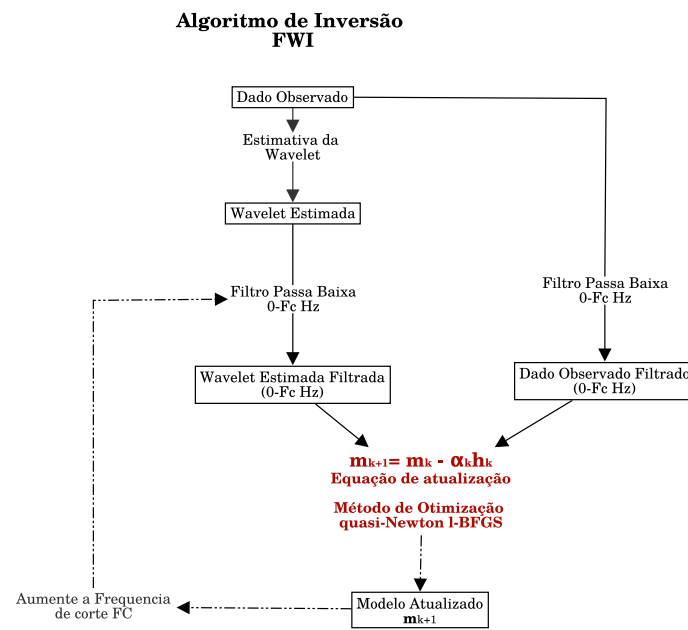


Figura 96 – Visão geral do algoritmo de FWI utilizando a abordagem multiescala no domínio da frequência.

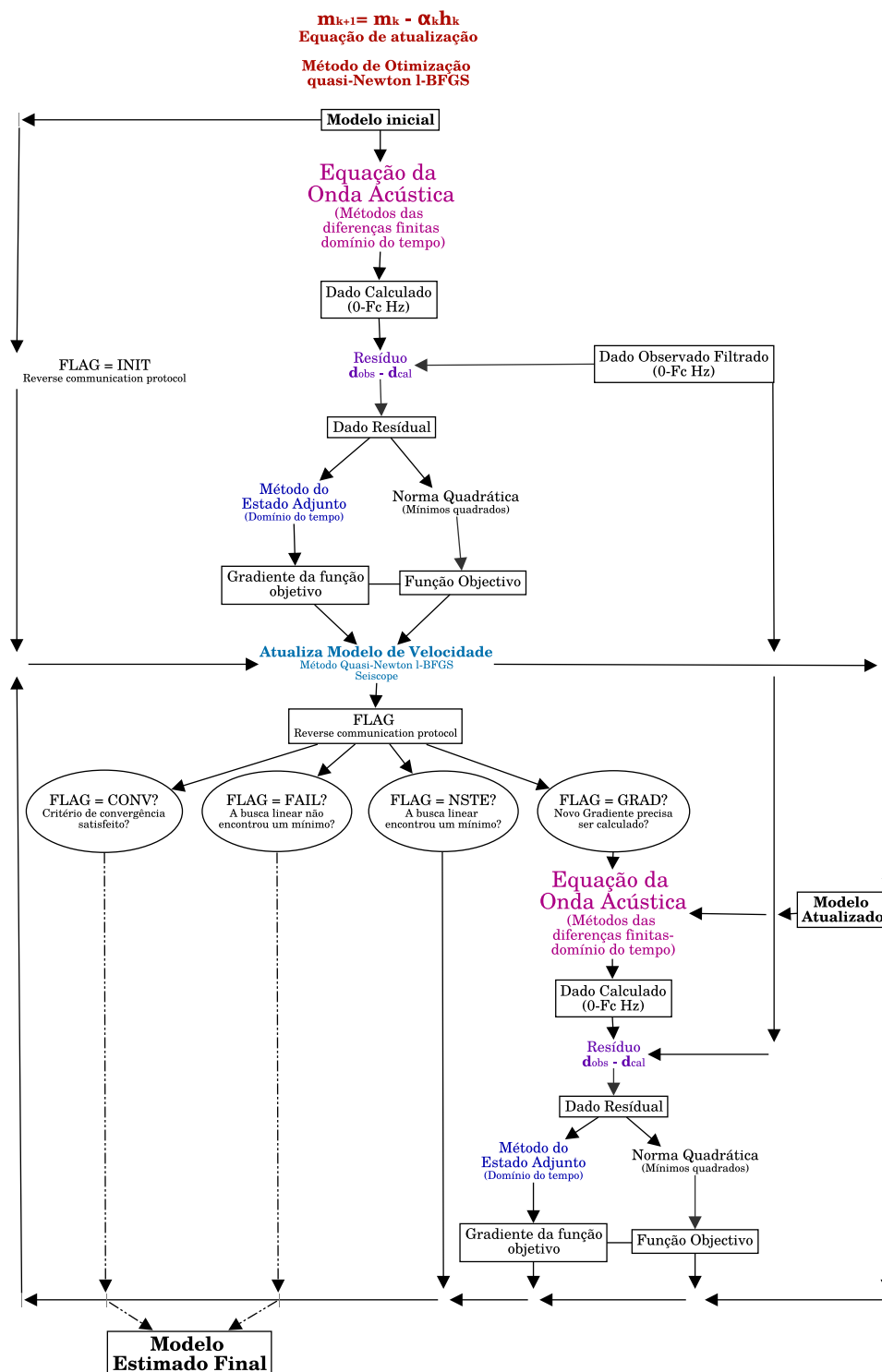


Figura 97 – Visão geral para o algoritmo implementado que atualiza o modelo de velocidade utilizando um método de otimização baseado no gradiente da função objetivo. O pacote disponibilizado pelo consórcio Seiscope foi utilizado para aplicação do método de otimização.

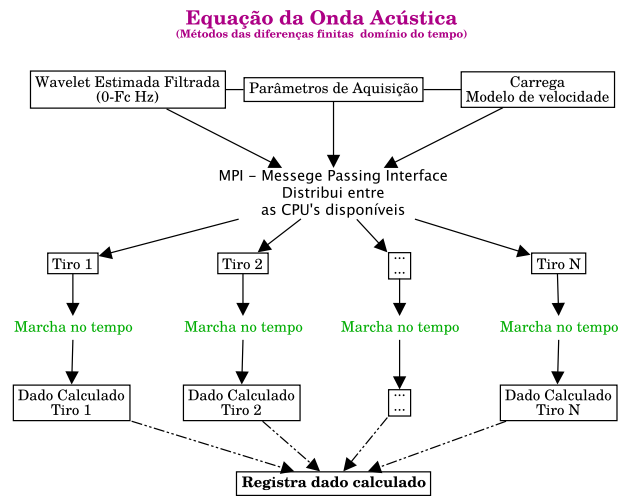


Figura 98 – Visão geral para o algoritmo implementado para a extrapolação do campo de onda para o caso acústico utilizando paralelização no domínio dos tiros.

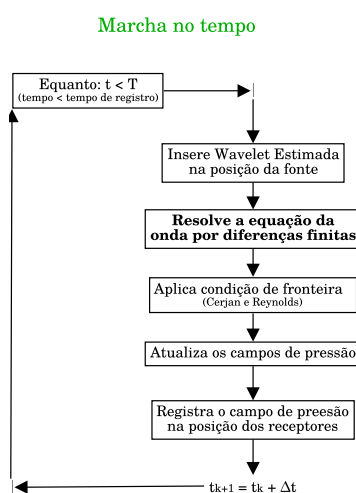


Figura 99 – Visão geral do algoritmo implementado de extrapolação do campo da onda

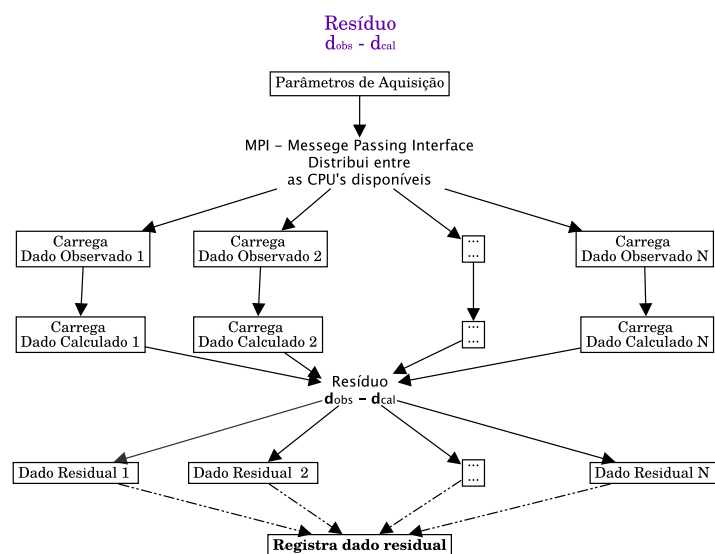


Figura 100 – Visão geral do algoritmo que calcula o resíduo utilizando a paralelização no domínio do tiro.

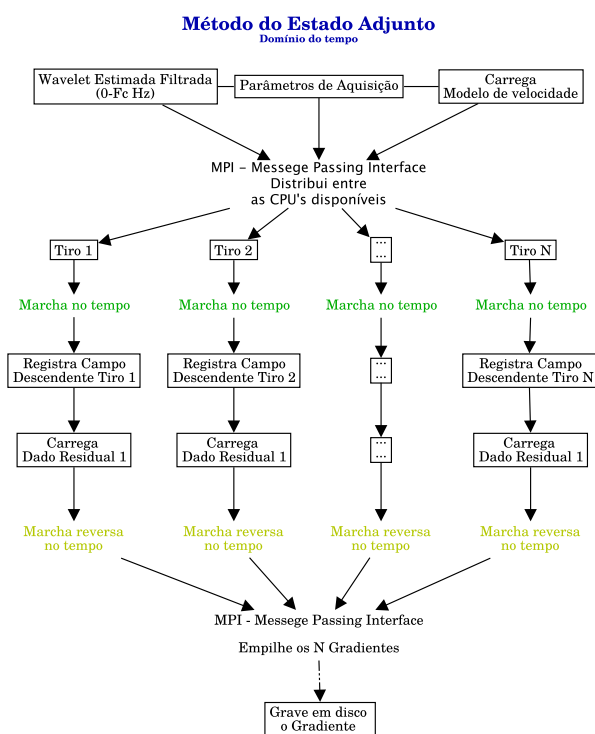


Figura 101 – Visão geral do algoritmo que calcula o gradiente da função objetivo através do método adjunto no domínio do tempo e utilizando a paralelização no domínio do tiro.

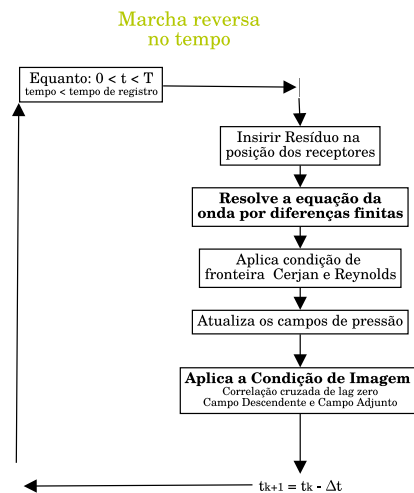


Figura 102 – Visão geral do algoritmo implementado de extrapolação da fonte adjunta (modelagem reversa no tempo).

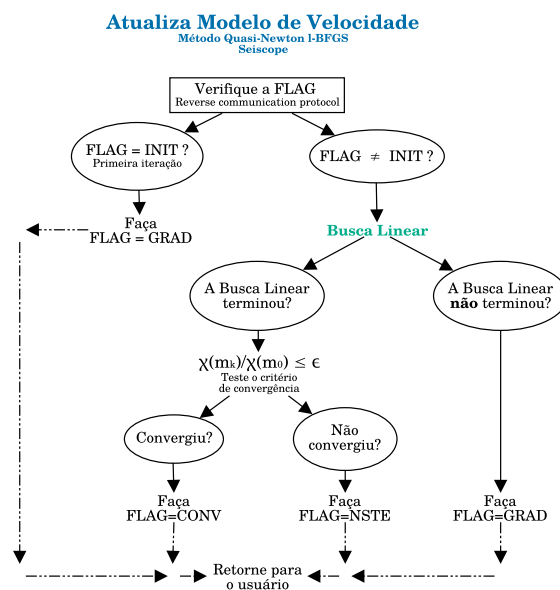


Figura 103 – Visão geral do algoritmo que atualiza o modelo através do método LBFGS. Pacote disponibilizado pelo consórcio Seiscope.

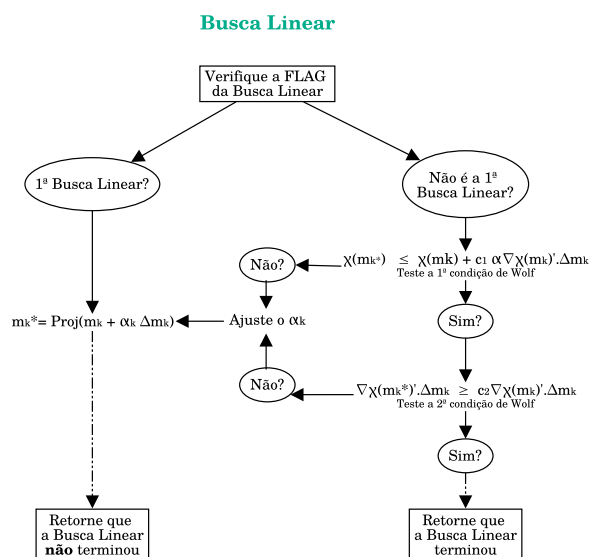


Figura 104 – Visão geral do algoritmo que realiza a busca linear. Algoritmo disponibilizado pelo consórcio Seiscope.

Anexos

ANEXO A – Wavenumbers illuminated by time-domain acoustic FWI using the L1 and L2 norms



Contents lists available at ScienceDirect

Journal of Applied Geophysics

journal homepage: www.elsevier.com/locate/jappgeo

Wavenumbers illuminated by time-domain acoustic FWI using the L1 and L2 norms

Felipe Timóteo da Costa^{a,*}, Marco Antonio Cetale Santos^a, Djalma Manoel Soares Filho^b

^a GLIS/DOT/Federal Fluminense University, Niterói, RJ, Brazil

^b CENPES/PETROBRAS, Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Article history:

Received 31 January 2019

Received in revised form 13 September 2019

Accepted 27 December 2019

Available online 03 January 2020

Keywords:

Seismic modeling

Seismic inversion

Optimization methods

FWI

ABSTRACT

The full waveform inversion (FWI) has been widely used for constructing high-resolution velocity models. The adjoint method was used to calculate the update direction and the optimization method L-BFGS was used to solve the nonlinear inverse problem. The wavenumbers retrieved by the gradient of the objective function were compared to the wavenumbers calculated analytically in the frequency domain. The good fittings achieved in the performed tests confirm the equivalence of calculating the gradient in the time or frequency domains. The time-domain acoustic FWI was applied to a 2D synthetic experiment and the 1D Fourier transform was used to quantify the wavenumbers illuminated by the L1 and L2 norms.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Recent advances in computer processing capabilities and acquisition of (low frequency) seismic data allowed nonlinear optimization methods to become increasingly popular to solve seismic problems with many parameters. Using optimization methods to solve such problems became feasible less than a decade ago due to the increasing computational power provided by the parallelization of the algorithms. As a result, techniques such as the full waveform inversion (FWI) stood out once again. Lailly (1983), Tarantola (1984) and Pratt (1999) published a few of the most important works for developing the FWI, contributing significantly to the resolution of nonlinear inverse problems of seismic data. More recently, Virieux and Operto (2009) provided an overview of this theory and presented the results obtained by Sirgue et al. (2010) for constructing a velocity model using only frequencies ranging from 3.0 Hz to 7.0 Hz while being able to identify geological structures in the velocity model directly. These results have significantly improved the depth migrated image of a reservoir in the North Sea. The FWI is a nonlinear optimization problem based on the fitting of the calculated data to the observed data. The solution to this problem allows constructing high quality and high-resolution velocity models (Ratcliffe et al., 2011). The calculated data are obtained through a modeling operator that uses the full waveform equation.

Seismic imaging is not very sensitive to intermediate wavelengths (Claerbout, 1985), especially due to the limitations of recorded offsets

and the limited frequency band of the source in the seismic experiment. As FWI contributes to filling this gap, the spatial resolution reconstructed by the FWI can be analyzed by the wavenumber vector contained in the gradient of the objective function (Virieux and Operto, 2009), where each frequency and source-receiver offset map a single wavenumber in the model. Jannane et al. (1989) determined which wavelengths are present in the seismic data by introducing different perturbations in the property model, and quantifying the hypothesis made by Claerbout, thus providing the wavelengths that can be recovered by the seismic data and verifying the gap in the intermediate lengths.

Sirgue and Pratt (2004) explained the relationship between the source frequency content and the inverse wavenumbers (inverse of wavelength) illuminated in the property model using the Claerbout (1971) cross-correlation imaging condition. Also, they showed that the wavenumbers retrieved by the gradient can be written as a function of the wavenumbers scattered across the source field and the receiver field. In addition, they provided an analytical expression for the 1D case using a single frequency. More recently, Soares et al. (2015) performed numerical experiments introducing 2D perturbations in the Santos Basin velocity model and analyzed the effects of these perturbations on the velocity model, estimated by FWI, thus identifying the sizes of the structures (wavelengths) that can be recovered for different offsets and initial models.

Therefore, understanding and quantifying the wavenumbers retrieved by the seismic inversion can contribute to the development of strategies that mitigate FWI convergence problems, such as cycle skipping (Virieux and Operto, 2009). In this study, we show that FWI contributes to filling the gap of the intermediate wavelengths present in

* Corresponding author.

E-mail addresses: felipetimoteo@id.uff.br (F.T. Costa), marcocetale@id.uff.br (M.A.C. Santos).

the seismic reflection data for a given numerical experiment. We also propose a method to estimate and determine the illuminated wavenumbers by the FWI and find the limit of the spatial resolution reached by the FWI using the L1 and L2 norms.

The article is organized as follows. First, the fundamental aspects of the acoustic time-domain FWI theory were reviewed. A strategy to determine how the wavenumbers of a perturbation in the model are recovered by the time-domain gradient of the objective function was proposed and compared to a numerical experiment performed in the frequency domain originally by [Sirgue and Pratt \(2004\)](#). A quasi-monochromatic wavelet was built to simulate a discrete frequency in the time domain. In the second part of the article, an FWI experiment using a region of the Marmousi model with low-velocity anomaly was carried out. The wavenumber recovered by the inversion using norm L1 and norm L2 was compared with the expected wavenumber for the low-velocity anomaly using the Fourier transform. And so, we determine the limits of the wavenumbers illuminated by the FWI.

2. Materials and methods

2.1. Review of forward problem in acoustic media

The wave equation, the inversion core, allows estimating the surface pressure field (calculated data), and then, comparing the observed data, to obtain the residual used to calculate the objective function, which in turn, is minimized iteratively by the nonlinear optimization method. In an acoustic, isotropic and heterogeneous medium with constant density, the wave equation is expressed as follows:

$$\frac{1}{c^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} p(\mathbf{x}, t) - \nabla^2 p(\mathbf{x}, t) = s(\mathbf{x}, t), \quad (1)$$

where the vector $\mathbf{x}$ is the spatial coordinates (in the 2D case, x and z); t time; c propagation velocity of the wave in the medium; p pressure field; and s source signature. The finite difference approximation ([Levander, 1989](#)) was used to solve the acoustic wave equation numerically and reproduce the seismic experiment.

2.2. The inverse problem and the optimization methods

In general, optimization methods are techniques used to solve the inverse problem that allows estimating a set of unknown parameters, from the observed data, in a pre-established physical experiment. In this case, the estimated parameter set is the subsurface velocity model; the observed data are the recorded seismograms, and the physical experiment is the multichannel seismic acquisition.

Defining the objective function (cost function, error function) is a fundamental step when solving inverse problems. The objective function can be understood as a measurement of the similarity between the observed and calculated data, which typically, in inverse problems, is the quadratic norm of the difference between the observed and calculated data ([Menke, 2012](#)). The objective function used in this work is expressed as:

$$\chi(\mathbf{m}) = \frac{1}{2} \int_T \int_{\partial\Omega_s} \int_{\partial\Omega_r} |(\mathbf{d}_{cal}(\mathbf{x}_r, \mathbf{x}_s, t)) - \mathbf{d}_{obs}(\mathbf{x}_r, \mathbf{x}_s, t)|^2 d\mathbf{x}_r d\mathbf{x}_s dt, \quad (2)$$

where $\mathbf{x}_r$, $\mathbf{x}_s$ and t are the positions of the receiver stations, sources and time, respectively; $\partial\Omega_r$ is the domain containing the positions of the receiver stations arranged per shot; $\partial\Omega_s$ is the domain containing the source positions; and, T the recording period per shot. The terms $\mathbf{d}_{obs}$ and $\mathbf{d}_{cal}$ are the observed and calculated data, whereas $\frac{1}{2}$ is a mathematical convenience.

[Nocedal and Wright \(2006\)](#) stated that the model must be estimated iteratively using a local optimization method based on the gradient of

the objective function due to the nonlinearity of the seismic problem and the large number of parameters. These methods only guarantee convergence to a local minimum of the objective function, so it is fundamental that the initial model is close to the global minimum.

Thus, the equation that estimates the velocity model iteratively is given by:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - [\nabla^2 \chi(\mathbf{m}_k)]^{-1} \nabla \chi(\mathbf{m}_k), \quad (3)$$

where $\mathbf{m}_{k+1}$ is the velocity model estimated for the k th iteration and $\mathbf{m}_k$, the initial model. $\nabla \chi(\mathbf{m}_k)$ is the gradient of the objective function with respect to the parameters, and $[\nabla^2 \chi(\mathbf{m}_k)]^{-1}$ is the inverse of the Hessian operator, whose elements represent the second derivative of the objective function with respect to the parameters. Normally, calculating the inverse of the Hessian matrix is computationally prohibitive for problems with a large number of parameters, therefore, methods for an approximate calculation are used. The approximation used to calculate the inverse of the Hessian matrix defines the optimization method ([Nocedal and Wright, 2006](#)).

The L-BFGS quasi-Newton method has been used extensively to solve the FWI ([Métivier et al., 2012](#)), being the most indicated among gradient-based methods due to the supra-linear convergence rate. In addition, it uses the approximation of the Hessian matrix obtained through the k th gradients of the previous iterations, calculated by finite differences iteratively, so that the inverse of the Hessian matrix is updated at each iteration ([Métivier and Brossier, 2016](#)). Therefore, the equation for updating the model parameters is:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k H_k \nabla \chi(\mathbf{m}_k) \quad (4)$$

where α_k is the step size and H_k is the approximation of the inverse of the Hessian matrix and the product $-H_k \nabla \chi(\mathbf{m}_k)$ is the update direction. Once the update direction is defined, it is necessary to define step size α_k . The step size should ensure that the objective function is minimal for a given update direction. This criterion can be reached using Wolfe's conditions: (1) The step must be calculated to reduce sufficiently the objective function for a given direction; (2) The step should not be too small, avoiding insignificant updates. In addition to defining step size, it is possible to use a constraint to ensure that the updated model is within a plausible domain ([Métivier and Brossier, 2016](#)).

2.3. Gradient calculation by the adjoint method

For problems with a large number of parameters, calculating the gradient of the objective function $\nabla \chi(\mathbf{m}_k)$ has a high computational cost, becoming prohibitive sometimes. This difficulty stems from the need to calculate Fréchet's derivatives, which can be interpreted as the wavefield recorded on the surface that was scattered to each of the points of the property model ([Martins, 2015](#)).

[Lailly \(1983\)](#) and [Tarantola \(1984\)](#) presented a way to obtain the gradient of the objective function without calculating the Fréchet's derivatives. The method, known as the adjoint state method, allows solving the inverse problem for the seismic case with several parameters. Later, [Plessix \(2006\)](#) revised the adjoint state method using the Lagrange multipliers to obtain the gradient of the objective function, allowing the method to become popular in the geophysical community.

In the acoustic case, where the parameter to be estimated is the velocity model ($\mathbf{m} = c(\mathbf{x})$), the gradient of the objective function for a single shot becomes ([Fichtner, 2011](#)):

$$\nabla \chi(c) = \frac{2}{c^3(\mathbf{x})} \int_T \int_D p^\dagger(\mathbf{x}, t) \left[\frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} \right] d\mathbf{x} dt, \quad (5)$$

where the integrals are evaluated for all instants of time and for the entire computational domain. The second derivative of the pressure field is

obtained by forward modeling using Eq. (1). And, the adjoint field, $p^\dagger(\mathbf{x}, t)$, is obtained by the reverse time modeling, using the final conditions and the adjoint source (Fichtner et al., 2008):

$$\frac{1}{c^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} p^\dagger(\mathbf{x}, t) - \nabla^2 p^\dagger(\mathbf{x}, t) = \frac{\partial \chi}{\partial p}, \quad (6)$$

in the case of the L2 norm (Eq. (2)), the adjoint source is given by the residual, difference between the observed and calculated data:

$$\frac{\partial \chi}{\partial p} = \Delta \mathbf{d} = \mathbf{d}_{cal}(\mathbf{x}_r, \mathbf{x}_s, t) - \mathbf{d}_{obs}(\mathbf{x}_r, \mathbf{x}_s, t). \quad (7)$$

It is observed that the operators used for calculating the pressure field and the adjoint field are the same except for the source term.

According to Tarantola (1984), the residuals contain the missing information in the current velocity model. The adjoint field can be understood as a sort of missing diffracted field generated by the information in the residuals. In this case, the correlation between the missing diffracted field and the second time derivative of the forward wavefield (Eq. (5)) provides the position of the necessary correction in the model that could explain the observed data. The calculation of $\nabla \chi(c)$ is similar to the imaging condition of the Reverse Time Migration (RTM) method proposed initially by Claerbout (1971), except for the constants and the time derivative factor.

The choice of the parameters used to solve the inverse problem can affect drastically the convergence properties of the optimization method, even if the parameters were mathematically equivalent (Tarantola, 1986). Carneiro et al. (2018) showed that if the inverse problem (Eq. (4)) was parameterized by the square of slowness:

$$\mathbf{m}(\mathbf{x}) = s(\mathbf{x})^2 = \frac{1}{c(\mathbf{x})^2}, \quad (8)$$

instead of velocity $c(\mathbf{x})$, the optimization method would have better convergence properties. The reparametrization was motivated by the dimensional analysis of the model update equation (Eq. (4)). Therefore, the calculation of the gradient of the objective function can be rewritten as:

$$\nabla \chi(c) = \frac{c^3(\mathbf{x})}{2} \int_T \int_D p^\dagger(\mathbf{x}, t) \left[\frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} \right] d\mathbf{x} dt. \quad (9)$$

2.4. Multiscale approach

The oscillatory nature of the seismic data implies that the residual, used in the calculation of the objective function, also tend to have an oscillatory behavior. Therefore, the objective function will have highly oscillatory behavior, as a consequence of the non-linearity of the seismic data, so it is natural to expect a large number of local minima in the objective function (Fig. 1a).

However, even choosing an initial velocity model close to the global minimum does not guarantee the convergence of the objective function to a property model that makes geological sense, which can be explained by cycle skipping. Virieux and Operto (2009) explained that this problem occurs when the phase difference between the calculated and observed data is greater than half period of the wavelength so that the optimization method adjusts the calculated data to one cycle of the observed seismic trace with one or more wavelength lag. Consequently, the accumulation of out-of-phase adjustments, between calculated and observed data, can insert spurious structures in the property model that make no geological sense, or in the worst case, structures that make geological sense but do not match reality.

Bunks et al. (1995) proposed the multiscale approach to avoid the cycle skipping problem. This approach consists of separating the

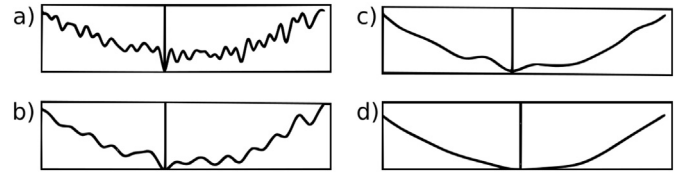


Fig. 1. Representation of the objective function and its oscillatory behavior for different scales. (a)–(d) illustrate the same objective function with increasing scales. The objective function has fewer local minima, the larger the scale. Modified from Bunks et al. (1995).

problem into different scales, so that the objective function oscillates less in the larger scales (Fig. 1d), allowing the local optimization algorithm to approach the global minimum of the objective function more consistently.

The multiscale approach can be applied to several domains such as frequency, source-receiver offset, and time window of observed data, among others, the most widely used is the frequency domain due to the nature of the seismic data. This approach optimizes the low frequencies (larger scales) at the beginning, optimizing progressively higher frequencies, up to a certain cutoff frequency. At low frequencies, the seismic data is expected to have fewer oscillations and lower frequency content, however, the objective function also oscillates less (Fig. 1d), contributing to the convergence of the optimization method to a value closer to the global minimum. With the progressive increase of the frequency content, the objective function tends to converge to the global minimum, avoiding the problem of the cycle skipping (Figs. 1d–1c–1b–1a).

2.5. Recovery of wavenumbers

The cross-correlation imaging principle formulated by Claerbout (1971) consists of three steps, (1) Extrapolation of the direct wavefield; (2) Reverse extrapolation of the recorded wavefield; and (3) Applying the image condition. As previously mentioned, calculating the gradient of the objective function in the time domain is very similar to calculating the image migrated in depth by the reverse time migration (RTM) method. Thus, wavelength illumination analysis is similar for both methods (Huang and Schuster, 2014).

The main difference is in the analyzed frequency content. For the gradient analysis, the frequency content is controlled by the multiscale approach (from low to high frequencies), the frequency content of the migrated image presents full spectrum in the observed seismogram, except for the direct wave.

The relationship between the frequency of the seismic source and the wavenumbers illuminated in the property model can be better understood in the frequency domain, as explained by Sirgue and Pratt (2004). The gradient calculated by the adjoint state method for a single temporal frequency is given by:

$$g(\omega) = \omega^2 \sum_r \sum_s \text{Re} [G_0^*(\mathbf{x}, \mathbf{x}_s) G_0(\mathbf{x}, \mathbf{x}_r) \Delta \mathbf{d}(\mathbf{x}_s, \mathbf{x}_r)], \quad (10)$$

where ω is the frequency injected by the source, $G_0^*(\mathbf{x}, \mathbf{x}_s)$ is the conjugate complex of the Green function for an excitation at the position of the sources, $G_0(\mathbf{x}, \mathbf{x}_r)$ is the conjugate complex of the Green function for an excitation at the position of the receivers, and $\Delta \mathbf{d}(\mathbf{x}_s, \mathbf{x}_r)$ is the residual. Comparing Eq. (10) to the time domain gradient calculation (Eq. (9)), the calculation of the Green functions equals the extrapolation of the forward wavefield (excitation at the source position, $G_0(\mathbf{x}, \mathbf{x}_s)$) and extrapolation of the reverse wavefield (excitation at the receiver position multiplied by the residual, $G_0^*(\mathbf{x}, \mathbf{x}_r) \Delta \mathbf{d}$). Woodward (1992) called the gradient operator, the wavepath.

Assuming that the amplitude effects can be ignored and the velocity model is homogeneous (c_0), and further, using the approach to distant

fields, the Green functions can be approximated by plane waves scattered by both the source and the receivers (Sirgue and Pratt, 2004):

$$G_o^s(\mathbf{x}, \mathbf{x}_s) \approx e^{-ik_0 \hat{\mathbf{s}} \cdot \mathbf{x}}; \quad G_o^r(\mathbf{x}, \mathbf{x}_r) \approx e^{-ik_0 \hat{\mathbf{r}} \cdot \mathbf{x}}; \quad (11)$$

allowing the gradient to be expressed in terms of the wavenumber vector scattered by the source and receivers, weighted by residuals:

$$g(\mathbf{x}) = \omega^2 \sum_r \sum_s \text{Re} \left[e^{-ik_0(\hat{\mathbf{s}} + \hat{\mathbf{r}}) \cdot \mathbf{x}} \Delta \mathbf{d} \right], \quad (12)$$

where $k_0 = \omega/c_0$ is the wavenumber generated by the source and i is the imaginary unit. Sirgue and Pratt (2004) reported that for a single frequency and single source-receiver pair, the retrieved image, located at the midpoint between source and receiver, is expected to have the following properties:

1. The existence of vertical wavenumber vectors only (Fig. 2);
2. Large offsets should recover small wavenumbers while short offsets should recover large wavenumbers; and,
3. Larger wavenumbers should be recovered for a given offset, the deeper the target.

2.6. Wavenumber vs. offset

The contribution of the vertical component of the wavenumber vector as a function of the source-receiver offset can be calculated analytically by a 1D constant velocity model. Using Eq. (12) that relates the gradient calculation with the wavenumbers scattered by the source and the receiver, and also the geometric interpretation of the retrieved wavenumber (Fig. 2), it can be seen that the vertical component of the wavenumber is formed by the sum of the wavenumber vectors scattered by both source and receiver. Therefore, we can write:

$$k_z = 2k_0 \cos\theta. \quad (13)$$

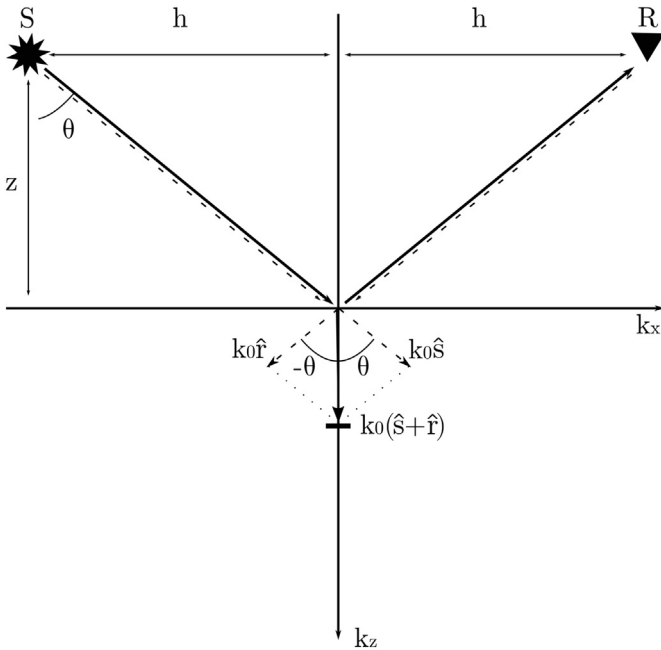


Fig. 2. Schematics showing the illumination of a point in the model, for a single source-receiver pair on a single frequency. Dashed lines are the fields scattered by the source and receiver, while solid lines are the wavepath. Note that for this choice of source-receiver pair the angles of incidence and reflection are the same, so only the wavenumber vertical components are imaged. Modified from Sirgue and Pratt (2004) and Schuster (2017).

Knowing that θ is the aperture angle of the source and receiver wavenumber vectors with respect to the midpoint, $\cos\theta$ can be rewritten in terms of the offset and reflector depth,

$$\cos\theta = \frac{z}{\sqrt{h^2 + z^2}} = \frac{1}{\sqrt{1 + R^2}}, \quad (14)$$

where z is the reflector depth, h is half of the source-receiver offset, and $R = h/z$. Thus, the vertical component at the midpoint between source-receiver can be written as:

$$k_z = 2\frac{\omega}{c_0} \frac{1}{\sqrt{1 + R^2}}. \quad (15)$$

Eq. (15) allows estimating the behavior of the illuminated wavenumbers as a function of the offset. This estimate can be compared to the wavenumbers illuminated by the time-domain adjoint state method as a function of the offset.

3. Results

3.1. Estimation of the wavenumber for a source-receiver pair

This work presented a methodology for the time domain gradient calculation as to reproduce the results obtained by Sirgue and Pratt (2004), originally carried out in the frequency domain. The numerical experiment used a single source-receiver pair. The background velocity model and the perturbed model are similar to the numerical experiments performed by Sirgue and Pratt (2004). The perturbed velocity model (Fig. 3a) is identical to the background model (Fig. 3b) except for the 250 m thick layer at a depth of 2000 m and 2700 m/s velocity. The velocities of the first and second layers are 2400 m/s and 2500 m/s, respectively.

The wavelet used as a source was constructed to have a narrow frequency spectrum, simulating a source with a single angular frequency ω , without compromising the stability of the modeling operator in the time domain. The wavelet used, known as Wide-Band B-spline wavelet, proposed by Cao and Han (2011) produces the desired property and is expressed as:

$$s(\mathbf{x}, t) = \frac{1}{q-p} \sqrt{f_b} \left(\text{sinc} \left(\frac{f_b t}{m} \right) \right)^m (q \text{sinc}(2qt) - p \text{sinc}(2pt)) \delta(\mathbf{x} - \mathbf{x}_r). \quad (16)$$

Eq. (16) was based on the theory of decomposition in the wavelet domain and the respective reconstruction in the time domain. This wavelet was formulated as a function of 4 parameters, p (initial frequency), q (final frequency), m (order), and f_b (bandwidth) that define their frequency content accurately.

In this experiment, frequencies centered on 5 Hz ($p = 4.5$, $q = 5.5$, $m = 1$, $f_b = 1$) and 10 Hz ($p = 9.5$, $q = 10.5$, $m = 1$, $f_b = 1$) are shown in Fig. 3c and d. The respective gradients were calculated using the adjoint method in the time domain with offsets of 10 km (Fig. 3e and g) and 3 km (Fig. 3f and h), the same configuration as the numerical experiment performed in the frequency domain by Sirgue and Pratt (2004).

The wavenumbers contained in the gradient in the interface region¹ were obtained by the Fourier transform 1D (Fig. 3i, j, k, and l), which was calculated only for the midpoint between source and receiver in each case.

The gradient analysis for all configurations (Fig. 3e, f, g and h) allows us to conclude that property 1, proposed by Sirgue and Pratt (2004), is

¹ It was applied a Hamming window on gradient centered in reflector region to isolate the illuminated wavenumbers in this region.

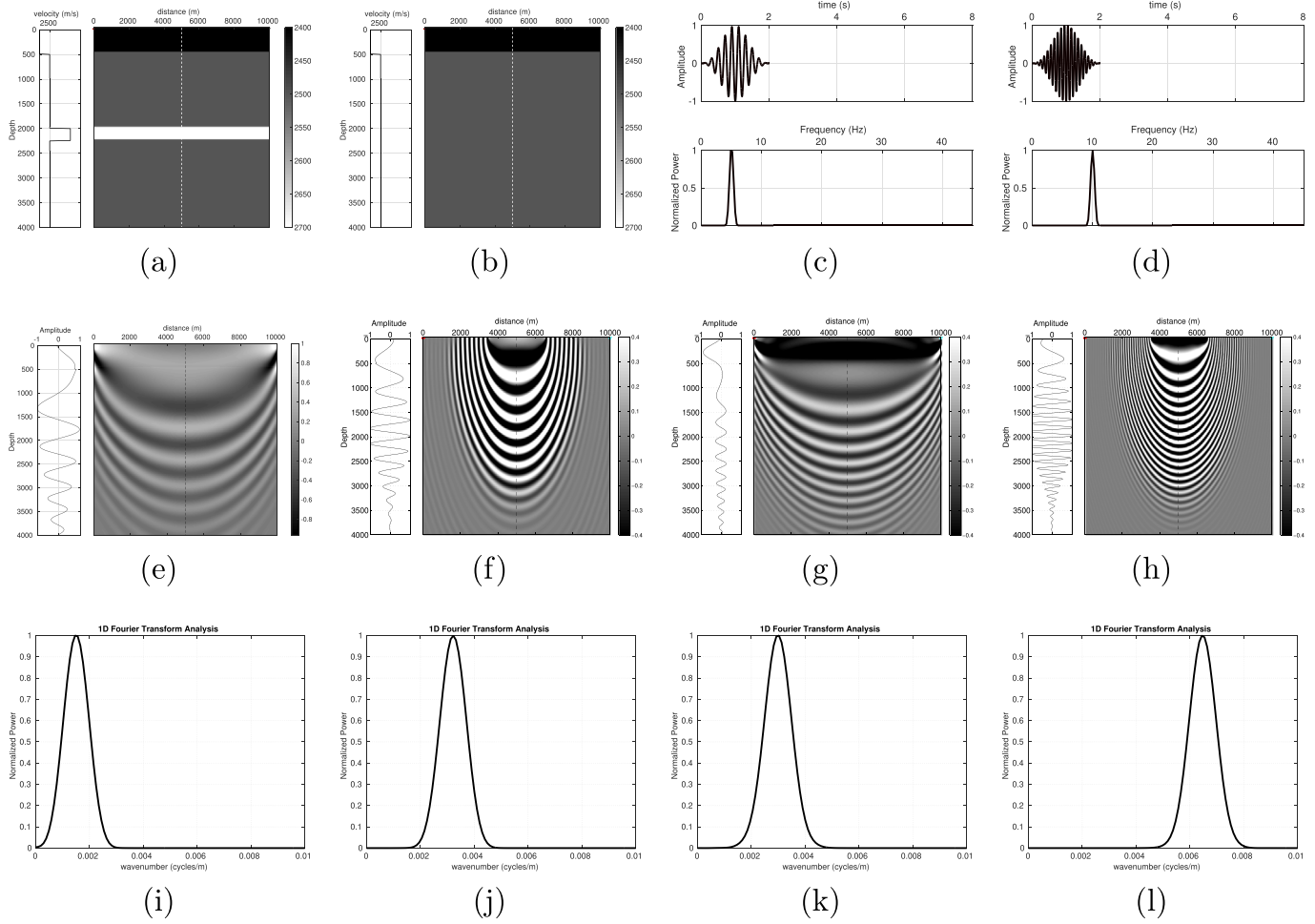


Fig. 3. (a) Perturbed velocity model. (b) Background velocity model. (c) Wavelet centered at 5 Hz. (d) Wavelet centered at 10 Hz. (e) Gradient for a single source-receiver pair with frequency centered at 5 Hz, and 10 km offset. (f) Gradient for a single source-receiver pair with frequency centered at 5 Hz, and 3 km offset. (g) Gradient for a single source-receiver pair with frequency centered at 10 Hz, and 10 km offset. (h) Gradient for a single source-receiver pair with frequency centered at 10 Hz, and 3 km offset. (i)–(j)–(k)–(l) Fourier transforms for the midpoint (dashed line) of the (e)–(f)–(g)–(h) gradients, respectively.

checked at the midpoint positions between source and receiver (dashed line). For each case, property 2 is also checked at the midpoint positions between the source and receiver. In addition, it is verified that for both frequencies (5 Hz and 10 Hz), the smaller offsets recover higher wavenumbers. Finally, the images of the gradients (Fig. 3e, f, g and h) show that with increasing depth, for the same midpoint, the wavenumbers increase (property 3) as the amplitude decreases. In addition, as expected, the comparison of the 5 Hz and 10 Hz cases shows that increasing the source frequency content allows retrieving larger wavenumbers.

An additional numerical experiment was done using a broadband wavelet (Fig. 4a). The start frequency of the wavelet was 5 Hz and the cutoff frequency was 10 Hz, the parameters used in B-spline wavelet were $p = 4$, $q = 11$, $m = 10$, and $f_b = 7$. Likewise, these results respect the properties 1, 2 and 3. The gradient using offsets of 10 km (Fig. 4c) contain the wavenumbers shown in Fig. 4e. The wavenumbers illuminated using the broadband wavelet contain the same wavenumbers of the previous case (Fig. 3i and k) using the wavelets centered in 5 and 10 Hz. In the case of 3 km offset, the gradient is shown in Fig. 4d. The wavenumbers illuminated in this case (Fig. 4f) have a gap between two peaks of wavenumbers. The gap between the wavenumber peaks means that even with a broadband wavelet, there are some wavenumbers that are not illuminated although these missing wavenumbers could be recovered by the different offsets.

3.2. Calculating the vertical wavenumber as a function of offset

The vertical component of the wavenumber vector was calculated analytically using Eq. (14) and compared to the wavenumbers obtained numerically by the adjoint gradient method in the time domain (Eq. (9)). Therefore, a numerical experiment like similar to the previous one was performed using all receivers with up to 10 km offset. The calculated gradient was obtained using 5 Hz and 10 Hz dominant frequency sources, which are shown in Fig. 5a and c.

Figs. 5b and 6d show the 1D Fourier transform of the gradients, trace-by-trace, which provide the vertical wavenumbers in the gradients. It is noteworthy that the vertical wavenumbers retrieved by the gradient adjust well to the theoretical curve (Eq. (15)) for both cases up to approximately 5 km offset, the midpoint between the source and the maximum offset. Despite the artifacts, mainly caused by the transmitted wave, this result shows that the assumptions used for estimating the wavenumbers in the frequency domain are also reproducible in the time domain using an approximately monochromatic wavelet. The shift for lower spatial frequencies as a function of offset is called image stretch (Sirgue and Pratt, 2004).

3.3. The 2D seismic experiment - the forward problem

In the performed numerical experiments, all observed data were obtained synthetically, using the same operator to obtain the calculated

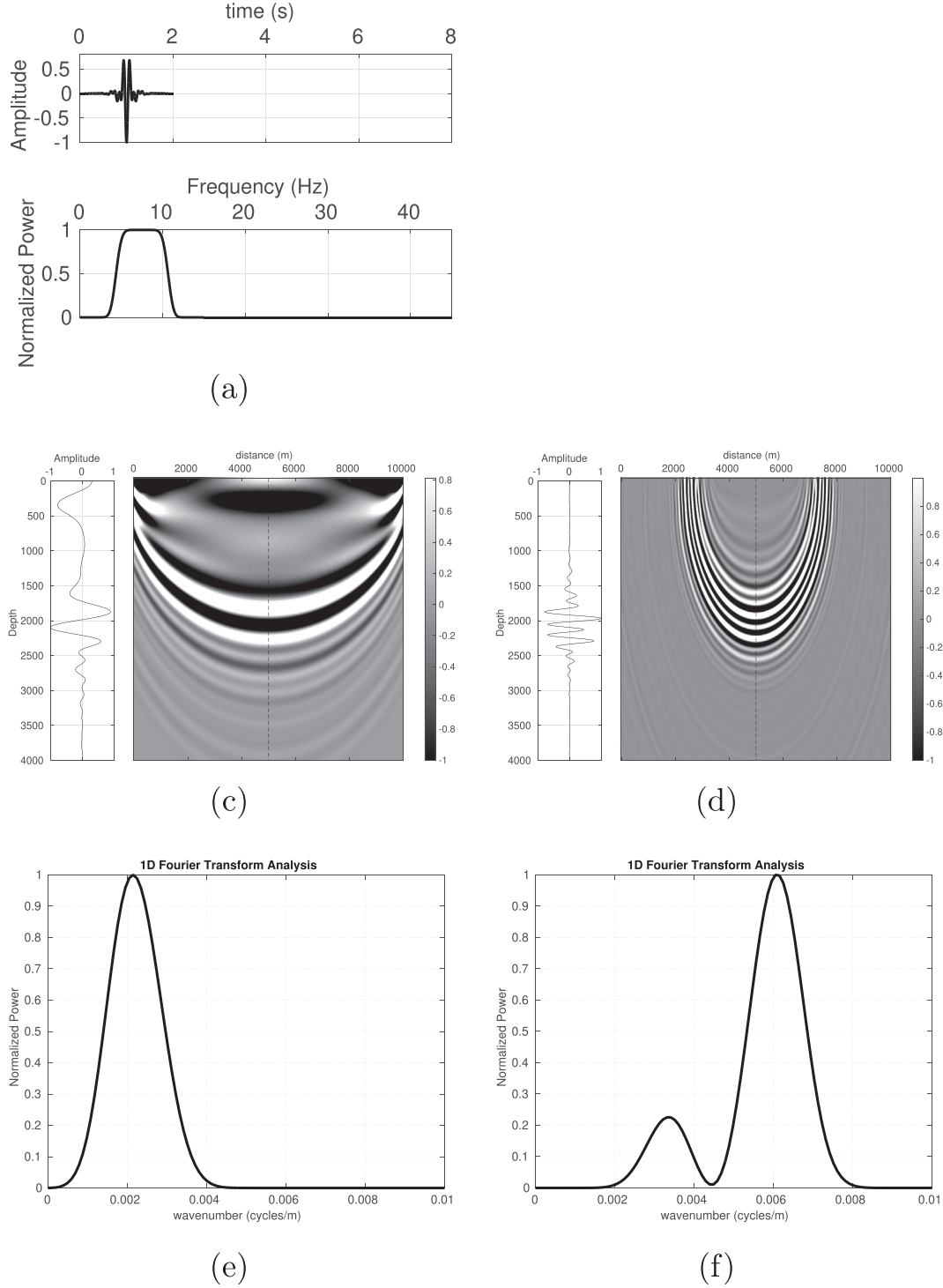


Fig. 4. (a) Broadband wavelet with a start frequency of 5 Hz and cutoff frequency of 10 Hz. (b) Gradient for a single source-receiver pair with broadband wavelet, and 10 km offset. (c) Gradient for a single source-receiver pair and 3 km offset. (d) and (e) Fourier transforms for the midpoint (dashed line) of the (a)–(b) gradients, respectively.

and observed data but different velocity models. The so-called true model was used for the observed data and the initial or estimated models for the calculated data. This process became known as the “inversion crime” (Wirgin, 2004).

In this study, the true model uses a region of the Marmousi II model, developed by Versteeg (1994) and updated by Martin et al. (2006) (Fig. 6b). The initial model (Fig. 6b) was obtained by smoothing the real model to simulate a velocity model originating from the seismic tomography.

A total of 64 shots with 38 m spacing were performed on the model surface. The receivers were positioned along the model at 450 m deep

with 5 m spacing. The source used to generate the observed data was the Ricker wavelet with 30 Hz, given by Eq. (17):

$$s(\mathbf{x}, t) = A_{\max} \left[2\pi(\pi f_c t_d)^2 - 1 \right] e^{-\pi(\pi f_c t_d)^2} \delta(\mathbf{x} - \mathbf{x}_s), \quad (17)$$

where the parameter f_c is related to the cutoff frequency f_{cut} and expressed as $f_c = \frac{f_{\text{cut}}}{3\sqrt{\pi}}$, t_d is the shifted time so that the wavelet is centered at zero. $A_{\max}$ is the maximum amplitude and $\mathbf{x}_s$ the position of the

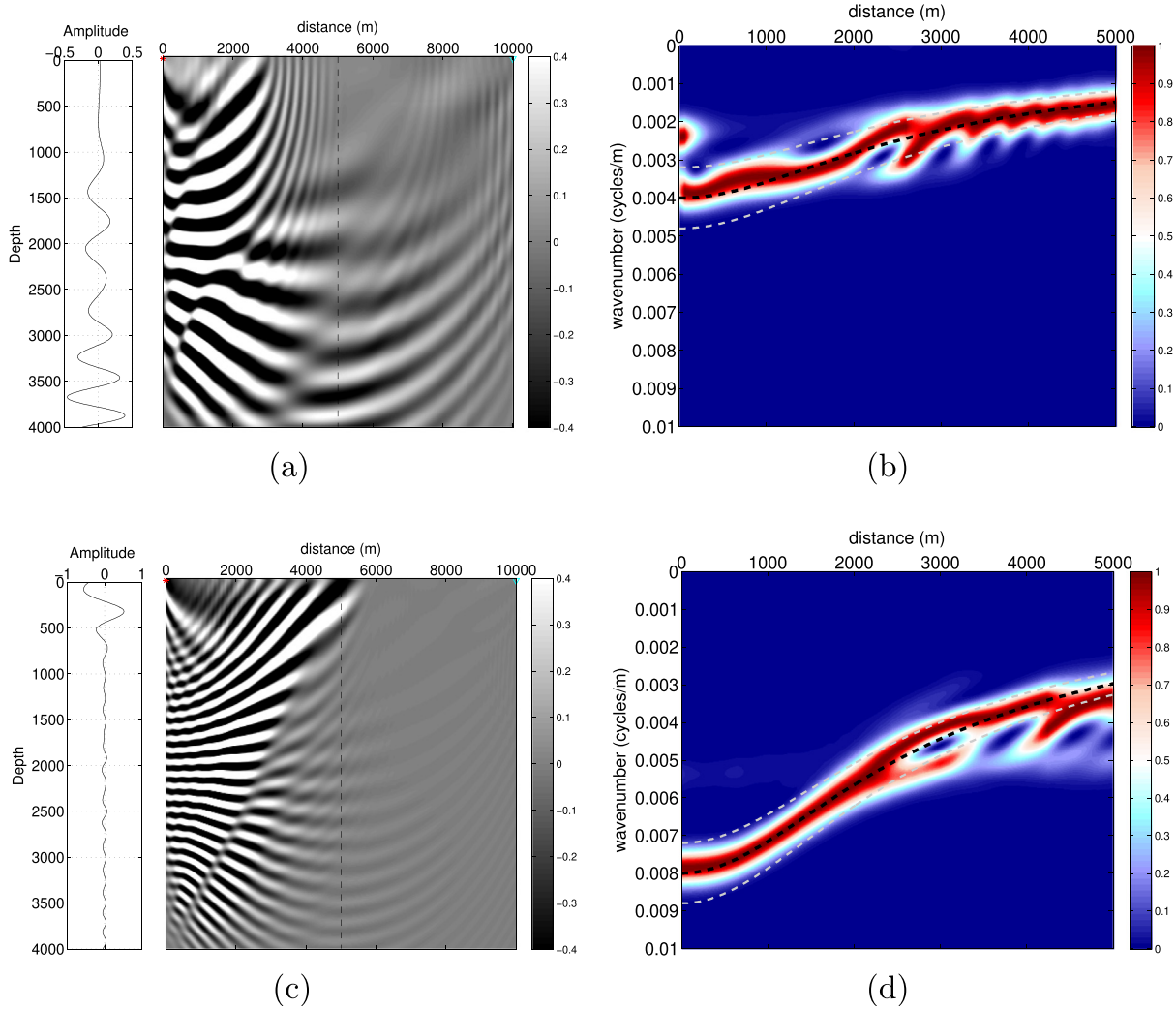


Fig. 5. (a) and (c) Gradient calculated using all source-receiver offsets for the central frequencies of 5 Hz and 10 Hz. The vertical dashed line indicates the position of the profile shown on the left. (b) and (d) Fourier transform, trace by trace, of the 5 Hz and 10 Hz gradients. The black dashed line is the vertical wavenumber as a function of offset given by Eq. (10). The other dashed lines indicate the upper and lower limits due to the wavelet source being almost monochromatic.

source. The Ricker wavelet is commonly used to solve seismic problems due to its similarity to the seismic signature of real data, both in form (Fig. 7a) and frequency spectrum (Fig. 7b).

3.4. The inverse problem

At this stage of the numerical experiment, the optimization method chosen to solve the inverse problem was the L-BFGS (Eq. (4)) due to its

supra-linear convergence rate. The gradient, a fundamental part for defining the updating direction, was calculated by the adjoint state method in the time domain parameterized in the square of slowness (Eq. (9)).

In addition, the multiscale strategy was employed to mitigate the nonlinearity of the problem. Initially, the low frequencies were inverted using the filtered Ricker wavelet with a cutoff frequency of 12 Hz. The velocity model resulting from this inversion was used in the initial

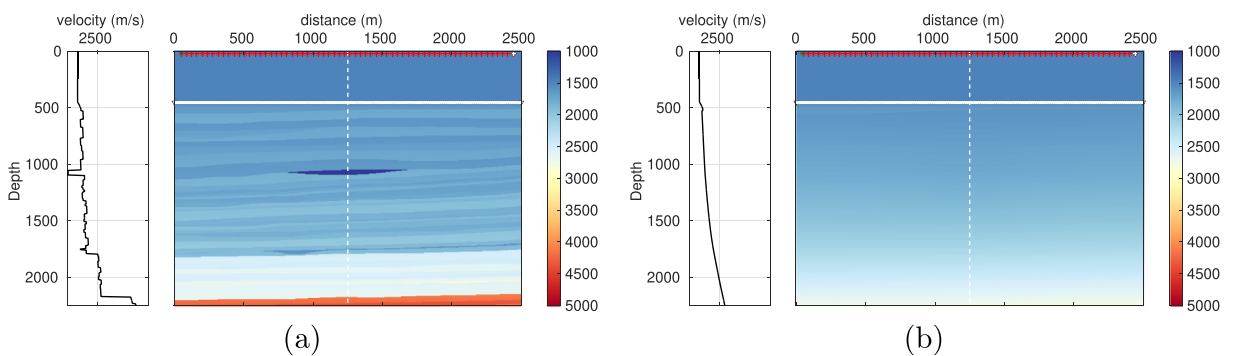


Fig. 6. (a) True model – a region of the Marmousi model. (b) Initial model. The dashed white lines indicate the position of the profile shown on the side of each model.

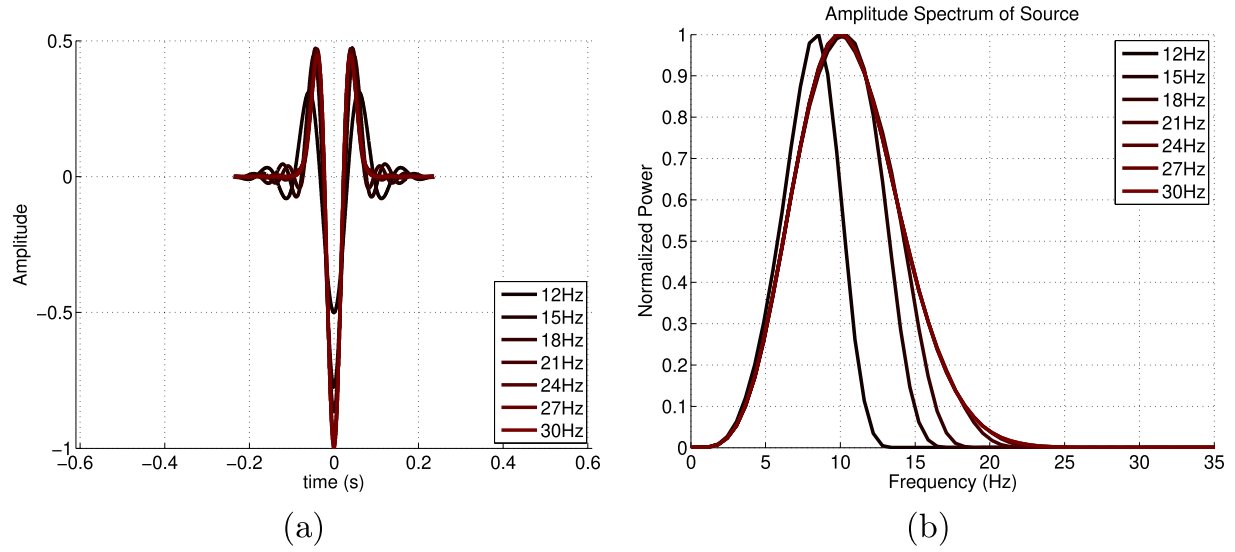


Fig. 7. (a) Wavelet Ricker with cutoff frequencies of 30 Hz in red and filtered wavelets with cutoff frequencies from 12 to 27 Hz. (b) Frequency spectrum of the wavelets in Figure (a). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

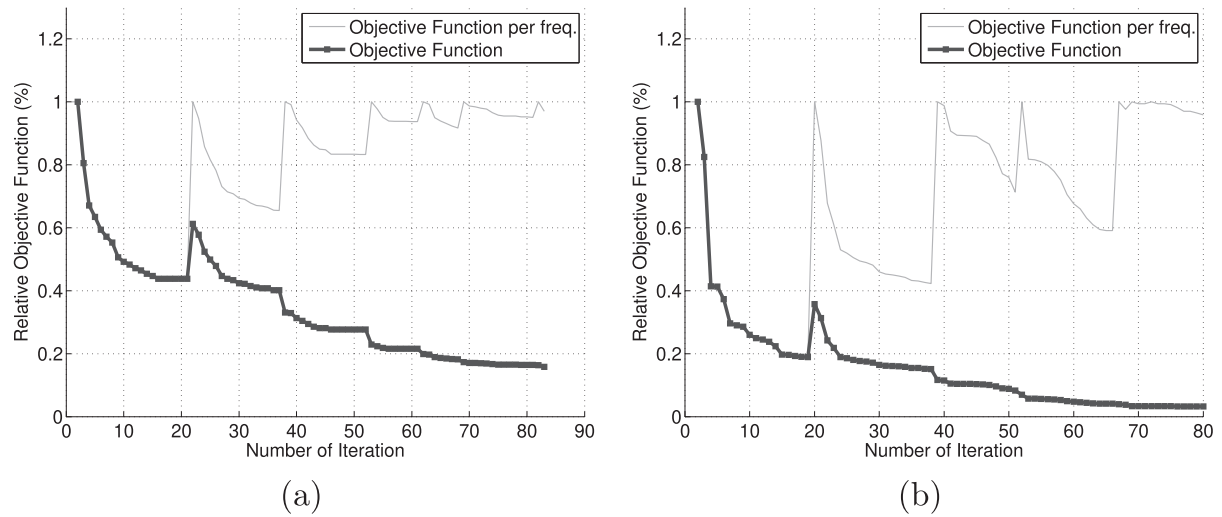


Fig. 8. Evolution of the normalized objective function. a) Norm L1 and b) Norm L2.

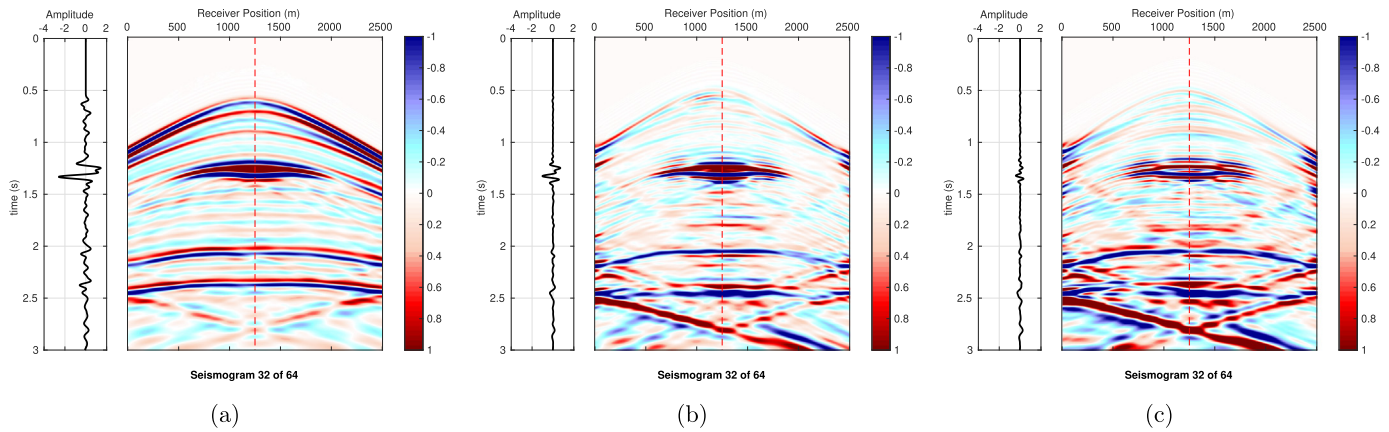


Fig. 9. Comparison between (a) initial, (b) estimated by norm L1 and (c) estimated by norm L2 residual seismograms. The dashed red line indicates the position of the shown profile on the side of each seismogram. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

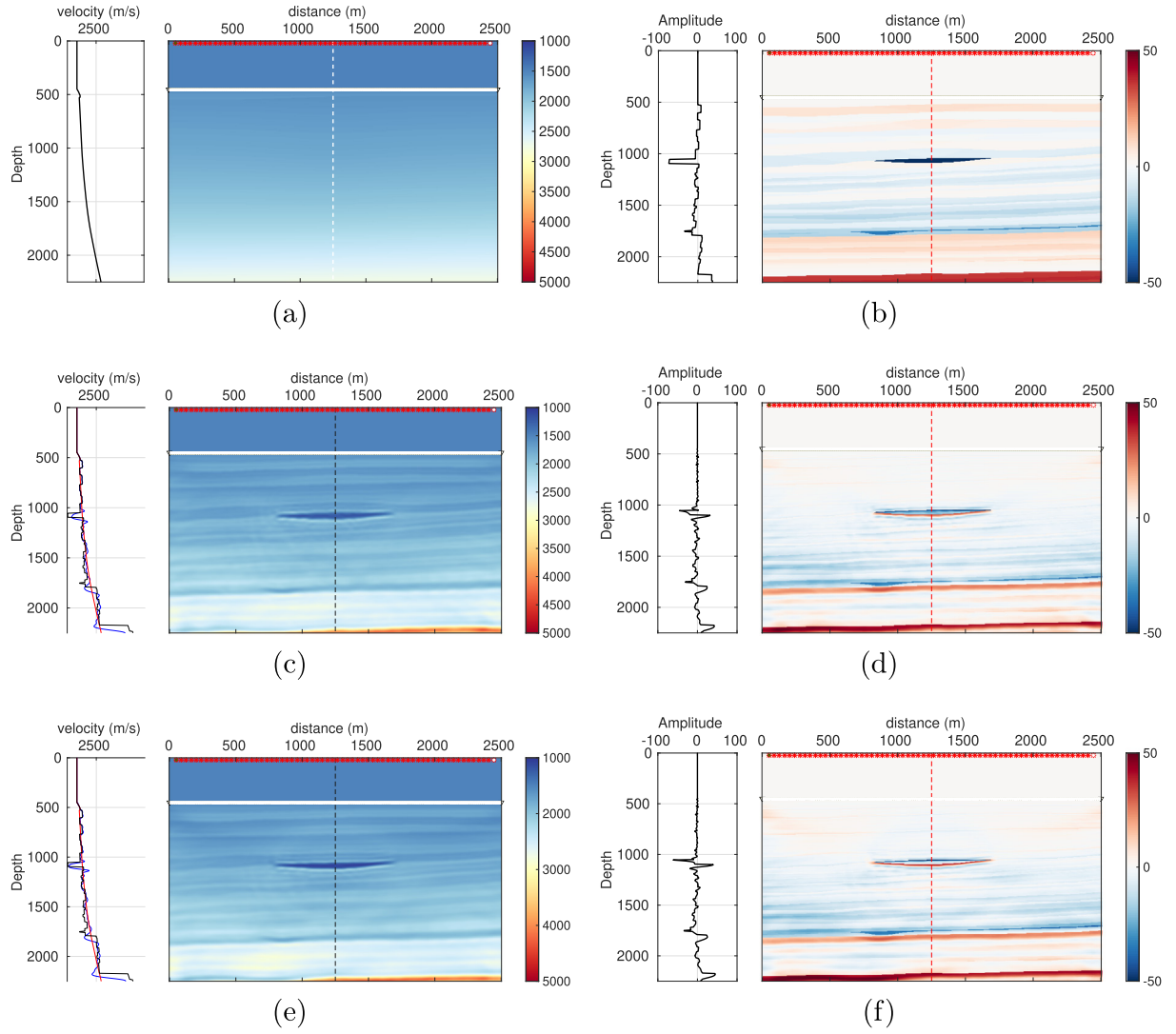


Fig. 10. Models estimated by the FWI using the norms (a) L1 and (b) L2. The dashed lines indicate the position of the shown profiles on the side of each model, the red line is the initial model profile, the black line is the real model profile, and the blue line the estimated model. The relative difference between c) true model and initial model, d) true model and estimated model by norm L1 and e) true model and estimated model by norm L2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

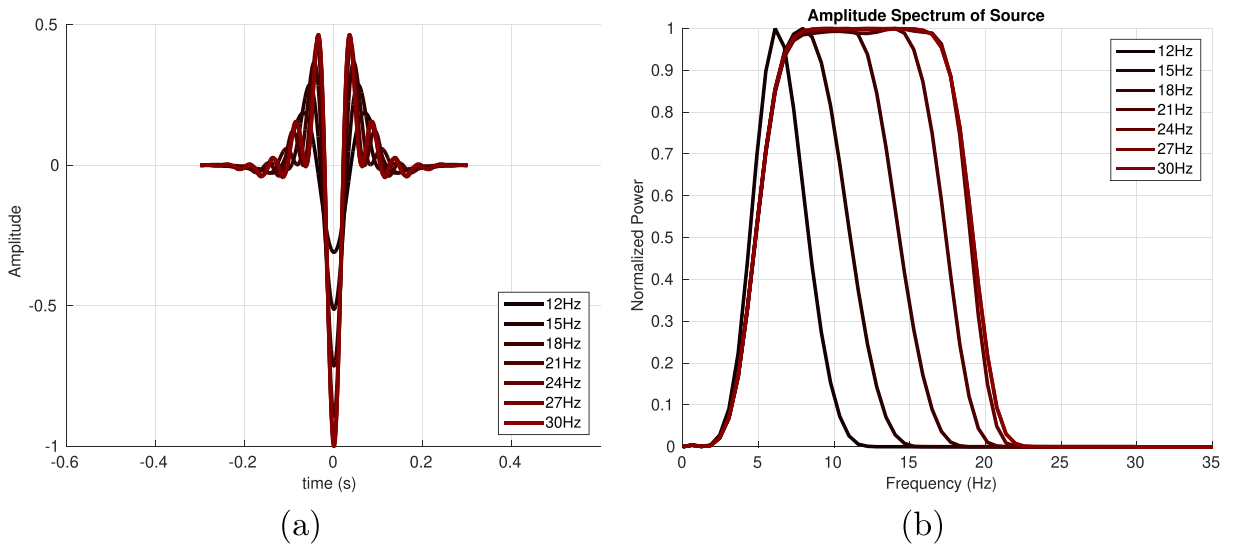


Fig. 11. (a) Wavelet Ricker with cutoff frequencies of 30 Hz in red and filtered wavelets with cutoff frequencies from 12 to 27 Hz. (b) Frequency spectrum of the wavelets in Figure (a). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

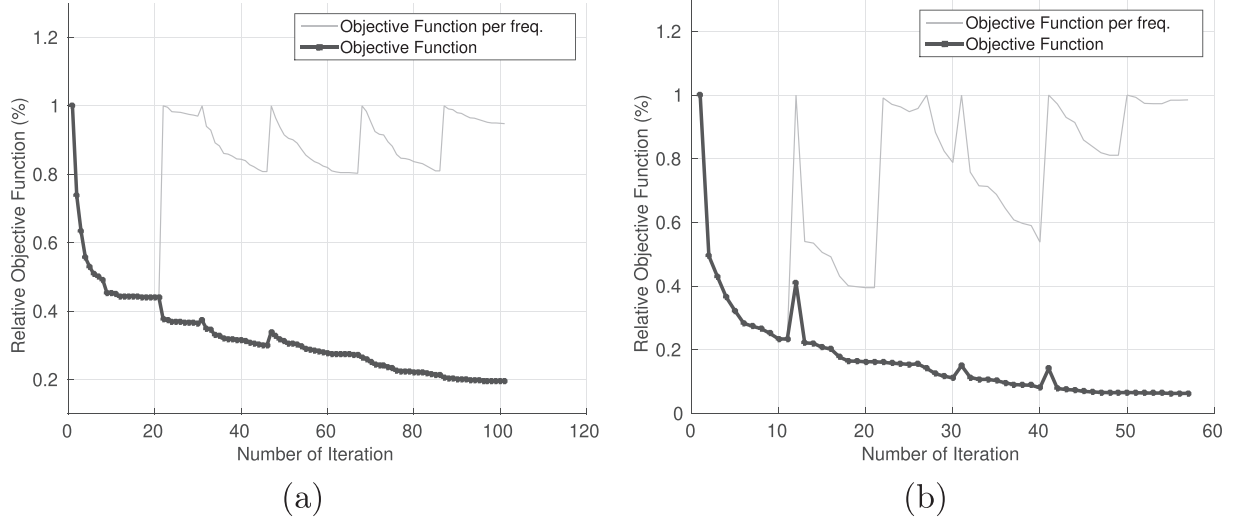


Fig. 12. Evolution of the normalized objective function using the broadband wavelet. a) Norm L1 and b) Norm L2.

model of the subsequent inversion in the filtered Ricker wavelet with higher cutoff frequency. This process was performed iteratively up to the cutoff frequency of the 30 Hz source (approximately 10 Hz center frequency).

Experiments were performed using the objective function with the L1 and L2 norms. The difference between the calculation of the gradient of the objective function is the choice of the adjoint source in Eq. (6). Where the adjoint source for the L2 norm is given by Eq. (7) (data residual), and the adjoint source for L1 norm is given by the normalized residual of the data:

$$\frac{\partial \chi}{\partial p} = \frac{\Delta d}{|\Delta d|}. \quad (18)$$

Fig. 8 shows the convergence evolution of the normalized objective function to the L1 and L2 norms. It is noteworthy the convergence consistency across the different frequency bands, indicating that the model estimated by FWI has converged to a minimum. The stopping criterion was based on the reduction of the relative objective function.

Fig. 9a shows the residual seismogram using the initial velocity model. The Fig. 9b and c show the final residual seismogram using the velocity model estimated by the norm L1 and the norm L2, respectively. It is possible to verify that the amplitudes decrease in the final residuals, indicating that the objective function reached a minimum in both cases.

Fig. 10c and e show the models estimated using the L1 and L2 norms, respectively. As can be seen, the final models obtained by the FWI are very close to the model used to generate the observed data, as shown in Fig. 6a. Fig. 10b, d, and f show the relative differences between the

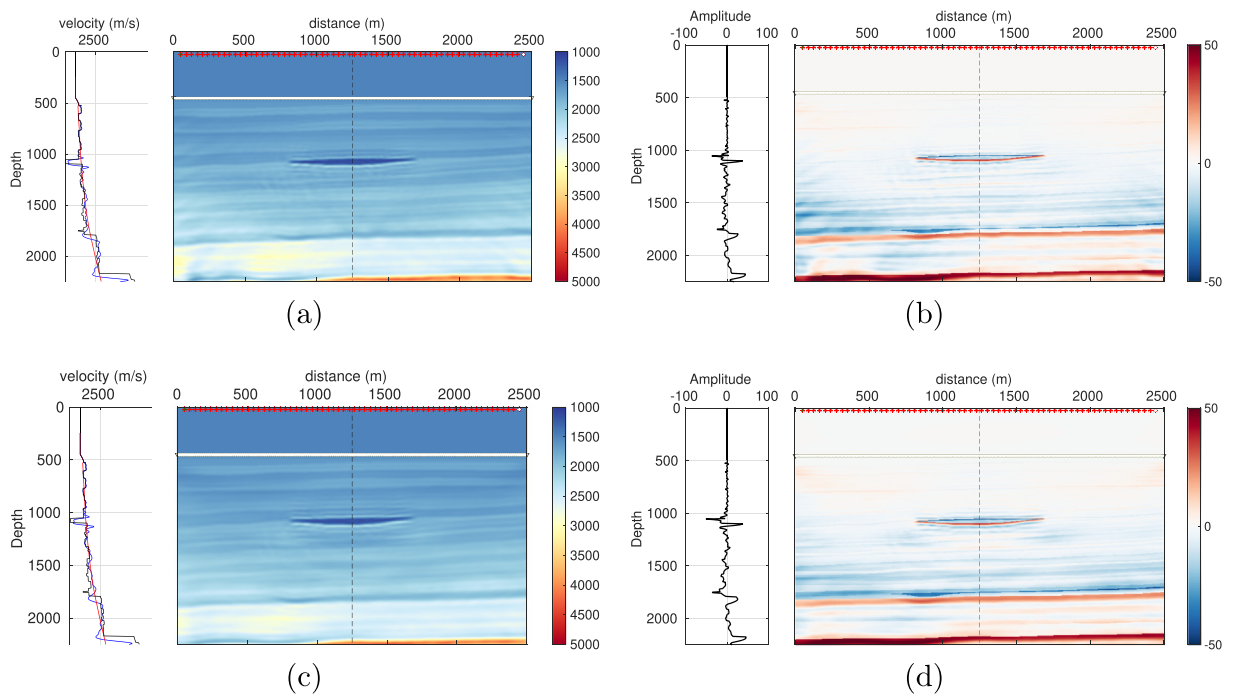


Fig. 13. Models estimated by the FWI using the broadband wavelets and the (a) L1 and (b) L2 norms. The dashed lines indicate the position of the shown profiles on the side of each model, the red line is the initial model profile, the black line is the real model profile, and the blue line the estimated model. The relative difference between c) true model and that estimated by norm L1 and d) true model and that estimated by norm L2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

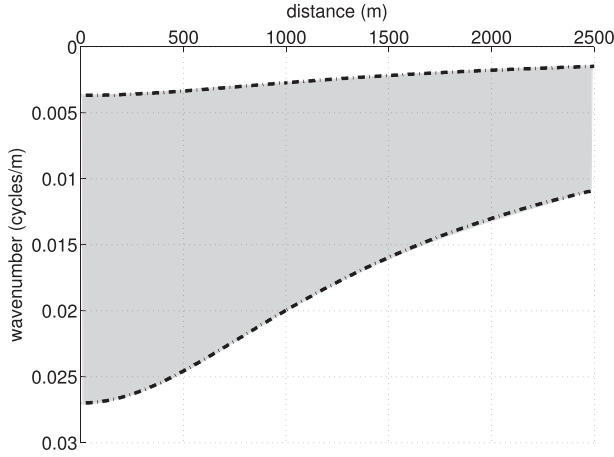


Fig. 14. The wavenumbers expected for the low-velocity anomaly present in the true velocity model (Fig. 6a) as a function of source-receiver offset.

true and initial models, estimated by L1 and L2 norms, respectively. The relative difference is lower in the estimated velocity models.

An additional inversion test was carried out using a broadband wavelet (Eq. (16)) and the parameters $p = 5$, $q = 25$, $fb = 20$ and

$m = 8$ to obtain the observed dataset. To apply the multiscale approach, the observed data and the wavelet were filtered with the same low-pass filter. The frequencies used in these tests started with a cutoff frequency of 12 Hz up to 30 Hz as shown in Fig. 11a and b.

The convergence evolution of the normalized objective function for the L1 and L2 Norms are shown in Fig. 12. Again, it is possible to see that the objective function achieves a minimum in both cases.

The models estimated using the broadband wavelets for L1 and L2 norms are shown in Fig. 13a and c, respectively. The relative differences between the true models and those estimated by both norm L1 (Fig. 13b) and norm L2 (Fig. 13d) confirm that the models converge to a good solution in both cases.

3.5. Analysis of the retrieved wavenumbers

Fig. 14 shows the expected vertical wavenumbers as a function of the source-receiver offset to low-velocity anomaly in the true model (Fig. 6a) at 1100 m depth. Eq. (15) was used choosing the minimum and maximum effective frequencies (3 Hz and 22 Hz, respectively) that are present in the source wavelet (Figs. 7b and 11b) to determine the lower and upper limits of the illuminated wavenumbers (0.001 cycles/m and 0.027 cycles/m, respectively). Additionally, the average velocity above the region of the anomaly, 1630 m/s, was used to calculate the wavenumbers.

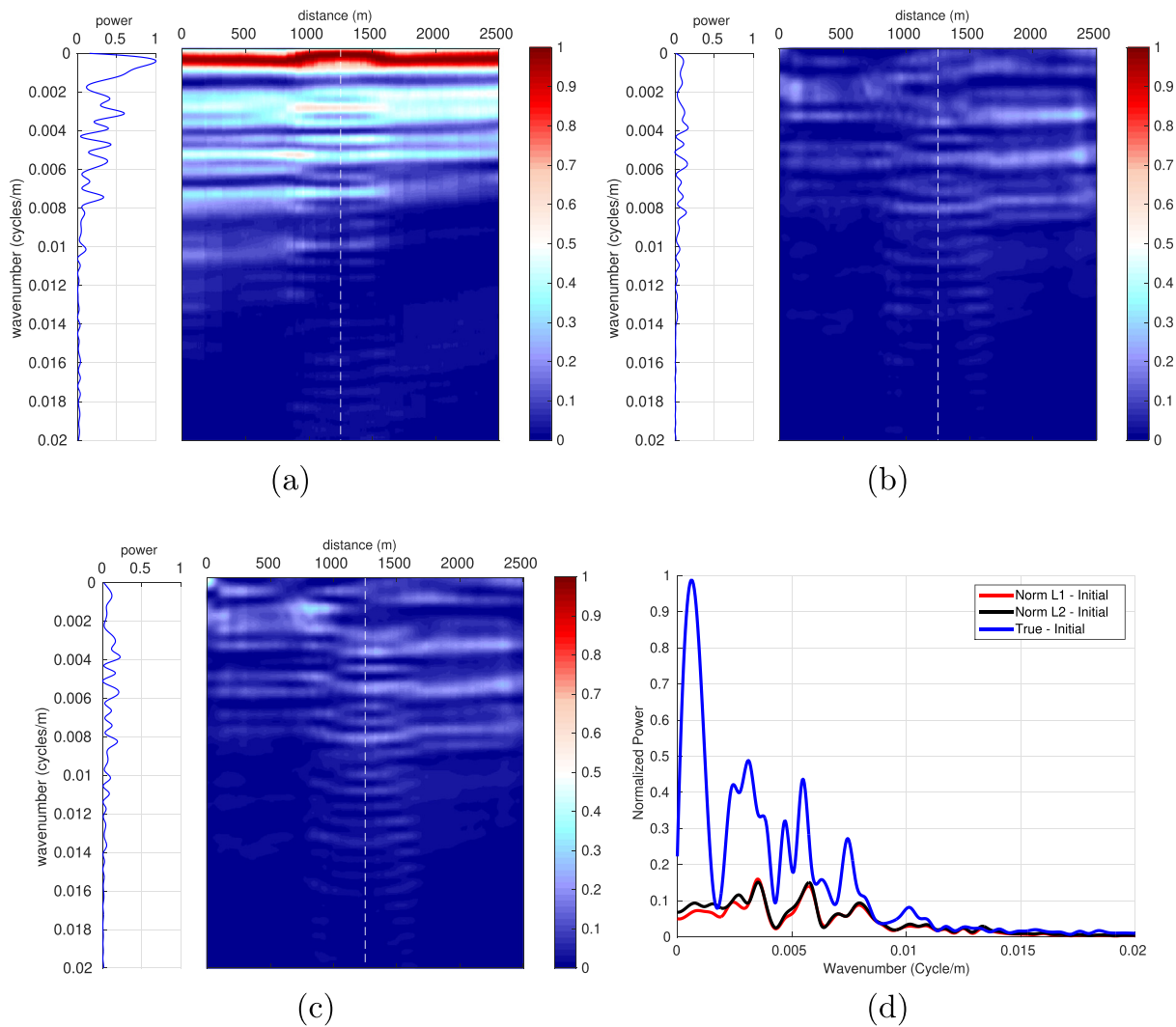


Fig. 15. 1D Fourier transform in the vertical direction of the difference between (a) true model and initial model; (b) model estimated by the L1 norm and initial model (c) model estimated by the L2 norm and initial model. The dashed line indicates the position of the shown wavenumber profile on the side of each Figure. (d) 1D Fourier transform of the stacked traces.

A Fourier domain analysis was performed to understand the illumination of the wavenumbers by the FWI, similar to section 3.1. However, in this numerical experiment, the Fourier domain analysis was performed directly in the velocity models, rather than analyzing the wavenumbers retrieved by the objective function gradient (Eq. (9)), the analysis in the Fourier domain was applied in the vertical direction on the difference between the true model and initial models, as well as the difference between the inverted models and the initial model. This process was performed to remove the low wavenumbers contained in the initial model and to analyze only the wavenumbers inserted by the FWI.

Fig. 15a shows the content of reference wavenumbers obtained by the Fourier transform of the difference between the true and initial models. The contents of wavenumbers illuminated by the inversion in the first numerical experiment using the L1 and L2 norms are shown, respectively, in Fig. 15b and c. The mean behavior of the recovered wavenumbers (Fig. 15d, black and red lines) suggests that the wavenumber recovery limit is approximately 0.02 cycles/m in this context since the average behavior of the wavenumbers of the inverted models tends to zero from that point. The blue line, used as reference, is the average behavior of the wavenumbers of the difference between the true and initial models.

Similarly to the previous numerical experiment, Fig. 16 compares the retrieved wavenumbers in the case of the broadband wavelet test. Fig. 16a shows the reference wavenumbers, Fig. 16b and c show the

retrieved wavenumber using the L1 and L2 norms, respectively. The average behavior of the retrieved wavenumbers is shown in Fig. 16d in the red and black lines.

4. Discussion and conclusion

The Fourier domain analysis allowed to determine the mean limit of the velocity model wavenumbers illuminated by the FWI in the time domain using the L1 and L2 norms. In addition, we studied the wavenumbers illuminated for the 1D model (Fig. 3a), giving continuity to the numerical experiment proposed by Sirgue and Pratt (2004). The results showed that equivalent wavenumbers were retrieved by the gradient calculated in both domains, frequency and time.

The results obtained in the time domain are similar to those obtained in the frequency domain by Sirgue and Pratt (2004), even though we know that the wavelets used to calculate the wavenumbers for a source-receiver pair (Fig. 3) do not have a discrete spectrum. Expression (15), which calculates the wavenumbers as a function of the source frequency and source-receiver offset, was compared to the wavenumbers (Fig. 5b and d) obtained numerically by the adjoint method in the time domain. The observed fitting between the numerically and analytically obtained wavenumbers confirms the equivalence of the different methods used for calculating the gradient of the objective function.

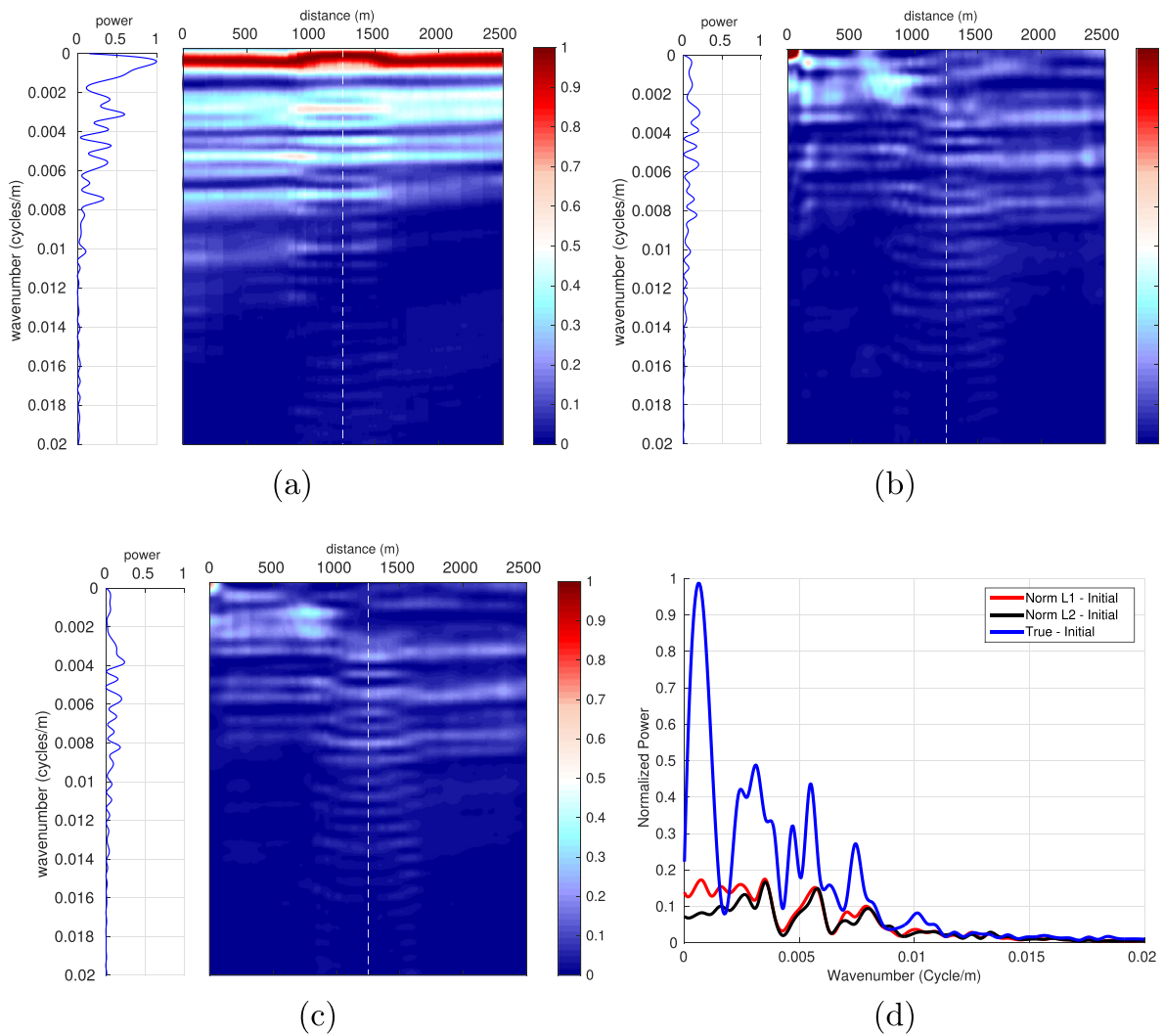


Fig. 16. 1D Fourier transform in the vertical direction of the difference between (a) true model and initial model; (b) model estimated by the L1 norm and initial model (c) model estimated by the L2 norm and initial model. The dashed line indicates the position of the shown wavenumber profile on the side of each Figure. (d) 1D Fourier transform of the stacked traces.

The inverted Marmousi models shown in Fig. 10c and e were used to analyze the wavenumbers illuminated by FWI (Fig. 15b and c, respectively). In both cases, the inversion was able to recover part of the wavenumbers present in the true model (Fig. 15a), mainly filling the intermediate wavelength (in this case, between 0.003 and 0.01). The comparison between the average behavior of the wavenumbers illuminated by the L1 and L2 norms (Fig. 15d) indicate that the L2 norm achieved lower wavenumbers in the test using the Ricker wavelets. Furthermore, the performed experiments allowed to identify the limit of wavenumbers recovered by the inversion, which is 0.02 cycles/m.

An additional test was carried out using a broadband wavelet (Fig. 13c and d) to compare the retrieved wavenumber using the L1 and L2 norms. In this case, on average, the lower wavenumbers were illuminated more by norm L1 compared to norm L2. But these lower wavenumbers could be associated with the artifacts generated on the left part of the model.

The expected wavenumbers (Fig. 14) agree with the wavenumbers present in the low-velocity anomaly region of the inverted models (Figs. 15b, c, 16b, and c). The illuminated lower wavenumbers can be related to the low temporal frequencies and the far offsets, as suggested in Fig. 14, while the higher wavenumbers can be related to higher temporal frequencies and short offsets.

The analysis of the wavenumbers illuminated by the inversion can be incorporated as a step in the quality control of the models estimated by FWI, even in 3D cases. This analysis allows a quantitative comparison of the vertical wavenumbers according to the position of the estimated model once the Fourier transform is applied only in the vertical direction. This procedure can be useful to determine the wavenumber of the geological structures illuminated by the FWI.

Acknowledgments

The authors would like to thank CAPES and PETROBRAS for funding the research, UFF and GISIS for the infrastructure and support, PGS for donating the computer clusters in which the tests were conducted, and the members of GISIS, especially Rodrigo Stern for the TI support.

References

- Bunks, C., Saleck, F.M., Zaleski, S., Chavent, G., 1995. Multiscale seismic waveform inversion. *Geophysics* 60 (5), 1457–1473.
- Cao, S., Han, D., 2011. Wide-band B-spline wavelet with four parameters. *SEG Technical Program Expanded Abstracts* 2011, pp. 3840–3844.
- Carneiro, M.R., Pereira-Dias, B., Soares Filho, D.M., Landau, L., 2018. On the scaling of the update direction for multi-parameter full waveform inversion: applications to 2D acoustic and elastic cases. *Pure Appl. Geophys.* 175 (1), 217–241.
- Claerbout, J.F., 1971. Toward a unified theory of reflector mapping. *Geophysics* 36 (3), 467–481.
- Claerbout, J.F., 1985. *Imaging the Earth's Interior*. 1. Blackwell Scientific Publications, Oxford.
- Fichtner, A., 2011. *Full Seismic Waveform Modelling and Inversion*. Springer Science & Business.
- Fichtner, A., Kennett, B.L., Igel, H., Bunge, H.P., 2008. Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. *Geophys. J. Int.* 175 (2), 665–685.
- Huang, Y., Schuster, G.T., 2014. Resolution limits for wave equation imaging. *J. Appl. Geophys.* 107, 137–148.
- Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., et al., 1989. Wavelengths of earth structures that can be resolved from seismic reflection data. *Geophysics* 54 (7), 906–910.
- Lailly, P., 1983. The Seismic Inverse Problem as a Sequence of Before Stack Migrations, *Conference on Inverse Scattering, Theory and Applications*. Society for Industrial and Applied Mathematics, pp. 206–220.
- Levander, A., 1989. Finite-difference forward modeling in seismology. *Encyclopedia of Solid Earth Geophysics* 410–431.
- Martin, G.S., Wiley, R., Marfurt, K.J., 2006. Marmousi2: an elastic upgrade for Marmousi. *Lead. Edge* 25 (2), 156–166.
- Martins, M., 2015. Estratégias para inversão do campo de onda Completo associado a fontes sísmicas reais e virtuais. Ph.D. thesis. Universidade Federal do Rio de Janeiro.
- Menke, W., 2012. *Geophysical Data Analysis: Discrete Inverse Theory: MATLAB Edition*. 45. Academic Press.
- Métivier, L., Brossier, R., 2016. The SEISCOPE optimization toolbox: a large-scale nonlinear optimization library based on reverse communication. *The SEISCOPE optimization toolbox*. *Geophysics* 81 (2), F1–F15.
- Métivier, L., Brossier, R., Virieux, J., Operto, S., 2012. Optimization schemes for Full Waveform Inversion : The preconditioned truncated Newton method. *SEG 2012 Annual Meeting*, pp. 1–5.
- Nocedal, J., Wright, S., 2006. *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering. Springer New York ISBN 9780387303031.
- Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophys. J. Int.* 167 (2), 495–503.
- Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics* 64 (3), 888–901.
- Ratcliffe, A., Win, C., Vinje, V., Conroy, G., Warner, M., Umpleby, A., Stekl, I., Nangoo, T., Bertrand, A., 2011. Full Waveform Inversion: A North Sea OBC Case Study. *SEG San Antonio 2011 Annual Meeting* 2384–2388.
- Schuster, G., 2017. *Seismic Inversion*. Society of Exploration Geophysicists.
- Sirgue, L., Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies. *Geophysics* 69 (1), 231–248.
- Sirgue, L., Barkved, O.I., Dellinger, J., Etgen, J., Albertin, U., Kommedal, J.H., 2010. Full waveform inversion: the next leap forward in imaging at Valhall. *First Break* 28 (4), 65–70.
- Soares, D.M.F., Souza, A.A.V.B., Bulcão, A., Dias, B.P., Martins, E.O., Farias, F.F., Loureiro, F.P., Santos, L.A., dos Santos, M., 2015. Full Waveform Inversion: investigating resolution on a Santos Basin model. 14th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, Brazil, 3–6 August 2015, pp. 323–326.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation. *Geophysics* 49 (8), 1259–1266.
- Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data. 56th Ann. Internat. Mtg. 51 (10) (Session: S12.4).
- Versteeg, R., 1994. The Marmousi experience: Velocity model determination on a synthetic complex data set. *Lead. Edge* 13 (9), 927–936.
- Virieux, J., Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics. *Geophysics* 74 (6), WCC1–WCC26.
- Wirgin, A., 2004. *The Inverse Crime*.
- Woodward, M.J., 1992. Wave-equation tomography. *Geophysics* 57 (1), 15–26.