UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GEOFÍSICA

DISSERTAÇÃO DE MESTRADO

Determinação de parâmetros ótimos do método inversão da forma de onda completa para o caso 1-D acústico

ELIANE SANTOS CARDOSO

Belém-Pará
2019

ELIANE SANTOS CARDOSO

Determinação de parâmetros ótimos do método inversão da forma de onda completa para o caso 1-D acústico

Dissertação apresentada ao Programa de Pós-Graduação em Geofísica do Instituto de Geociências da Universidade Federal do Pará, para obtenção do título de Mestre em Geofísica.

Área de Concentração: Métodos Sísmicos

Linha de Pesquisa: Inversão e Processamento de dados Geofísicos

Orientador: Prof. Dr. João Carlos Ribeiro Cruz

Belém-Pará

2019

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

C268d Cardoso, Eliane Santos
Determinação de parâmetros ótimos do método inversão da forma de onda completa para o caso 1-D acústico / Eliane Santos Cardoso. — 2019.
75 f. : il. color.

Orientador(a): Prof. Dr. João Carlos Ribeiro Cruz
Dissertação (Mestrado) - Programa de Pós-Graduação Geofísica, Instituto de Geociências, Universidade Federal do Pará, Belém, 2019.
1. Inversão. 2. FWI. 3. Modelo de velocidade. I. Título.

CDD 550

ELIANE SANTOS CARDOSO

Determinação de parâmetros ótimos do método inversão da forma de onda completa para o caso 1-D acústico

Dissertação apresentada ao Programa de Pós-Graduação em Geofísica do Instituto de Geociências da Universidade Federal do Pará, para obtenção do título de Mestre em Geofísica.

Data de aprovação: 27 de junho de 2019

Banca Examinadora:

Prof. Dr. João Carlos Ribeiro Cruz (Orientador)
Universidade Federal do Pará

Manuel de Jesus dos S. Costa.

Prof. Dr. Manuel de Jesus dos Santos Costa
Universidade Federal do Pará

Dr. Paulo Eduardo Miranda Cunha
Petrobras

AGRADECIMENTOS

À Deus por me dar disciplina e força de vontade para finalizar o trabalho.

Ao prof. Dr. João Carlos Ribeiro Cruz pela orientação e oportunidade, pelo suporte teórico ao longo do trabalho.

À minha mãe, Elaine Santos dos Santos, por impor desde cedo a prática dos estudos. Essa dissertação é uma consequência direta disso.

Agradeço aos meus tios e tias, Rui Fernando Santos, Paulo Sérgio Cardoso, Marilene Cardoso, Danielle Santos e Lilian Santos por todo o apoio durante o processo.

Ao Mateus Winker por me apoiar e incentivar nos momentos mais difíceis.

Às amigas Diana Dax e Maria Clara Acacio pela fidelidade e apoio de sempre.

À amiga Bruna Boulhosa pela amizade e incentivo.

Ao Akira Takeda pelo apoio e dedicação.

Aos Profs. Dr. Marcelo Jorge Luz Mesquita e Dr. Manuel de Jesus dos Santos Costa pelo apoio e sugestões valiosas.

Aos secretarios Benildes Lopes Rodrigues de Souza e Kleberson Moura do Programa de Pós Graduação em Geofísica pela disponibilidade em me ajudar nos mais diversos problemas durante esses dois anos.

Aos amigos Daniele Pantoja Monteiro, Danilo da Silva e Jozinei Ferreira Lopes pelo apoio, amizade e companhia para tomar café.

Ao amigo Murilo Santiago Vale Rodrigues pela ajuda nos problemas computacionais.

Agradeço à CAPES pelo financiamento do projeto mediante bolsa de mestrado CAPES-DS.

À Petrobras pelo apoio na infraestrutura do laboratório de inversão da onda sísmica do Instituto de Geociências.

Ao Prof. Dr. Thomas Bohlen da universidade de Karlsruhe por disponibilizar o software de inversão IFOS2D.

De um modo geral agradeço a todos que ajudaram de alguma forma para a realização desta dissertação de mestrado.

RESUMO

Este trabalho é um estudo sobre o método de inversão da forma de onda completa, do inglês "Full Waveform Inversion"(FWI) que será a denominação adotada durante esta dissertação, por ser a mais utilizada na literatura. O método FWI foi aplicado utilizando dados sintéticos de reflexão para o meio acústico. O FWI foi estudado com o objetivo de obter a melhor estimativa do modelo de velocidade da onda P. A metodologia de inversão foi desenvolvida através do programa IFOS2D, que realizou busca local através do método L-BFGS que calcula a aproximação da matriz Hessiana a fim de estimar os parâmetros físicos do meio. Foram realizados testes com parâmetros da inversão como a escolha da aproximação inicial a fim de avaliar a convergência do método, e a parametrização do dado observado através da filtragem de frequências e janelamento de afastamentos a fim de melhorar a precisão e acurácia do método. Os resultados mostraram que o sucesso do método é fortemente dependente da escolha do modelo inicial e da escolha das bandas de frequências e intervalos de afastamentos para o janelamento do dado durante o processo de inversão. Também foram realizados testes com a adição de ruído ao dado sintético a fim de avaliar a estabilidade do método, o que mostrou a necessidade do pré-processamento para aumentar a razão sinal-ruído. O modelo de densidade e a "wavelet"fonte são tidos como conhecidos em todos os experimentos. Concluindo, os resultados apresentados na dissertação mostram que o FWI é aplicável para a melhoria do modelo de velocidade da onda P desde que se tenha disponível informações a priori para suprimir as dificuldade do método, em especial a reconstrução das partes mais profundas do modelo.

Palavras-chaves: Inversão. FWI. Modelo de velocidade.

ABSTRACT

This work is a study of the full waveform inversion method (FWI), which will be the denomination adopted during this dissertation because it is the most used in the literature. The method was applied using synthetics for the acoustic medium. The inversion methodology was developed through the IFOS2D program, which performs local search using the L-BFGS method that calculates the approximation of the Hessian matrix in order to estimate the physical parameters of the medium. The FWI was studied in order to obtain the best estimate of the P wave velocity model. We performed tests with inversion parameters such as the choice of the initial approximation in order to evaluate the convergence of the method and the parameterization of the observed data through filtering of frequencies and windowing of spacings in order to improve the accuracy of the method. The results showed that the success of the method is strongly dependent on the choice of the initial model and the choice of the frequencies and offsets intervals for windowing the data during the inversion process. We also performed tests with the addition of noise to the synthetic data in order to evaluate the stability of the method, which showed the need for pre-processing to increase the signal-to-noise ratio. The density model and the source wavelet are as known in all experiments. In conclusion, the results presented in dissertation show that the FWI is applicable for the improvement of the velocity model of the P wave provided information is available a priori to suppress the difficulty method to reconstruct the deeper parts of the model.

Keywords: inversion. FWI. velocity model.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A indústria de óleo e gás desenvolve continuamente ferramentas para obter informações sobre os reservatórios em subsuperfície de forma indireta a fim de evitar a perfuração de poços e baratear o custo da exploração. Nesse contexto, a inversão sísmica é um método indireto amplamente utilizado. A caracterização de reservatórios através da inversão sísmica tem como objetivo gerar um modelo de parâmetros físico a partir dos dados sísmicos observados. Vários métodos de inversão foram desenvolvidos, dentre os quais, a inversão da forma da onda completa (FWI) é eficiente pra calcular modelos multiparâmetros (velocidade da onda p, velocidade da onda s, densidade) (Przebindowska, 2013).

O método de inversão da forma da onda completa (FWI) é uma técnica de imageamento de alta resolução que trabalha utilizando todo o conteúdo contido nos sismogramas observados (fase e amplitude) para obter os parâmetros físicos de interesse. Esses parâmetros são calculados pelo método adjunto com o objetivo de fornecer o menor valor de desajuste entre os dados observados e modelados. A característica diferencial do método FWI é a capacidade de vizualizar estruturas na subsuperficie menores que o comprimento de onda.

Efeitos elásticos, atenuação, ruído presente nos dados e uma assinatura de fonte desconhecida são fatores que reduzem o desempenho do algoritmo de inversão. Para eliminar esses problemas, é necessário pré-processar os dados, estimar um modelo inicial preciso e definir a "wavelet"da fonte. Também são utilizados métodos de pré-condicionamentos e a inclusão de informações a priori sobre os parâmetros do modelo para garantir a convergência do algoritmo e eliminar efeitos indesejados.

Em 1984, A. Tarantola propôs a teoria da inversão da forma de onda completa acústico no domínio do tempo, seguido da aproximação elástica e para o caso viscoelástico (Tarantola, 1984; Tarantola, 1986; Mora, 1987; Tarantola, 1988).

No início dos anos 90, Pratt (1990) e Pratt and Worthington (1990) propôs o método no domínio da frequência. O FWI elástico é mais complexo que a abordagem acústica, pois é necessário otimizar simultaneamente três parâmetros elásticos acoplados (velocidade da onda P, velocidade da onda S e densidade), o que requer maior poder computacional.

Existem diversas aplicações da FWI para escala de petróleo e para a sismologia (Virieux and Operto, 2009), mas o interesse pela escala rasa surgiu recentemente, Romdhane et al. (2011) apresenta um algoritmo de FWI 2D no domínio da frequência para avaliar uma situação topografia complexa através de dados sintéticos. Os trabalhos Köhn (2011), Kurzmann (2012) e Przebindowska (2013) mostram de forma aprofundada a abordagem multiescala através da filtragem de frequências para reduzir a não-linearidade do problema inverso. Shah et al. (2012) mostra a aplicação do FWI em um conjunto de dados de campo 3D com significativa anisotropia de velocidade.

O trabalho Spadini and Diogo (2018) utiliza busca global para avaliar a sensibilidade da função objetivo para cada parâmetro do modelo em diferentes janelas de afastamentos em relação à fonte. Koehene (2018) apresenta a implementação multiescala utilizando o método de inversão L-BFGS para minimizar a função objeto.

O objetivo desta dissertação é fazer um estudo sobre o método FWI executando a inversão em dados sintéticos para avaliar, inicialmente, a dependência da convergência do método FWI com a escolha do modelo inicial, seguido do estudo da abordagem multi-escala através de testes de regularizações no dado através da filtragem de frequências e janelamento de afastamentos para obter o modelo de velocidade da onda P. Além disso, é feita a análise da estabilidade do método através da inversão do dado sintético com adição de ruído.

Esta dissertação está dividida em quatro capítulos e dois anexos. No capítulo 2 é apresentada a teoria da inversão de problemas não-lineares para otimização local. Em seguida, é apresentada a teoria do algoritmo de inversão da forma de onda completa no domínio do tempo com abordagem adjunta e aproximação acústica. O capítulo é finalizado com a descrição das exigências de aplicação do método.

No capítulo 3 é apresentada aplicação do método utilizando dados sintéticos. Foram realizados testes para avaliar a dependência da convergência do método com a escolha do modelo inicial. Também são realizados testes para avaliar a estratégia de inversão multiescala no dado observado,ou seja, todos os experimentos realizados neste trabalho foram feitos usando o pacote de programas IFOS2D disponibilizado através do site `https://git.scc.kit.edu/GPIAG-Software/IFOS2D` da universidade de Karlsruhe, Alemanha, sob coordenação do Prof. Thomas Bohlen. A estratégia é avaliada através da filtragem de frequências e janelamento de afastamentos. O capítulo é finalizado com a comparação do resultado de inversão para diferentes valores de ruído aleatório no dado observado.

No capítulo 4 são apresentadas as conclusões. O anexo A apresenta a solução da equação de onda acústica 3D e 2D segundo a implementação de camadas perfeitamente combinadas (PML) e o anexo B apresenta os métodos de otimização: gradiente conjugado, método de Newton e quasi-Newton.

2 TEORIA DO FWI E IMPLEMENTAÇÃO

Os dados de fundo oceânico são especialmente adequados para o FWI, pois costumam apresentar informações de longos afastamentos associados às baixas freqüências e uma gama completa de azimutes. Além disso, após a aplicação da reciprocidade fonte-receptor, o número total de fontes efetivas não é maior e, portanto, esses conjuntos de dados podem ser invertidos com relativa eficiência usando algoritmos para os quais o custo total é dimensionado com o número de fontes efetivas.

A figura 2.1 mostra cortes horizontais em profundidade através de dois modelos de velocidade isotrópica sobre o campo de Vahall no Mar do Norte. Os dados usados para gerar os modelos de velocidade foram adquiridos usando uma matriz de cerca de 2400 sensores de pressão no fundo do oceano registrando cerca de 100.000 tiros de ar comprimido. Após as aplicações de reciprocidade, isso representa apenas 2.400 fontes de pressão efetiva.

Na figura 2.1, os painéis superiores mostram o modelo de velocidade obtido a partir modelo de velocidade convencional, enquanto que os painéis inferiores mostram os mesmo cortes gerados pelo FWI. Os cortes da esquerda correspondem à parte rasa na seção, próximas ao fundo do mar. Os cortes do lado direito correspondem às profundidades intermediárias de cerca de um quilômetro, onde interceptam uma região de gás que cobre o reservatório de petróleo mais profundo. A melhoria na resolução espacial obtida pelo FWI é intensamente aparente.

A figura 2.2 mostra os resultados do uso desses dois modelos de velocidade para migrar em profundidade os dados de reflexão de Vahall. A imagem migrada a partir do modelo do FWI é claramente superior em todas as profundidades, pois mostra mais clareza no deliamento de estruturas, o que é necssário para migrar os dados corretamente.

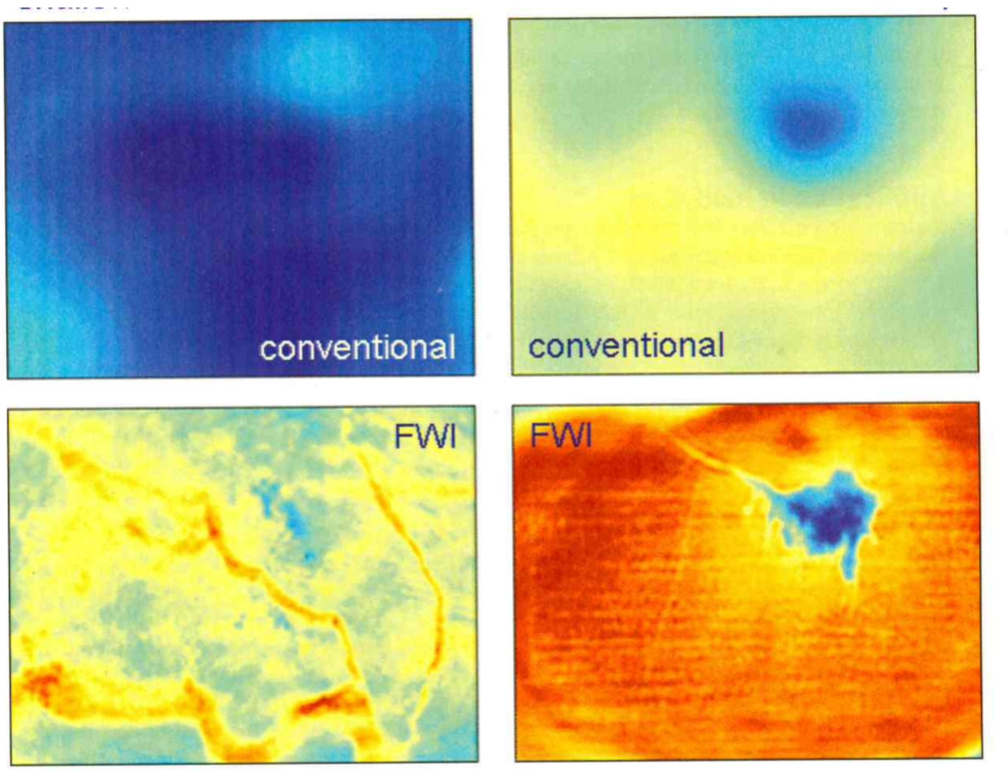
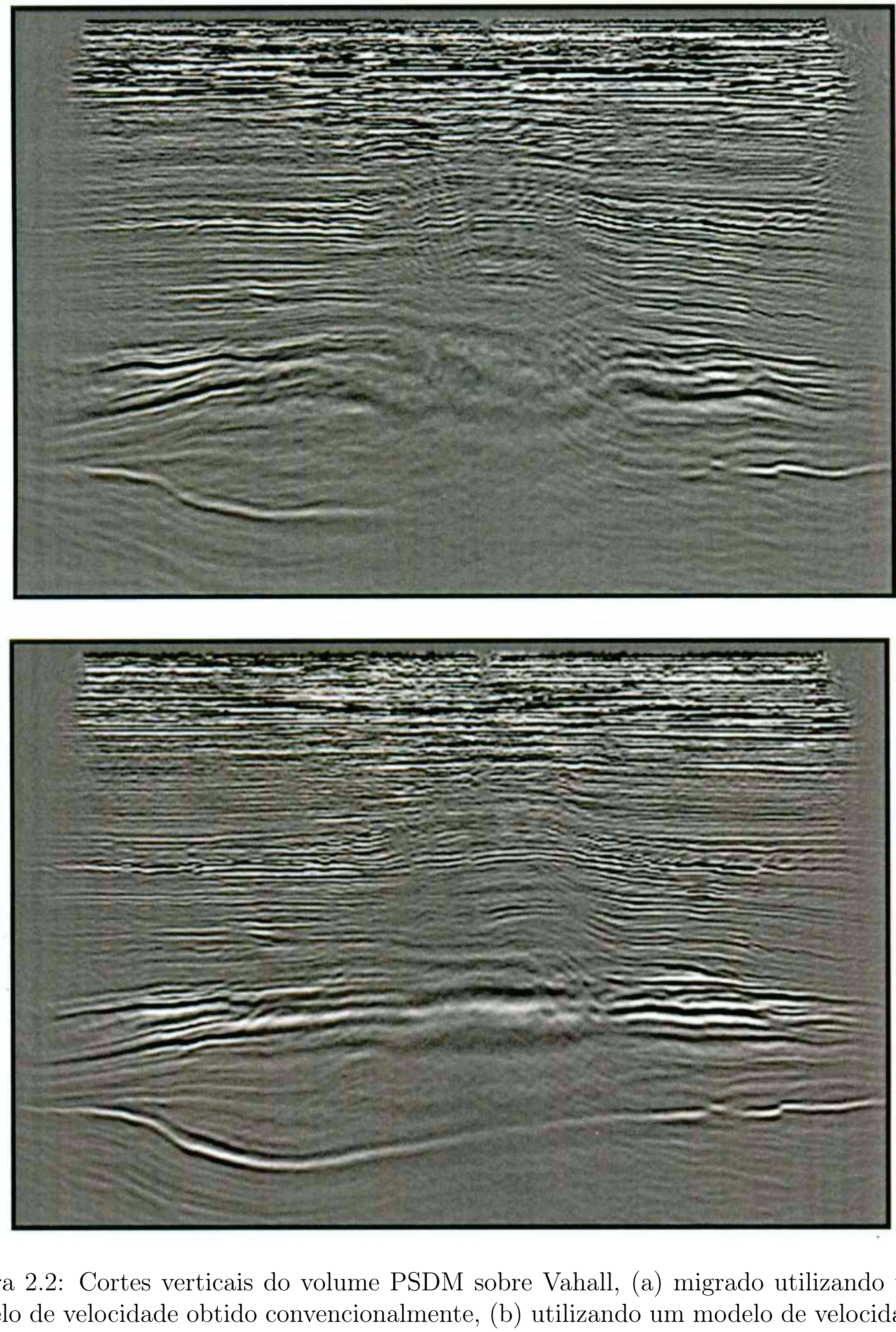


Figura 2.1: Cortes horizontais através do modelo de velocidade de Vahall. Painéis superiores: métodos convencional. Painéis inferiores: FWI. Warner (2012)



(a)

(b)

Figura 2.2: Cortes verticais do volume PSDM sobre Vahall, (a) migrado utilizando um modelo de velocidade obtido convencionalmente, (b) utilizando um modelo de velocidade obtido a partir do FWI. As seções tem cerca de 13 km de comprimento por 4500 m de profundidade. Warner (2012)

2.1 TEORIA DA INVERSÃO SÍSMICA

A inversão sísmica tem por objetivo encontrar um grupo de parâmetros físicos $\mathbf{m}$ do modelo capazes de descrever a subsuperfície da Terra a partir de dados sísmicos $\mathbf{d}$. Esse problema é complicado, pois as ondas sísmicas injetadas na Terra se relacionam de forma não-linear com os parâmetros desconhecidos do meio. Esse problema é descrito pela expressão do problema direto não-linear:

$$\mathbf{d} = f(\mathbf{m}), \tag{2.1}$$

onde f é o operador que descreve a relação físico-matemática entre o campo e os parâmetros do modelo. Na inversão clássica, esse problema seria facilmente resolvido calculando o operador f^{-1} para obter $\mathbf{m}$ através da relação $\mathbf{m} = f^{-1}(\mathbf{d})$, entretanto a alta não-linearidade do problema sísmico impede a resolução analítica do problema. A situação mais simples do problema (2.1) corresponde à aproximação acústica monoparamétrica, que corresponde ao caso abordado nesta dissertação, em que $\mathbf{m}$ corresponde à velocidade da onda P.

Na sísmica de exploração, os receptores posicionados na superfície da Terra registram o campo $\mathbf{d}$ ao longo do tempo, o que resulta nos sismogramas, ou dado observado $\mathbf{d}_{obs}$. Além disso, existe a aproximação inicial $\mathbf{m}_0$ que corresponde à informação a priori sobre geologia do local. Nesta dissertação m_0 corresponde à uma aproximação do modelo verdadeiro $\mathbf{m}$, pois só foram realizados testes sintéticos. Assim, é construída a formulação de mínimos quadrados da FWI, que consiste em calcular o melhor modelo $\mathbf{m}$ a partir da aproximação inicial $\mathbf{m}_0$ em iterações sucessivas para minimizar a função objeto, ou função desajuste.

O desajuste de dados é definido como a diferença entre os dados modelados e os dados observados

$$\delta\mathbf{d} = \mathbf{d}_{mod} - \mathbf{d}_{obs} = f(\mathbf{m}) - \mathbf{d}_{obs}, \tag{2.2}$$

onde $\delta\mathbf{d}$ é o vetor desajuste de dimensão N, e $\mathbf{m}$ é o vetor modelo de dimensão M. A fim de ajustar os dados modelados aos dados oservados, é comum utilizar como critério a minimização da função objeto descrita pela norma L2 do erro entre os dados modelados e observados.

$$E(\mathbf{m}) = \frac{1}{2}\delta\mathbf{d}^T\delta\mathbf{d} = \frac{1}{2}\sum_{i=1}^{N}\delta d_i^2 \tag{2.3}$$

onde E($\mathbf{m}$) é a função objeto (também chamada de função desajuste), e o sobrescrito T denota a transposição da matriz. O somatório é realizado sobre o número de pares fonte-receptor e o número de amostras de tempo. A função objeto mede a correspondência entre os dados observados e calculados quando um parâmetro do modelo é alterado. Dessa forma, o principal objetivo do FWI é encontrar o mínimo da função objeto a partir de um modelo inicial, como é mostrado na Figura 2.3.

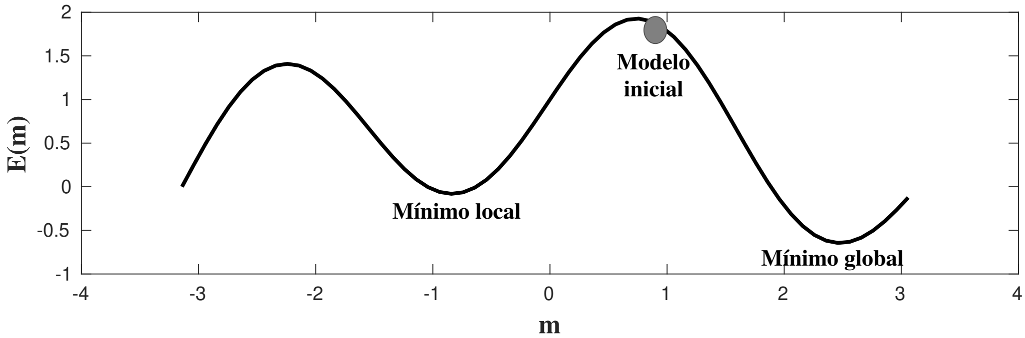


Figura 2.3: Representação da função objeto.

A presença de mínimos locais é um problema associado à alta não-linearidade do problema inversão. O modelo de aproximação inicial necessita estar proximo o suficiente do mínimo global para evitar que o modelo inverso fique preso no mínimo local e gere soluções incorretas de parâmetros do modelo. Dessa forma, O FWI pode ser tratado como um problema de otimização local iterativo através da teoria da perturbação e aproximação de Born (Virieux and Operto, 2009). O método da perturbação se baseia na idéia de que uma perturbação linear no espaço do modelo causa uma perturbação linear no espaço dos dados. Assim, a abordagem de Born procura o melhor modelo de ajuste $\mathbf{m}$ na vizinhança do modelo inicial m_0 tal que:

$$\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m}, \tag{2.4}$$

onde $\delta\mathbf{m}$ é a perturbação do modelo inicial. Significa que $\mathbf{m}$ é a perturbação de um meio de referência $\mathbf{m}_0$. Uma dedução detalhada por teoria da perturbação, à luz da aproximação de Born, para FWI acústica, pode ser encontrada em Yang et al. (2015).

A formulação de Tarantola (1984) diz que a perturbação dos parâmetros do modelo $\delta\mathbf{m}$ deve ser fixado na função objeto (2.3) através da aproximação de $E(\mathbf{m}_0 + \delta\mathbf{m})$ em torno de $E(\mathbf{m}_0)$ utilizando a expansão de primeira ordem da série Taylor:

$$E(\mathbf{m}_0 + \delta\mathbf{m}) = E(\mathbf{m}) = E(\mathbf{m}_0) + \frac{\partial E(\mathbf{m}_0)}{\partial\mathbf{m}}\delta\mathbf{m} + \vartheta(\mathbf{m}^2). \tag{2.5}$$

Os resíduos de ordens superiores, $\vartheta(m^2)$, são negligenciados.

Para encontrar o mínimo da função objeto E($\mathbf{m}$) como mostrado na Figura 2.3, a derivada da Equação (2.5) em relação aos parâmetros do modelo deve ser igual a zero, de modo que

$$\frac{\partial E(\mathbf{m})}{\partial\mathbf{m}} = \frac{\partial E(\mathbf{m}_0)}{\partial\mathbf{m}} + \frac{\partial^2 E(\mathbf{m}_0)}{\partial\mathbf{m}^2}\delta\mathbf{m} = 0 \tag{2.6}$$

O que leva ao modelo de perturbação

$$\delta\mathbf{m} = \left[\frac{\partial^2 E(\mathbf{m}_0)}{\partial \mathbf{m}^2}\right]^{-1} \frac{\partial E(\mathbf{m}_0)}{\partial \mathbf{m}} \tag{2.7}$$

O gradiente negativo da função objeto representa a direção de descida mais íngreme. Pode-se concluir então que a perturbação $\delta\mathbf{m}$, que leva ao mínimo do problema mínimos quadrados local, é "iniciada"na direção de descida mais íngrime, e sofre uma transformação linear através da inversa da matriz Hessiana $\mathbf{H}^{-1}$, gerando a direção que minimiza o problema local (Koehene, 2018). onde $\mathbf{H}$ é a matriz Hessiana de dimensão $M \times M$, que define a curvatura da função objeto.

A primeira derivada da função objeto (2.5) em relação aos parâmetros do modelo é dada

$$\frac{\partial E(\mathbf{m}_0)}{\partial \mathbf{m}} = \left(\frac{\partial \mathbf{d}_{mod}(\mathbf{m}_0)}{\partial \mathbf{m}}\right)^T (\mathbf{d}_{mod}(\mathbf{m}_0) - \mathbf{d}_{obs}) = \mathbf{J}^T \delta\mathbf{d}, \tag{2.8}$$

onde $\mathbf{J}^T$ é a transposição da matriz Jacobiana $\mathbf{J}$ também chamada de matriz derivada de Fréchet. Os elementos individuais da matriz são

$$J_{ij} = \frac{\partial d_{mod_i}}{\partial m_j}(i = 1, 2, .., N), (j = 1, 2, .., M), \tag{2.9}$$

onde N é o número de observações e M é o número de parâmetros do modelo. A segunda derivada da função objeto em relação aos parâmetros do modelo é

$$\frac{\partial^2 E(\mathbf{m}_0)}{\partial \mathbf{m}^2} = \mathbf{H} = \mathbf{J}^T\mathbf{J} + \left(\frac{\partial \mathbf{J}}{\partial \mathbf{m}}\right)^T \delta\mathbf{d}, \tag{2.10}$$

Substituindo a Equação (2.8) e (2.10) na Equação (2.7) obtem-se:

$$\delta\mathbf{m} = -\mathbf{H}^{-1}\mathbf{J}^T\delta\mathbf{d} \tag{2.11}$$

A Equação (2.11) representa o cálculo de δm através do método de newton. É inviável utilizar esse método computacionalmente devido a necessidade de N perturbações e N modelagens para o cálculo de $\nabla E(\mathbf{m})$, além de N^2 modelagens para $\mathbf{H}$. Assim, é necessário utilizar métodos que calculem $\delta\mathbf{m}$ de maneira aproximada, como os métodos Gauss-Newton, quasi-Newton e Steepest-Descent.

Os métodos de quasi-Newton aproximam a Hessiana inversa nos cálculos de primeira derivada e de segunda derivada, a presença dessas informações de segunda derivada tornam o método mais preciso que os métodos de Guass-Newton e Steepest-Descent. Nos experimentos dessa dissertação foi utilizado o método quasi-Newton L-BFGS apresentado no Anexo B.

2.2 INVERSÃO DE FORMA DE ONDA COMPLETA NO DOMÍNIO DO TEMPO

O código de inversão da forma de onda completa acústica, utilizado nesta dissertação, é baseado na abordagem geral de (Tarantola, 1984) e (Mora, 1987) com formulação no domínio do tempo, mas também existe a possibilidade de realizar a abordagem no domínio da frequência (Pratt and Worthington, 1990;Pratt, 1990). Os resíduos entre os dados modelados e os observados são minimizados em um processo iterativo com o objetivo de encontrar um modelo de subsuperfície. Neste método, a maior dificuldade é calcular a Matriz de derivada de Fréchet $\mathbf{J}$, pois o seu cálculo explícito exigiria a modelagem do campo de onda correspondente à cada perturbação individual dos parâmetros do modelo. Assim, uma única simulação do campo de ondas exigiria N modelagens diretas. A fim de diminuir o custo computacional, a abordagem adjunta (Mora, 1987; Tarantola, 1984) é utilizada para calcular o gradiente da função objeto $\nabla E(\mathbf{m})$ de forma implícita, em que a direção do gradiente é obtido pela correlação cruzada entre o campo de onda propagado para frente e o campo de onda residual retropropagado. Dessa forma, o gradiente da função objeto é calculado de forma eficiente com $2 \times N$ simulações do campo de onda.

2.2.1 A abordagem adjunta

No caso anterior, a linearização do problema é feito na função objeto E($\mathbf{m}$), agora a resolução do problema inverso se inicia com a linearização do modelo inicial na vizinha do modelo de referência. O novo modelo $\mathbf{m}$ é definido como uma combinação linear entre o modelo de referência $\mathbf{m}_0$ e o modelo perturbado $\delta\mathbf{m}$, assim a relação entre os dados modelados e o modelo $\delta\mathbf{m}$ pode ser escrita como

$$\mathbf{d}_{mod} = f(\mathbf{m}) = f(\mathbf{m}_0 + \delta\mathbf{m}) \tag{2.12}$$

O operador direto não linear f pode ser aproximado usando a expansão da série de Taylor de primeira ordem:

$$f(\mathbf{m}_0 + \delta\mathbf{m}) = f(\mathbf{m}) = f(\mathbf{m}_0) + \frac{\partial f(\mathbf{m}_0)}{\partial \mathbf{m}}\delta\mathbf{m} + \vartheta(\mathbf{m}^2). \tag{2.13}$$

Uma pequena perturbação no espaço de dados $\delta\mathbf{d}$ que resulta de uma pequena perturbação nos parâmetros do modelo $\delta\mathbf{m}$ pode ser definida como:

$$\delta\mathbf{d} = f(\mathbf{m}_0 + \delta\mathbf{m}) - f(\mathbf{m}_0). \tag{2.14}$$

Substituindo a equação (2.13) na equação (2.14), temos:

$$\delta\mathbf{d} = f(\mathbf{m}_0) + \frac{\partial f(\mathbf{m}_0)}{\partial \mathbf{m}}\delta\mathbf{m} - f(\mathbf{m}_0) = \frac{\partial f(\mathbf{m}_0)}{\partial \mathbf{m}}\delta\mathbf{m} = \mathbf{J}\delta\mathbf{m}, \tag{2.15}$$

onde $\delta\mathbf{d}$ representa os dados residuais obtidos pela diferença entre os dados observados

$\mathbf{d}_{obs}$ e os dados calculados $\mathbf{d}_{mod}(\mathbf{m})$ pelo problema direto a partir de um modelo $\mathbf{m}$ que produz o melhor ajuste. A partir da Equação (2.15), observa-se que os dados residuais $\delta\mathbf{d}$ se comportam de forma linear com a perturbação do modelo $\delta\mathbf{m}$, e o operador linear $\mathbf{J}$ é a matriz derivada de Fréchet.

A Equação (2.15) pode ser escrita na forma contínua (Mora,1987)

$$\delta\mathbf{d}(D) = \int_M dM \frac{\partial\mathbf{d}(D)}{\partial\mathbf{m}}\delta\mathbf{m}(M), \tag{2.16}$$

onde M e D indicam o espaço do modelo e dos dados. Através da Equação (2.16), percebe-se que as perturbações no espaço dos dados podem ser calculadas através da integração das pequenas perturbações nos parâmetros do modelo, se a matriz derivada de Fréchet for conhecida.

O gradiente da função objeto (2.8), utilizado para atualizar os parâmetros do modelo no método do gradiente, pode ser escrito na forma contínua:

$$\nabla E(M) = \int_D dD \Big[\frac{\partial\mathbf{d}(D)}{\partial\mathbf{m}}\Big]^*\delta\mathbf{d}(D), \tag{2.17}$$

onde, o sobrescrito (*) significa adjunto.

2.2.2 O problema adjunto na aproximação acústica

Nesta dissertação, o método FWI é aplicado utilizando a aproximação acústica, por esse motivo será apresentado somente a abordagem adjunta para o caso acústico (Przebindowska, 2013). Essa abordagem necessita do fator da fonte para realizar a retropropagação do resíduo, assim será utilizada a equação de onda acústica de densidade variável não homogênea de segunda ordem 2D dada por:

$$\frac{1}{K(x,z)}\frac{\partial^2 p(x,z,t)}{\partial t^2} - \nabla.\Big(\frac{1}{\rho(x,z)}\nabla p(x,z,t)\Big) = s(x,z,t), \tag{2.18}$$

onde $K = \rho v_p^2$ é o módulo de Bulk, ρ é a densidade, v_p é a velocidade da onda P, p é a o campo de pressão, e s é o termo correspondente à fonte.

Seja um dado modelo $\mathbf{m}$ com a fonte localizada em $\mathbf{x} = \mathbf{x}_s$ e o receptor em $\mathbf{x} = \mathbf{x}_r$, a componente i do sismograma de pressão p no tempo t é definido como $d_i = f(\mathbf{m}) = p(\mathbf{x}_r, t; \mathbf{x}_s)$. O termo da fonte é:

$$s(\mathbf{x}, t; \mathbf{x}_s) = \delta(\mathbf{x} - \mathbf{x}_s)S(t), \tag{2.19}$$

onde $S(t)$ é a função de tempo fonte, e $\delta(\mathbf{x})$ é a função delta de Dirac.

Na formulação acústica, a Equação (2.16) assume a forma:

$$\delta d_i(\mathbf{x}_r, \mathbf{x}_s, t) = \int_V dV(\mathbf{x})\frac{\partial d_i(\mathbf{x}_r, \mathbf{x}_s, t)}{\partial\mathbf{m}(\mathbf{x})}\delta\mathbf{m}(\mathbf{x}), \tag{2.20}$$

enquanto o problema adjunto (2.17) é:

$$\nabla E(\mathbf{x}) = \sum_s \int dt \sum_r \frac{\partial d_i(\mathbf{x}_r, \mathbf{x}_s, t)}{\partial \mathbf{m}(\mathbf{x})} \delta d_i(\mathbf{x}_r, \mathbf{x}_s, t). \tag{2.21}$$

O gradiente da função objeto é a integral sobre o espaço de dados da derivada Fréchet multiplicada pelos dados residuais. O cálculo da expressão integral (2.20) no problema direto, que define a perturbação no campo de pressão δd_i devido perturbações nos parâmetros do modelo $\delta \mathbf{m}$, é o suficiente para obter a formulação adjunda (2.21). Os gradientes dos parâmetros módulo de bulk K e densidade ρ são obtidos a partir das equações (Tarantola, 1984):

$$\begin{aligned} \delta K(\mathbf{x}) &= \frac{1}{K(\mathbf{x})^2} \sum_s \int_t dt \frac{\partial p(\mathbf{x}, t; \mathbf{x}_s)}{\partial t} \frac{\partial p'(\mathbf{x}, t; \mathbf{x}_s)}{\partial t} \\ \delta \rho(\mathbf{x}) &= \frac{1}{\rho(\mathbf{x})^2} \sum_s \int_t dt \nabla p(\mathbf{x}, t; \mathbf{x}_s) . \nabla p'(\mathbf{x}, t; \mathbf{x}_s) \end{aligned} \tag{2.22}$$

onde $p(\mathbf{x}, t; \mathbf{x}_s)$ e e $p'(\mathbf{x}, t; \mathbf{x}_s)$ representam o campo de onda de pressão propagado para frente e o campo de onda residual retropagado, respectivamente, sendo que $p(\mathbf{x}, t; \mathbf{x}_s)$ é gerado por uma fonte posicionada em $\mathbf{x}_s$.

O campo de onda $p(\mathbf{x}, t; \mathbf{x}_s)$ é definido por:

$$p(\mathbf{x}, t; \mathbf{x}_s) = \int_V dV G(\mathbf{x}, t; \mathbf{x}_s, 0) * s(\mathbf{x}, t; \mathbf{x}_s), \tag{2.23}$$

em que $G(\mathbf{x}, t; \mathbf{x}_s, 0)$ é a função de Green acústica convoluída com a equação da onda (2.18). O símbolo (*) indica convolução no tempo.

O campo de onda residual retropropagado $p'(\mathbf{x}, t; \mathbf{x}_s)$ é calculado através dos dados residuais retropropagados no tempo, como segue:

$$p'(\mathbf{x}, t; \mathbf{x}_s) = \sum_r G(\mathbf{x}, -t; \mathbf{x}_r, 0) * \delta d_i(\mathbf{x}_r, \mathbf{x}_s, t). \tag{2.24}$$

Os resíduos são retropropagados de todos os pontos receptores referentes à fonte na posição $\mathbf{x}_s$. O campo de onda residual é definido como o campo difratado ausente no modelo atual e seu cálculo requer uma única modelagem direta, sendo o termo fonte substituído pelos dados residuais. Considerando uma única fonte, o cálculo do gradiente do módulo de bulk (2.22) é realizado pela correlação cruzada de defasagem zero da derivada temporal do campo de onda p de modelagem direta e da derivada temporal do campo de onda residual retropagado p'. O cálculo do gradiente da densidade é semelhante, mas as derivadas tenporais dos campos de onda diretos e residuais são substituídas por derivadas espaciais. O gradiente final é resultado da soma dos gradientes produzidos pelos tiros individuais.

2.3 ALGORITMO DE INVERSÃO DE FORMA DE ONDA COMPLETA

Nesta seção são apresentados as principais etapas do Algoritmo de inversão utilizado neste trabalho (Kurzmann, 2012).

2.3.1 Modelagem direta

A equação de onda acústica de primeira ordem com uma densidade variável é expressa como:

$$\begin{aligned} \frac{\partial p}{\partial t} &= K\nabla \mathbf{v} \\ \frac{\partial \mathbf{v}}{\partial t} &= \frac{1}{\rho}\nabla p, \end{aligned} \tag{2.25}$$

onde $\mathbf{v}$ é a velocidade da partícula, p é o campo de pressão, K é o módulo de bulk e ρ é a densidade. Para a equação de onda acústica 2D, obtemos:

$$\begin{cases} \frac{\partial p(x,z,t)}{\partial t} = K(x,z)\left(\frac{\partial v_x(x,z,t)}{\partial x} + \frac{\partial v_z(x,z,t)}{\partial z}\right) \\ \frac{\partial v_x(x,z,t)}{\partial t} = \frac{1}{\rho(x,z)}\frac{\partial p(x,z,t)}{\partial x} \\ \frac{\partial v_z(x,z,t)}{\partial t} = \frac{1}{\rho(x,z)}\frac{\partial p(x,z,t)}{\partial z} \end{cases} \tag{2.26}$$

A fim de resolver o problema direto, a equação (2.26) é discretizada nos domínios do espaço e tempo através da aproximação de segunda ordem de diferenças finitas, de modo que:

$$\begin{aligned} x &= i\Delta h \qquad & i &= (1, 2, ..., N_x), \\ z &= j\Delta h \qquad & j &= (1, 2, ..., N_z), \\ t &= n\Delta t \qquad & k &= (1, 2, ..., N_t), \end{aligned} \tag{2.27}$$

onde N_x, N_z indicam o número de pontos da malha nas direções x e z, e N_t é o número total de etapas de tempo.

A ténica de escalonamento (Virieux, 1986) é utlizada para calcular as derivadas espaciais das variáveis do campo de onda na posição correta da malha. Dessa forma, a pressão p é definida nos pontos da malha de referência (j,i), enquanto que a velocidade da partícula $\mathbf{v}$ é definido nas posições de meio ponto fora da malha de referência. Além disso, as derivadas temporais da pressão são calculadas no instante n, e a velocidade da partícula é calculada em meia etapa de tempo.

Discretizar a equação da onda significa substituir as derivadas parciais por operadores de diferenças finitas. A equação da onda acústica 2D é discretizada através do sistema:

$$\begin{cases} \frac{\partial v_x(x,z,t)}{\partial t} \approx \frac{v^{n+\frac{1}{2}}_{x|j,i+\frac{1}{2}} - v^{n-\frac{1}{2}}_{x|j,i+\frac{1}{2}}}{\Delta t} = \frac{1}{\rho_{j,i+\frac{1}{2}}} \frac{p^n_{j,i+1} - p^n_{j,i}}{\Delta h} \\ \frac{\partial v_z(x,z,t)}{\partial t} \approx \frac{v^{n+\frac{1}{2}}_{z|j+\frac{1}{2},i} - v^{n-\frac{1}{2}}_{z|j+\frac{1}{2},i}}{\Delta t} = \frac{1}{\rho_{j+\frac{1}{2},i}} \frac{p^n_{j+1,i} - p^n_{j,i}}{\Delta h} \\ \frac{\partial p(x,z,t)}{\partial t} \approx \frac{p^{n+1}_{j,i} - p^n_{j,i}}{\Delta t} = k_{j,i} \frac{v^{n+\frac{1}{2}}_{x|j,i+\frac{1}{2}} - v^{n+\frac{1}{2}}_{x|j,i-\frac{1}{2}} + v^{n+\frac{1}{2}}_{z|j+\frac{1}{2},i} - v^{n+\frac{1}{2}}_{z|j-\frac{1}{2},i}}{\Delta h} \end{cases} \tag{2.28}$$

2.3.1.1 Condições iniciais e de contorno

As condições iniciais e de contorno devem ser definidas para a resolução da equação da onda. Considerando um meio em repouso, antes da excitação, a condição inicial correspondente é

$$p(x, z, t = 0) = \partial_t p(x, z, t = 0) = 0 \tag{2.29}$$

Existem dois tipos de condições de contorno, a condição de superfície livre e o limite absorvente.

- Condição de superfície livre: A modelagem dessa condição utiliza o formalismo a vácuo (Bohlen and Saenger, 2006; Zahradník et al., 1993), em que o meio acima da superfície livre é tratado como vácuo através da atribuição de valores zero aos parâmetros acústicos e próximos de zero à densidade para evitar divisões indefinidas. A utilização do formalismo a vácuo é ideal em situações de superfície livre planar. Essa condição também pode ser implementada através da técnica de espelhamento (Levander, 1988), mas seu uso não é tão popular devido variações na implementação da técnica nos códigos de de modelagem direta (acústica 2D, acústica 3D, elástica 2D).

- Limite absorvente: Condição aplicada para atenuar as reflexões indesejadas nas bordas do modelo (Berenger, 1994). O código de modelagem acústica por diferenças finitas também utiliza como condição de contorno camadas perfeitamente combinadas (PML), o que requer equações adicionais para atualizar o campo de onda de pressão (Kurzmann, 2012).

2.3.1.2 Precisão e estabilidade

Condições de amostragem espacial e temporal devem ser satisfeitas para evitar artefatos e instabilidades numéricas. A dispersão da malha corre devido o truncamento da série de Taylor quando se aproxima de derivadas espaciais. A condição utilizada para não ocorrer a dispersão relaciona o número de pontos da malha n com o comprimento de onda mínimo λ_{min} através da equação

$$\Delta h \leq \frac{\lambda_{min}}{n} = \frac{v_{pmin}}{n f_{max}}, \tag{2.30}$$

onde f_{max} é a frequência máxima do campo de onda e v_{pmin} é a velocidade mínima da onda P. A escolha do comprimento e do tipo de operador de DF determina o número mínimo de pontos da malha por comprimento de onda (Köhn, 2011).

O critério de amostragem (Courant et al., 1928) para a modelagem 2D é definida através da razão entre os intervalos de amostragem temporal e espacial

$$\Delta t \leq \frac{\Delta h}{\sqrt{2} v_{pmax}} \tag{2.31}$$

onde v_{pmax} é a velocidade máxima da onda P no modelo. Esse critério determina que o intervalo de tempo Δt deve ser menor que o tempo de propagação entre dois pontos vizinhos da malha.

2.3.2 Inversão

O processo de inversão é baseado na minimização iterativa da função objeto para a atualização do modelo. Para isso, é necessária a informação da matriz Hessiana. Existem diversos métodos de otimização para obter a matriz Hessiana, o método utilizado nesta dissertação foi o quasi-Newto L-BFGS. No Anexo B são descritos os métodos de otimização: Newton, quasi-Newton e gradiente conjugado.

2.3.2.1 Método L-BFGS

Nessa seção são apresentadas as principais equações do método quasi-Newton L-BFGS, baseado em Koehene (2018). O estudo aprofundado sobre o método pode ser encontrado em Nocedal and Wright (2006).

Considerando $\mathbf{B}_k$ uma aproximação da Hessiana, as abordagens quasi-Newton utilizam as informações de gradiente das iterações anteriores, uma vez que as mudanças apresentadas no gradiente ao longo das iterações carregam informação sobre as derivadas de segunda ordem de E($\mathbf{m}$)(Nocedal and Wright, 2006). Dessa forma, a atualização de $\mathbf{B}_k$ na iteração k utiliza as informações anteriores de gradiente e modelos para aproximar a matriz Hessiana.

O método BFGS define a atualização da matriz $\mathbf{B}_{k+1}^{-1}$(aproximação da Hessiana inversa), na k-ésima iteração, como:

$$\mathbf{B}_{k+1}^{-1} = \mathbf{B}_k^{-1} - \frac{\mathbf{B}_k^{-1}\mathbf{y}_k\mathbf{y}_k^T\mathbf{B}_k^{-1}}{\mathbf{y}_k^T\mathbf{B}_k^{-1}\mathbf{y}_k^T} + \frac{\mathbf{s}_k\mathbf{s}_k^T}{\mathbf{y}_k^T\mathbf{s}_k} \tag{2.32}$$

Entretando, a matriz $\mathbf{B}$ apresentará N^2 elementos, sendo N o número de parâmetros do modelo. Este valor excede a capacidade de armazenamento de dados dos computadores atuais. Para resolver esse problema, uma versão de memória limitada do BFGS foi criada para que pudesse ser aplicada a problemas de larga escala. Essa versão ficou conhecida como L-BFGS e nela não é preciso ter a matriz **B** explicitamente armazenada, apenas um número m de iterações passadas, ou seja, os modelos e gradientes. Portanto, o custo de armazenamento é reduzido de N^2 para $2 \times m \times N$. Normalmente m é um número menor que 30 (é comum utilizar as 10 iterações anteriores, $m = 10$). Antes de iniciar o método L-BFGS, é necessário fazer uma atualização com o método de descida mais íngrime para obter um primeiro conjunto de diferenças de modelo e gradiente necessárias para iniciar o L-BFGS.

Para a k-ésima iteração,

$$\begin{aligned} \tilde{\mathbf{x}}_k &= \mathbf{m}_{k+1} - \mathbf{m}_k \\ \tilde{\mathbf{y}}_k &= \mathbf{g}_{k+1} - \mathbf{g}_k, \end{aligned} \tag{2.33}$$

onde $\mathbf{m}$ é o modelo de parâmetros e $\mathbf{g}$ o gradiente, e $\tilde{\mathbf{x}}_k$ e $\tilde{\mathbf{y}}_k$ são cubos contendo m(número de iterações anteriores) diferenças entre modelos e gradientes consecutivos, respectivamente. No algoritmo para o L-BFGS (Nocedal and Wright, 2006) os loops sempre começam e terminam em $k1$, assim as expressões (2.33) sempre tratam da informação na iteração atual k e da iteração anterior $k1$. A atualização do modelo é dada pela expressão (Koehene, 2018):

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \mathbf{v}_k, \tag{2.34}$$

onde α_k é o comprimento do passo e $\mathbf{v}_k$ é a direção de busca. O método L-BFGS pode ser abordado como passo unitário ou passo variável, nesta dissertação foi utilizada a abordagem de passo variável(Nocedal and Wright, 2006).

2.3.2.2 Algoritmo do FWI

Nesta subseção é apresentado o Algoritmo generalizado do método FWI (Warner, 2012):

1. Obtenção do campo de onda, modelo inicial e fonte;

2. Modelagem direta → predizer o campo de ondas:
   Para cada fonte presente no conjunto de dados é resolvida a equação da onda para gerar o campo de onda calculado, isto é o campo de onda direto;

3. Calcular o campo de ondas residual na posição dos receptores:
   Para cada receptor presente no conjunto de dados, é calculada a diferença entre o dado observado e o dado calculado. Essa diferença fornece os dados residuais. O objetivo da inversão é reduzir esses resíduos modificando o modelo de velocidade e potencializando também a wavelet fonte. A norma L-2 do dado residual fornece a funçaõ objeto que é iterativamente minimizada;

4. Retropropagar o campos de ondas residual:
   Para cada fonte original, o dado residual é retropagado a partir do receptor para a fonte. É gerado o campo de onda residual retropagado;

5. Correlação cruzada → atualização do modelo não escalonado:
   Para cada fonte de origem e em cada ponto no modelo, é feita a correlação cruzada do campo de onda direto e retropagado. O resultado gera uma atualização do modelo de velocidade para todos os pontos no modelo. Essas atualizações são somadas em todas as fontes de origem. Nesse estágio, a atualização do modelo de velocidade é não-dimensionado, isso significa que sabemos a direção de atualização, mas não sabemos o quanto. A atualização do modelo não-dimensionado fornece a direção do gradiente necessário para minimizar a função objeto;

6. Cálculo do passo → atualizar o modelo escalonado:
   Para a escala do gradiente, o modelo d velocidade é perturbado em pequena quantidade na direção do gradiente, o campo de onda direto para o novo modelo é calculado, assim como um novo grupo de dados residuais e um novo valor de função objeto são calculados. A partir da comparação entre esse novo valor com o original, um escalar idela pode ser encontrado para minimizar a função objeto, esse escalar ideal é o comprimento do passo;

7. Atualizar o modelo e voltar para a etapa 2 até atingir o critério de convergência:
   Usando o gradiente e o comprimento do passo, o modelo de velocidade original é atualizado até atingir o critério de convergência definido.

2.3.2.3 Paralelização

Os custos computacionais do FWI estão associados à três fatores: capacidade da modelagem direta, tamanho dos dados e tamanho do domínio computacional. A modelagem direta e o tamanho dos dados são mutuamente definidos pelo número de tiros, enquanto que o tamanho do domínio computacional é definido pelos valores máximos de deslocamento dos dados, profundidade da imagem e intervalo de frequência. Para reduzir esses custos, a computação paralela é utilizada (Pica et al., 1990).

Existem dois níveis de paralelização utilizados no código FWI acústico que são baseados no príncipio da decomposição do domínio (Bohlen, 2002) e tiros paralelos (Bohlen et al., 2009). A estratégia de decomposição do domínio divide o modelo em subdomínios que são distribuídos para vários processadores, o que exige uma comunicação ponto-a-ponto para trocar valores de campo de onda entre subdomínios vizinhos. A troca de informações entre subdomínios é realizada através do *Message Passing Interface* (MPI). O segundo nível de Paralelização, tiros paralelos, não exige comunicação pois os tiros são distribuídos para grupos de processadores. Através da combinação dos dois tipos de paralelização o algoritmo de inversão apresenta uma aceleração significativa.

2.4 PRÉ-REQUISITOS PARA EFICIÊNCIA DO FWI

A inversão da forma da onda completa é um problema inverso altamente não linear e uma das consequências disso é a presença de vários mínimos locais na forma da função objeto. Tal característica torna o sucesso do método FWI altamente dependente da escolha do modelo inicial, pois a escolha adequada deve garantir a convergência do algoritmo para o mínimo global da função objeto. Dessa forma, um modelo inicial adequado é encontrando quando os dados calculados pelo modelo inicial correspondem aos dados observados em meio ciclo da freqüência mínima considerada (Sun and McMechan, 1992). Se essa exigência não for satisfeita, ocorre o salto de ciclo, em que a inversão tenta ajustar os dados calculados ao ciclo errado dos dados observados como mostrado na Figura 2.4, assim o algoritmo converge para um mínimo local da função objeto.

Segundo Pratt (2008), o critério de convergência também pode ser expresso em função de erros de tempo de trânsito e distâncias de propagação medidas em comprimentos de onda, em que o erro de tempo de trânsito δt para um dado evento sísmico tem que ser menor do que meio comprimento de onda, ou seja:

$$\frac{\delta t}{T} < \frac{\lambda}{2cT} \qquad \frac{\delta t}{T} < \frac{1}{2k_z} \tag{2.35}$$

onde T é o tempo total de chegada do evento, k_z é o número de ondas e c é a velocidade de referência. Essa exigência pode ser facilitada utilizando frequências mais baixas ou reduzindo o afastamento, o que significa diminuir o número de ondas k_z entre a fonte e o receptor. Dessa forma, existem três maneiras de reduzir o erro de tempo de trânito e melhorar a convergência do FWI:

- melhorar a precisão do modelo inicial,
- reduzir os afastamento,
- reduzir as frequências iniciais.

Para testes sintéticos, é muito tentador reduzir δt (melhorando artificialmente a precisão do modelo inicial) ou reduzir k_z (reduzindo os afastamento ou reduzindo as frequências iniciais). Qualquer uma dessas abordagens pode levar a resultados artificialmente otimistas, e deve sempre ser lembrado que parte do valor dos testes sintéticos é gerar resultados que ilustrem não apenas o potencial benefícios, mas potenciais armadilhas (Pratt, 2008).

Segundo Virieux and Operto (2009) aplicar a inversão multiescala também é uma alternativa para reduzir a forte não-linearidade do problema inverso. Isso significa iniciar a inversão com baixas frequências e incrementar as frequências mais altas gradualmente durante o processo de inversão. Não existe uma regra geral para aplicar a inversão multiescala, a abordagem é adaptada para cada problema inverso particular.

Na ausência de frequências baixas no dado, o modelo inicial precisa ser mais preciso e conter estruturas de comprimento de onda longo e por isso existem diferentes métodos para obter um modelo inicial, como por exemplo: "First-arrival traveltime tomography"(FATT) (Bleibinhaus et al., 2009); "Reflection traveltime tomography"(RTT)(Prieux et al., 2013); "Stereotomography"(Prieux et al., 2013); "Laplace domain inversion"(Shin and Ho Cha, 2009); "FO-CRS tomography"(Mesquita et al., 2019).

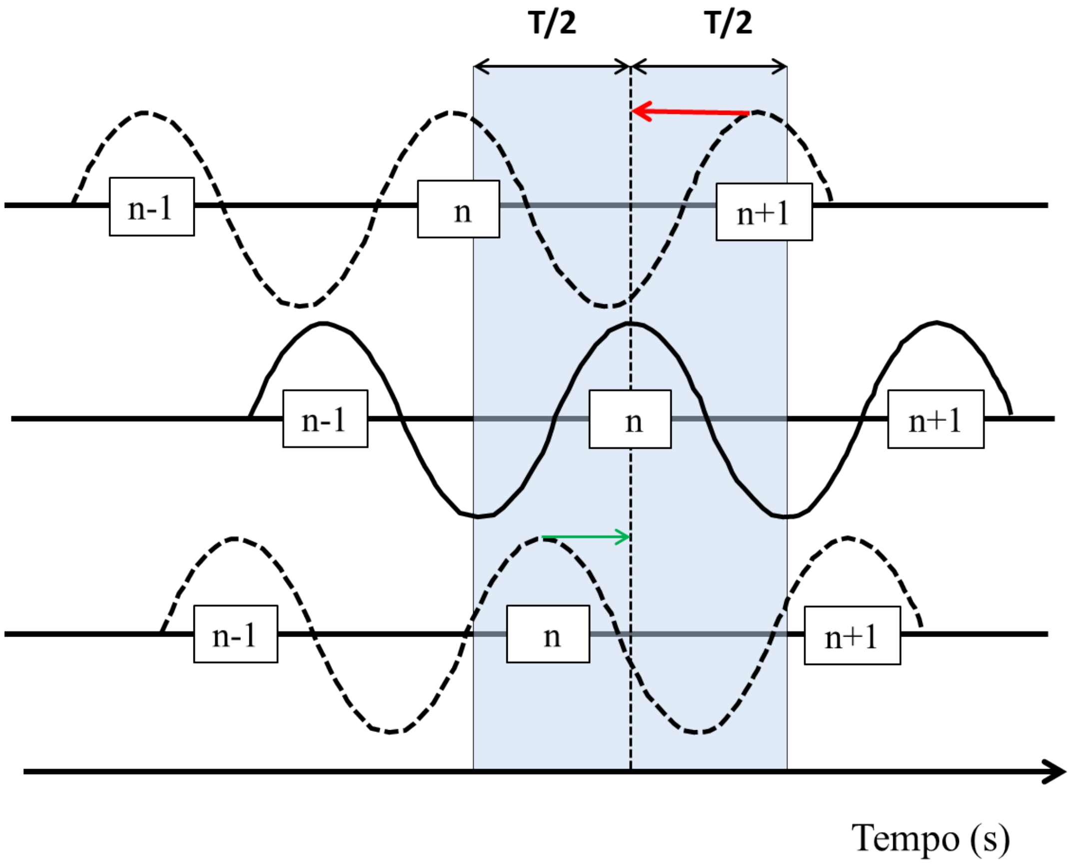


Figura 2.4: O problema do salto de ciclo. A linha sólida representa um sismograma monocromático. As linhas tracejadas representam os sismogramas calculados. O sismograma calculado no topo tem um atraso de tempo maior que $T/2$ e o FWI tentará ajustar os dados calculados ao ciclo errado do sismograma observado. modificado de Virieux and Operto (2009).

3 APLICAÇÃO: ESTUDO DE PARÂMETROS

Este capítulo apresenta um estudo sobre o FWI utilizando um modelo da subsuperfície sintético para investigar o efeito da escolha do modelo inicial na convergência do método e o efeito da parametrização do problema na acurácia e precisão do resultado da inversão, de tal forma que foram avaliados os efeitos da filtragem de frequências e janelamento de afastamentos no dado observado. Também foi apresentado o estudo da estabilidade do método através de dados contaminados por ruído aletório. Para evitar efeitos adicionais aos resultados da inversão, o estudo foi desenvolvido com a função fonte “wavelet”conhecida e a inversão limitada ao parâmetro v_p, ou seja, a densidade é negligenciada.

3.1 CONFIGURAÇÃO BÁSICA DO EXPERIMENTO

O modelo utilizado é acústico de três camadas com 1500 m de largura e 1500 m de profundidade com variação apenas na direção vertical. A camada de água tem uma espessura de 500 m. A velocidade média da onda P é 1826 m/s, com a velocidade mínima de 1480 m/s na camada de água e a velocidade máxima de 2500 m/s. A modelagem direta é realizada pelo método de diferenças finitas acústicas 2D, em que foram simulados 5 tiros com 6 segundos de dados. Os modelos de velocidade da onda P e densidade são mostrados nas Figuras 3.1 e 3.2, respectivamente.

O objetivo é simular as conições de medição da sísmica de reflexão marinha com a fonte de pressão posicionada em 20 m abaixo da interface ar-água e espaçadas em 50 m. A “wavelet”fonte é do tipo Ricker com uma frequência central de 4 Hz e foi filtrada com um filtro passa-baixa “Butterworth”com uma frequência de corte de 3 Hz Figura. Em 50 m de profundidade estão posicionados 15 hidrofones. A configuração fonte-hidrofone apresenta afastamento máximo de 750 m e afastamento mínimo de 50 m. A Tabela 3.1 mostra de modo geral a configuração básica utilizada no experimento, incluindo a geometria de aquisição e os parâmetros gerais para a modelagem e inversão por diferenças finitas. As famílias de tiro comum são mostrados na Figura 3.3.

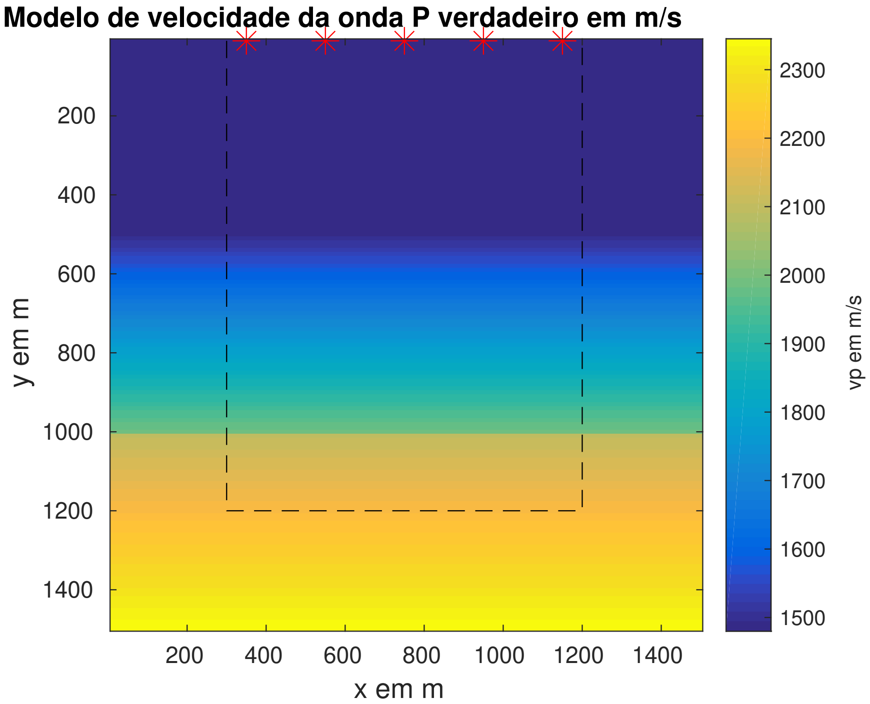


Figura 3.1: Modelo v_p verdadeiro

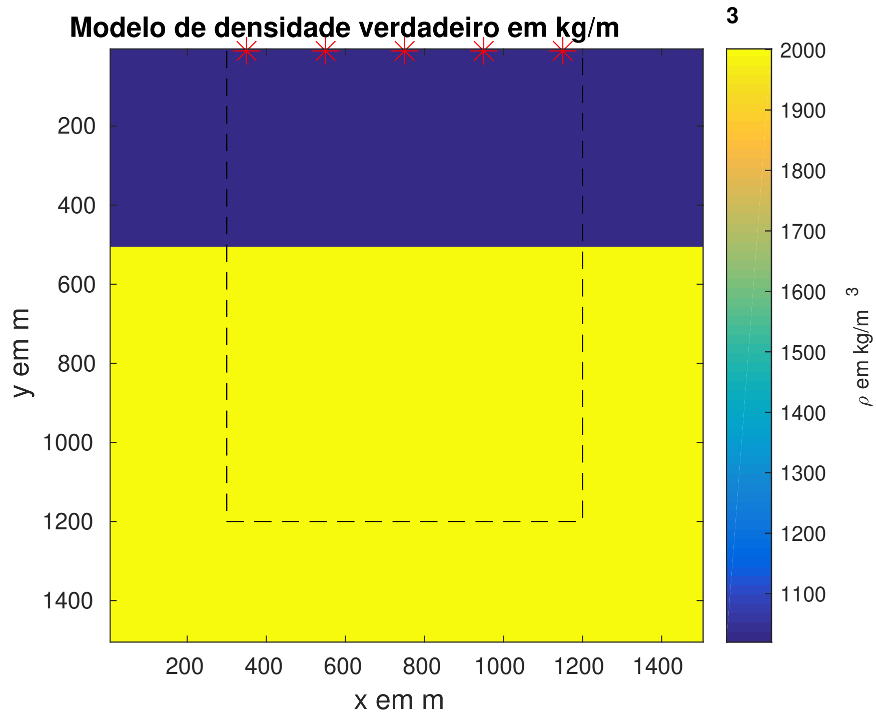


Figura 3.2: Modelo de densidade

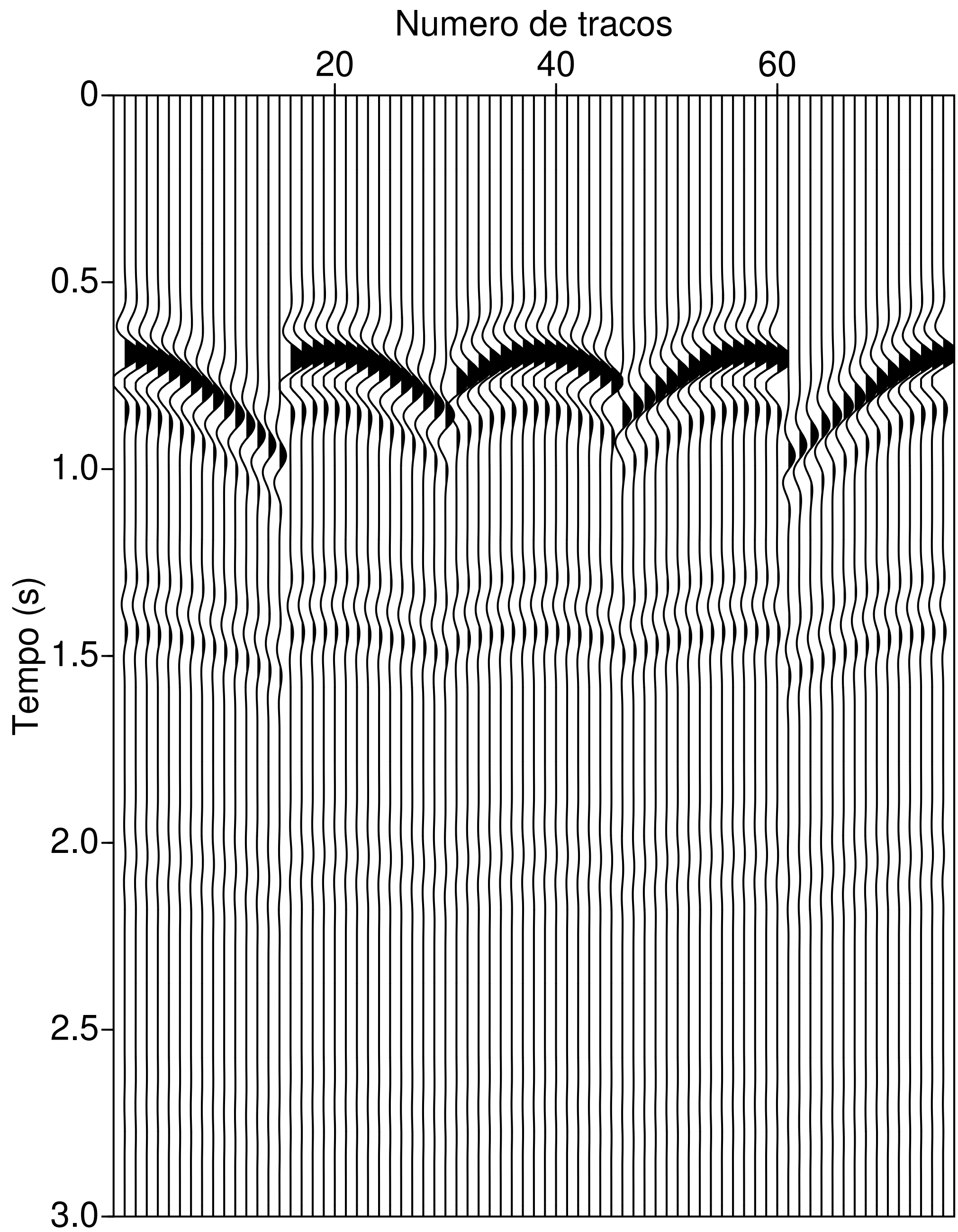


Figura 3.3: Dado sísmico sintético na configuração fonte comum, usado como dados observados no programa IFOS2D do método FWI.

Tabela 3.1: Parâmetros de modelagem e aquisição do conjunto de dados sintéticos.

Modelo sintético	
Tamanho do modelo	1500 m x 1500 m
v_p Média	1826 m/s
v_p Mínimo	1480 m/s
v_p Máximo	2500 m/s
Espessura da camada de água	500 m
Parametros de aquisição	
Número de tiros	5
Espaçamento dos tiros	200 m
profundidade dos tiros	20 m
Número de hidrofones	15
Espaçamento dos hidrofones	50 m
Profundidade dos hidrofones	50 m
Afastamento mínimo	50 m
Afastamento máximo	750 m
Parametros da modelagem	
Tamanho da malha	150 m x 150 m
Espaçamento da malha	10 m
Tempo de amostragem	1e-3 s
Número de amostras de tempo	1111
Comprimento de gravação	6 s
“wavelet”fonte	Ricker 4 Hz

3.2 ESTUDO DO EFEITO DA APROXIMAÇÃO INICIAL NA CONVERGÊNCIA DO MÉTODO FWI

3.2.1 Seleção da aproximação inicial

A fim de avaliar a convergência do método de inversão, foram feitos quatro exprimentos com diferentes modelos iniciais. Os modelos iniciais utilizados nas simulações foram obtidos pela perturbação de 25%(B), 35%(C) e 50%(D) das velocidades da segunda e terceira camada de um modelo inicial de referência A. O primeiro teste de inversão utilizou o modelo A, enquanto que o segundo, terceiro e quarto teste utilizaram os modelos B, C e D, respectivamente. Nenhum tipo de filtragem de dados foi aplicada e o modelo de densidade utilizado se manteve constante.

As Tabelas 3.2 e 3.3 mostram as velocidades de topo v_{topo} e base v_{base} e a Tabela 3.4 mostra as profundidades de topo e base z_{topo} e z_{base} utilizados no calculo do gradiente (Eq. 3.2) para a construção dos modelos A, B, C e D utilizados nos testes de inversão. A construção do modelo de velocidade é dado pela expressão:

$$v = v_0 + g_z z, \tag{3.1}$$

Onde z é a profundidade e g_z é o gradiente calculado por:

$$g = \frac{v_{topo} - v_{base}}{z_{topo} - z_{base}}. \tag{3.2}$$

A Figura 3.4 apresenta o resultado do primeiro teste de inversão, em que utilizamos o modelo A como modelo de velocidade inicial. Os perfis de velocidade vertical em x=500 m e x=750 m conseguem se ajustar perfeitamente à curva da velocidade inicial (preto tracejado), confirmando a convergência do algoritmo, mas não se aproximou da curva de velocidade verdadeira (cinza), o que pode indicar que o gradiente está preso em um mínimo local que não corresponde ao mínimo global.

As Figuras 3.5, 3.6 e 3.7 apresentam os resultados de inversão para os modelos B, C e D, respectivamente. Nos três casos observamos, através dos perfis de velocidade vertical em x=500 m e x=750 m, que o algoritmo converge para um mínimo. Assim como no primeiro teste, não conseguiram se aproximar do modelo real.

A Figura 3.8 (a) mostra a convergência das quatro curvas da função objeto obtidas pelas simulações do FWI com os respectivos modelos A, B, C e D. A Figura 3.8 (b) mostra os modelos de velocidade inicial correspondentes às curvas da função objeto. Notou-se que a curva da função objeto referente ao modelo de velocidade inicial (azul) apresenta a convergência mais rápida. Enquanto que, as curvas referentes aos modelos de velocidades com aumento progressivo de perturbações na segunda e terceira camada, se afastaram proporcionalmente do caminho de convergência.

A partir da Figura 3.8 (a), nota-se que o algoritmo consegue alcançar o objetivo de minimizar a função objeto, entretanto, não obteve um modelo invertido próximo do modelo verdadeiro (cinza), o que reforça a presença do problema de forte não-linearidade na inversão. Na próxima seção será apresentada a estratégia da inversão multiescala através da filtragem de frequências do dado observado com o propósito de diminuir a não-linearidade na inversão.

Tabela 3.2: Velocidades (Km/s) utilizadas no cálculo dos modelos A e B. Cada camada tem duas velocidades utilizadas na Equação (3.2), a primeira é a velocidade do topo v_{topo} e a segunda é a velocidade da base v_{base}.

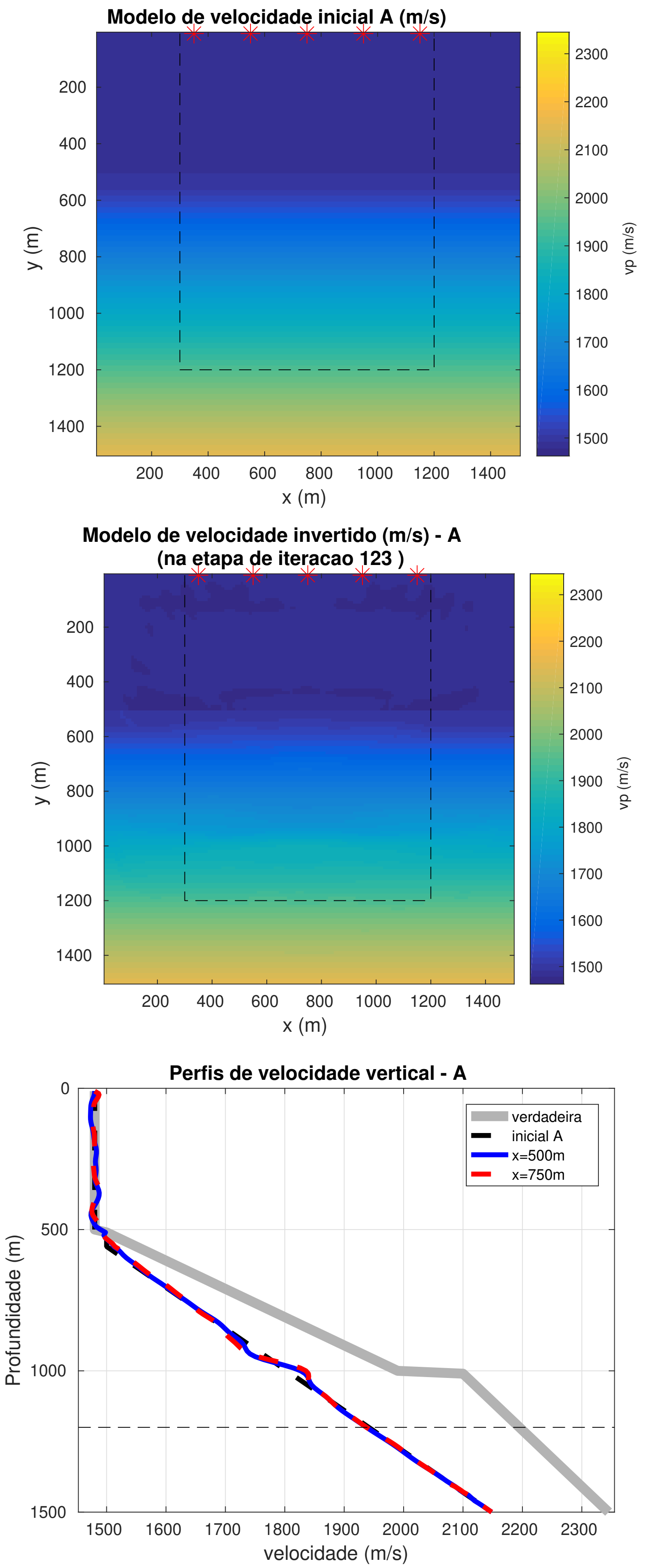
	Modelo A	Modelo B
Camada 1	1400 - 1400	1400 - 1400
Camada 2	1480 - 1500	1480 - 1500
Camada 3	1500 - 1500	1500 - 1845
Camada 4	1500 - 2500	1845 - 3125

Tabela 3.3: Velocidades (Km/s) utilizadas no cálculo dos modelos C e D. Cada camada tem duas velocidades utilizadas na Equação (3.2), a primeira é a velocidade do topo v_{topo} e a segunda é a velocidade da base v_{base}.

	Modelo C	Modelo D
Camada 1	1400 - 1400	1400 - 1400
Camada 2	1480 - 1500	1480 - 1500
Camada 3	1500 - 1950	1500 - 2175
Camada 4	1950 - 3375	2175 - 3750

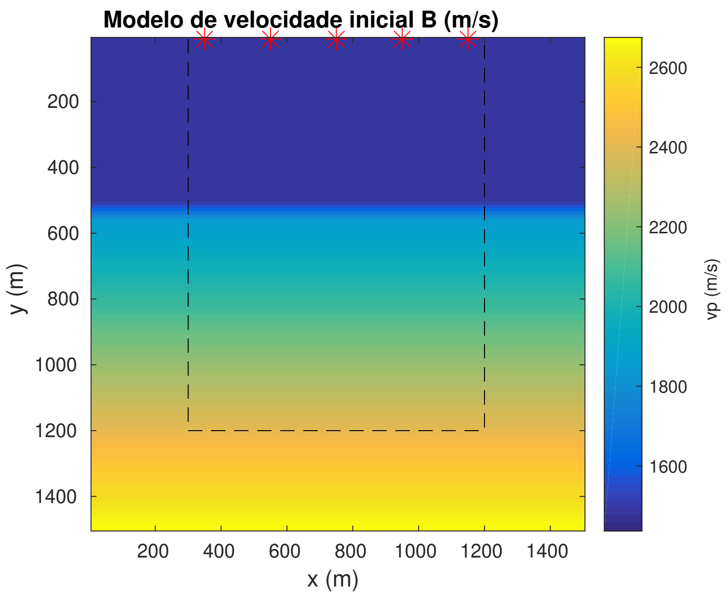
Tabela 3.4: Profundidades (m) utilizadas na construção dos modelos A, B, C e D. Cada camada é definida pelo valor de duas profundidades z utilizadas na Equação (3.2), a primeira é a profundidade do topo z_{topo} e a segunda é a profundidade da base z_{base}.

	Profundidade (m)
Camada 1	0 - 500
Camada 2	500 - 500
Camada 3	500 - 550
Camada 4	550 - 2000

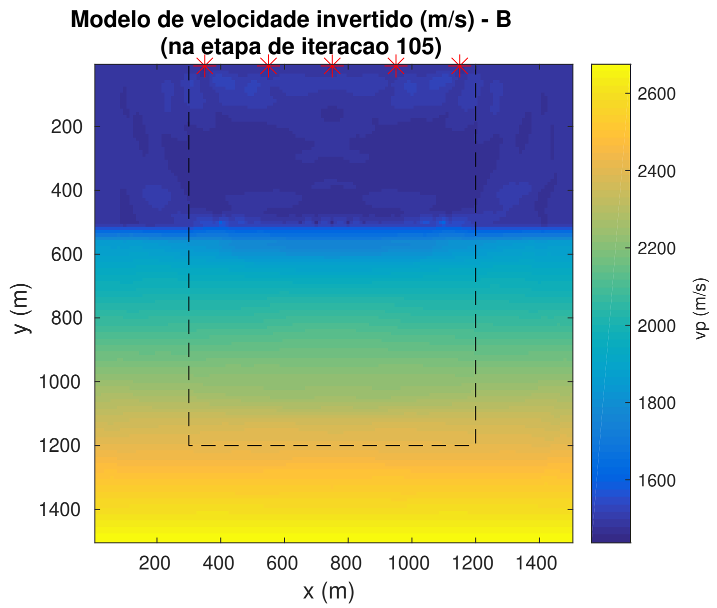


(a) (b) (c)

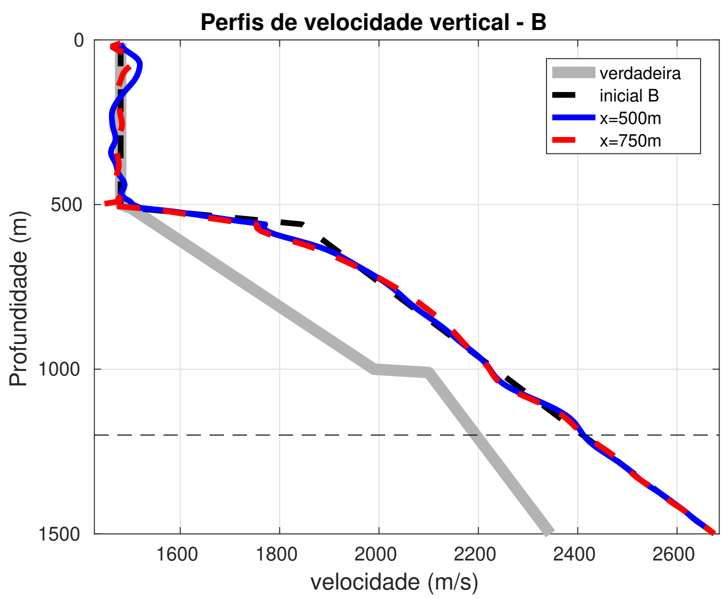
Figura 3.4: (a) Modelo de velocidade inicial A. (b) Modelo de velocidade obtido a partir do modelo A. (c) Perfis de velocidade. O retângulo tracejado representa a área de inversão dos parâmetros.



(a)

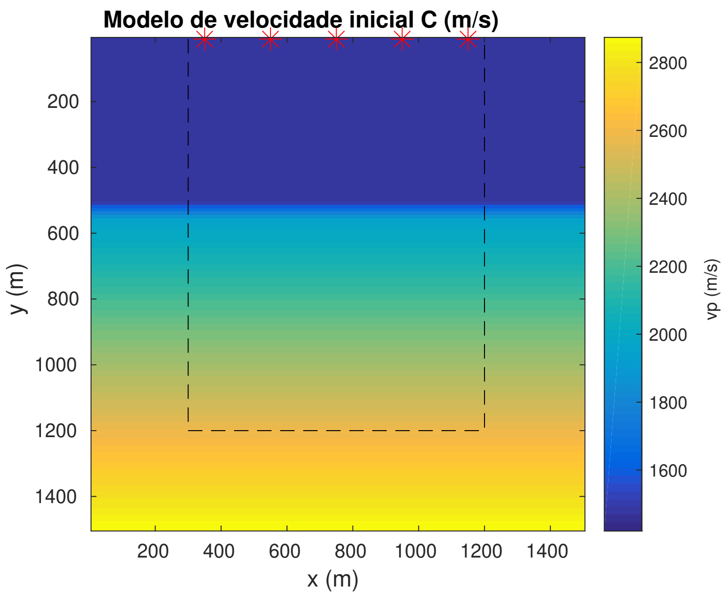


(b)

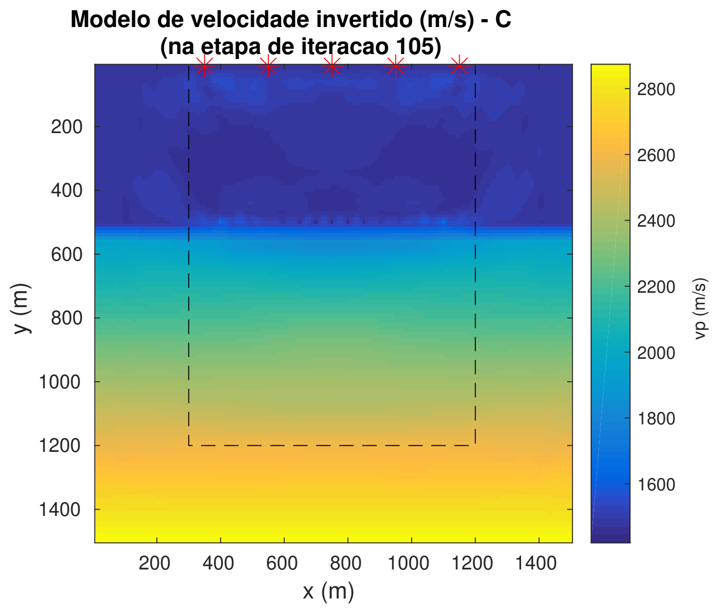


(c)

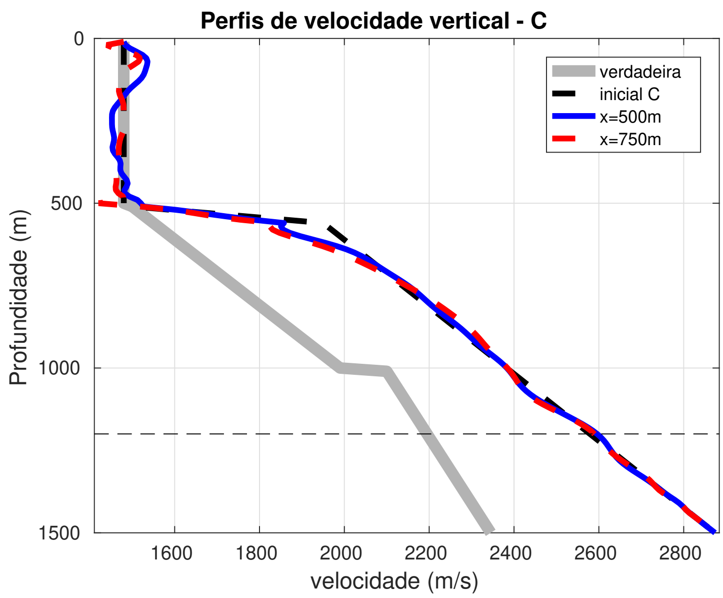
Figura 3.5: (a) Modelo de velocidade inicial B. (b) Modelo de velocidade obtido a partir do modelo B. (c) Perfis de velocidade. O retângulo tracejado representa a área de inversão dos parâmetros.



(a)

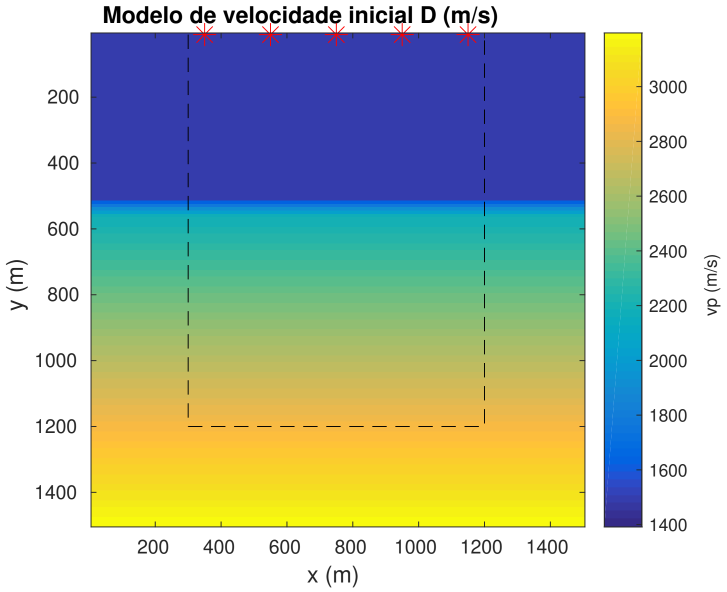


(b)

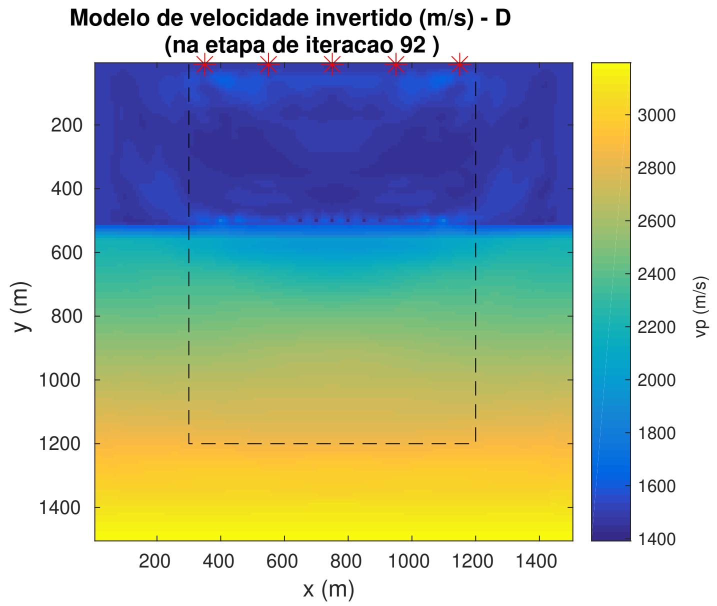


(c)

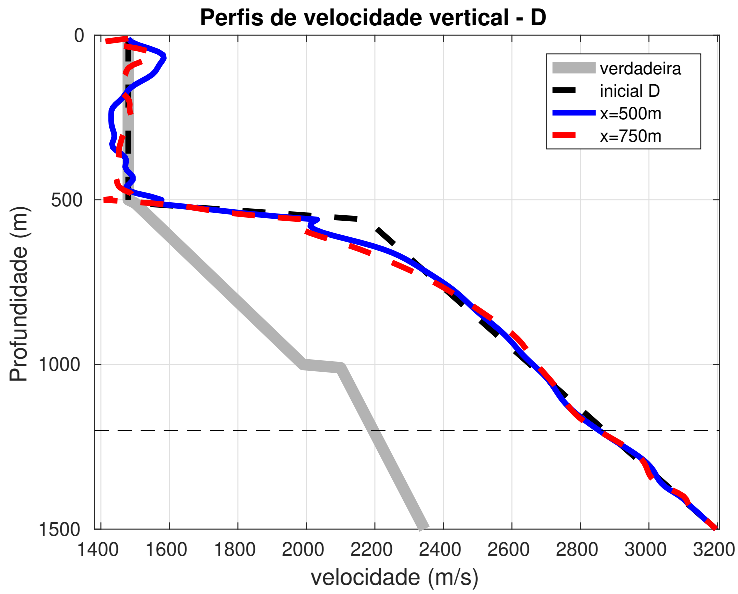
Figura 3.6: (a) Modelo de velocidade inicial C. (b) Modelo de velocidade obtido a partir do modelo C. (c) Perfis de velocidade. O retângulo tracejado representa a área de inversão dos parâmetros.



(a)

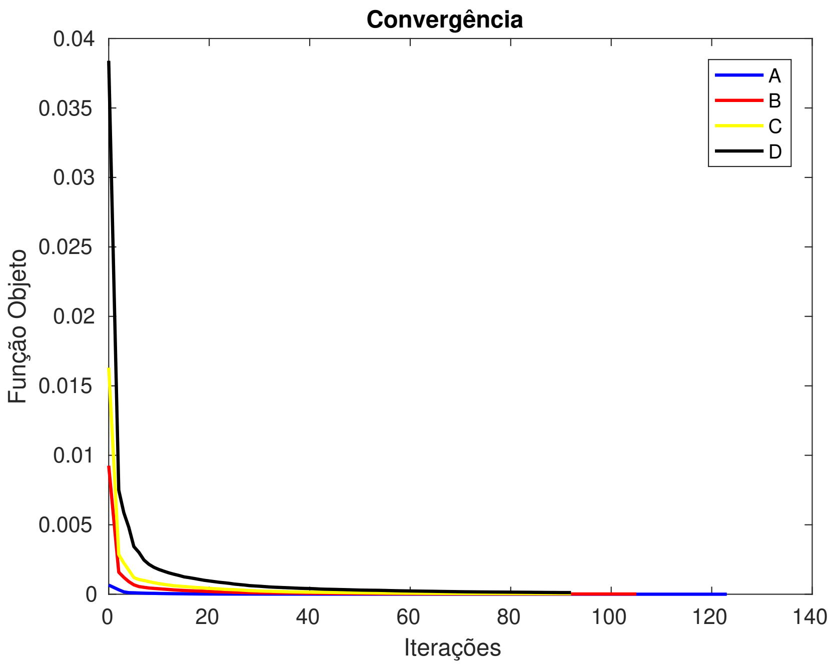


(b)

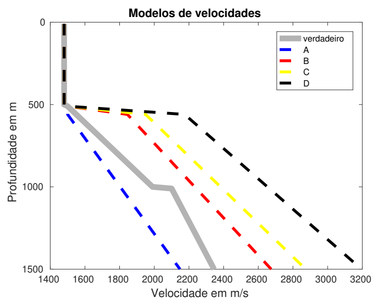


(c)

Figura 3.7: (a) Modelo de velocidade inicial D. (b) Modelo de velocidade obtido a partir do modelo D. (c) Perfis de velocidade. O retângulo tracejado representa a área de inversão dos parâmetros.



(a)



(b)

Figura 3.8: (a) Gráfico da convergência da função objeto para diferentes modelos de velocidades de partida em função do número de iterações. (b) Modelos de velocidades de partida utilizados em cada simulação do FWI.

3.3 INVERSÃO MULTIESCALA

Segundo Beydoun and Tarantola (1988) a aproximação de Born, ao calcular as derivadas parciais do campo de onda, exige que os dados calculados a partir do modelo inicial coincidam com os dados observados dentro de meio ciclo da frequência dominante. Essa exigência evita que o salto de ciclo ocorra e assim, a inversão não converge para um mínimo local. É mais fácil atingir essa condição quando a frequência dominante nos dados é baixa, pois a função objeto em baixas frequências é mais linear com relação às perturbações do modelo (Przebindowska, 2013).

Para reduzir a não-linearidade do problema inverso, ou seja, a existência de mínimos locais, a abordagem de inversão multiescala é aplicada (Bunks et al., 1995; Sirgue and Pratt, 2004). Assim, a inversão é executada com o acréscimo gradual do conteúdo de frequência, ou seja, o processo inicia em baixas frequências e os valores altos são adicionados consecutivamente. No domínio do tempo, a estratégia é aplicada através da filtragem de dados utilizando filtro passa-baixa (Bunks et al., 1995) ou passa-banda (Boonyasiriwat et al., 2009), enquanto que no domínio da frequência é aplicada a abordagem de múltiplas frequências (Virieux and Operto, 2009).

É válido destacar que a abordagem multiescala não é um método de inversão diferente, trata-se de uma forma de organizar e selecionar os parâmetros do modelo e do dado observado. A escala utilizada para isso é a escala do número de onda k_z, de k_{zmin} a k_{zmax}, o que significa decompor a velocidade em v_{min} a v_{max} e/ou a frequência em f_{min} a f_{max}, visto que $k_z = \frac{f}{v}$.

Segundo Virieux and Operto (2009), o campo de onda responde a variações na subsuperfície cujo tamanho mínimo é aproximadamente metade do comprimento de onda $\lambda(2\pi/k_z)$. Ao iniciar a abordagem multiescala com valores pequenos de número de onda, é esperado que sejam delineadas as feições maiores do modelo, e progressivamente as feições menores, à medida que se aumente a escala de k_z. Para controlar essa escala durante a inversão, a frequência f do dado observado é delimitada.

Nesta seção, eu faço um estudo sobre a abordagem multiescala com foco na escolha da primeira banda de frequência, janelamento de afastamentos e no número de bandas de frequência utilizadas no processo de inversão. Foram aplicados diversos testes de inversão no modelo acústico da figura 3.1 com o objetivo de obter a melhor acurácia e precisão dos resultados a partir dos parâmetros escolhidos.

3.3.1 Seleção das bandas de frequência

A estratégia de seleção de banda de frequência utilizada é a proposta por Sirgue and Pratt (2004), em que para um dada frequência f_n e um intervalo de afastamento são definidos os valores $[k_{zmin}(fn)$,$k_{zmax}(fn)]$ de cobertura do número de onda vertical k_z para um caso 1D. O cálculo de $k_{zmin}(fn)$ e $k_{zmax}(fn)$ é dado por

$$
\begin{aligned}
k_{zmin}(f_n) &= 4\pi f_n \alpha_{min}/c_0, \\
k_{zmax}(f_n) &= 4\pi f_n/c_0,
\end{aligned} \tag{3.3}
$$

onde c_0 é a velocidade de referência homogênea. A expressão para α_{min} é dada por:

$$
\begin{aligned}
\alpha_{min} &= \frac{1}{\sqrt{1+R_{max}^2}}, \\
R_{max} &= \frac{h_{max}}{z}.
\end{aligned} \tag{3.4}
$$

sendo R_{max} a razão entre meio afastamento e profundidade, h_{max} é o máximo meio afastamento e z é a profundidade do alvo. A imagem do gradiente é formada pela contribuição de diferentes números de onda verticais referentes a diferentes pares fonte-receptor, assim o incremento do afastamento produz o número de onda mínimo, enquanto que o número de onda máximo é produzido pelo afastamento mais próximo. A seleção da frequência deve possibilitar uma amostragem contínua do espectro do número de onda tal que

$$
k_{zmin}(f(n+1)) = k_{zmax}(f_n). \tag{3.5}
$$

Seja f_{n+1} a próxima frequência selecionada, que é calculada utilizando as equações 3.3 e 3.5 através da relação:

$$
f_{n+1} = \frac{f_n}{\alpha_{min}} \tag{3.6}
$$

A equação 3.6 mostra que o incremento da frequência aumenta com a frequência.

- **Filtro Butterworth** A partir da estratégia descrita acima, a frequência máxima de cada banda de frequência f_n é escolhida e aplicada como frequência de corte para o filtro passa-baixa (resposta ao impulso infinito) Butterworth. Esse filtro se comporta de modo que a resposta de amplitude é maximamente plana na banda passante e monotônica na banda rejeitada, e os dados são filtrados no domínio do tempo sem a necessidade de transformá-los para o domínio da frequência. A forma do filtro Butterworth é (Oppenheim and Willsky, 1992):

$$
|B_u(f)|^2 = \frac{1}{1+(f/f_c)^{2n}} \tag{3.7}
$$

Onde B_u(f) é a função transfência, sendo f a frequência angular do sinal em radianos por segundo e f_c a frequência de corte. A ordem do filtro é definida por n, neste trabalho a ordem utilizada é n=6. Além do filtro Butterworth, também é utilizado um filtro passa-alta de 3 Hz no dado e na wavelet.

3.3.1.1 Escolha da primeira banda de frequência

A fim de escolher a primeira banda de frequência, utilizei o espectro de amplitudes da Figura 3.9 correspondente ao tiro 4 para selecionar as frequências de teste da primeira banda (f1). Foram aplicados três testes de inversão utilizando as frequências 3 hz, 4 hz e 5 hz nas posições de primeira banda. As frequências utilizadas nas bandas seguintes foram calculada através da Equação 3.6, o qual utilizou $\alpha_{min} = 0.8$ que corresponde à profundidade do modelo z = 1500 m e ao valor do máximo meio-afastamento $h_{max} = 375$ m.

As bandas de frequência utilizadas nos testes representam a frequência máxima de corte para o filtro passa baixa Butterworth, com uma ordem n = 6. Foram utilizadas oito bandas de frequência em cada teste e, exceto para o último estágio da inversão que cobre o conteúdo de freqüência total dos dados, são realizadas 20 iterações para cada banda de frequência. Assim, o resultado de cada banda de frequência inferior é usado como o modelo inicial para a próxima inversão de frequência mais alta. O número máximo de iteração utilizado foi 300.

Para cada valor de banda de frequência selecionada na etapa da inversão multiescala vou ter os valores de k_{zmin} e k_{zmax} associados, e valores de λ_{min} e λ_{max} através da relação $k_z = \frac{2\pi}{\lambda}$. A tabela 3.5 mostra os valores mínimos e máximos do número de onda vertical k_z e comprimento de onda λ associados às frequências 3 Hz, 4 Hz e 5 Hz utilizadas nos testes de inversão. A velocidade média do modelo verdadeirocom é $c_0 = 1826$ m/s. A Figura 3.10 mostra os resultados obtidos nos três testes de inversão. O primeiro experimento (Figura 3.10a) utilizou como primeira banda de frequência $f(1) = 3$ hz. No segundo experimento (Figura 3.10b) a primeira banda de frequência utilizada foi $f(1) = 4$ hz e no terceiro experimento (Figura 3.10c) a primeira frequência escolhida foi $f(1) = 5$ hz. A tabela 3.6 mostra as bandas de frequência utilizadas em cada teste.

É possível observar que o teste 2 que utilizou $f_{max}(1)$= 4 Hz (Figura 3.10b) proporcionou a melhor reconstrução do modelo de velocidades da onda P quando comparado aos outros dois testes, isso significa que os dados iniciais desse teste cumprem o critério de meio-comprimento de onda (2.35) e foge do problema de saltos de ciclo. No entanto, a reconstrução do modelo falha com o aumento da profundidade em todos os testes, isso se deve ao fato do aumento da não-linearidade do problema e aumento na dificuldade de encontrar o mínimo global da função objeto.

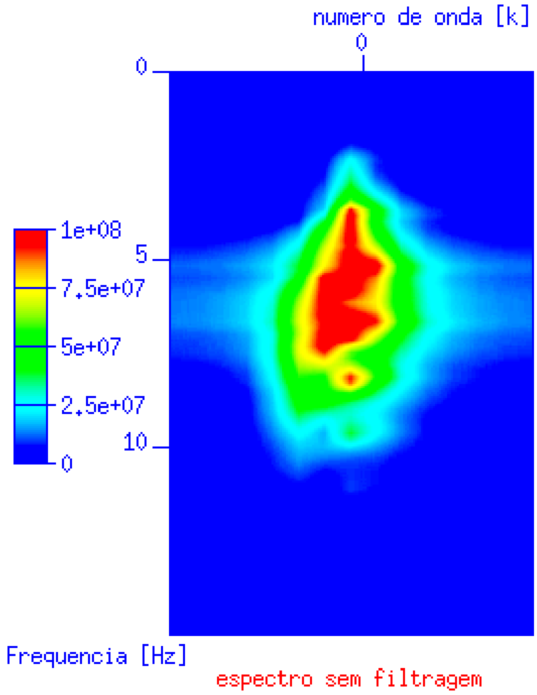


Figura 3.9: Espectro de amplitudes do tiro 4 localizado em x=950 m

Tabela 3.5: Valores máximos e mínimos dos números de onda verticais k_z e comprimentos de onda λ utilizadas nos três testes de inversão para escolha da primeira banda de frequência.

	3 Hz	4 Hz	5 Hz
k_{zmin} $(rad.m^{-1})$	0,016	0,022	0,027
k_{zmax} $(rad.m^{-1})$	0,020	0,027	0,034
λ_{min} (m)	316	232	185
λ_{max} (m)	395	285	234

Tabela 3.6: Valores das bandas de frequências utilizadas nos três testes de inversão para escolha da primeira banda de frequência.

$Teste1 \quad (f1 = 3Hz)$	3, 3,75, 4,6, 5,8, 7,3, 9,5, 11,8, 14,8
$Teste2 \quad (f1 = 4Hz)$	4, 5, 6,25, 7,8, 9,7, 12,2, 15,2, 19
$Teste3 \quad (f1 = 5Hz)$	5, 6, 7,8, 9,7, 12,2, 15,25, 19, 23,75

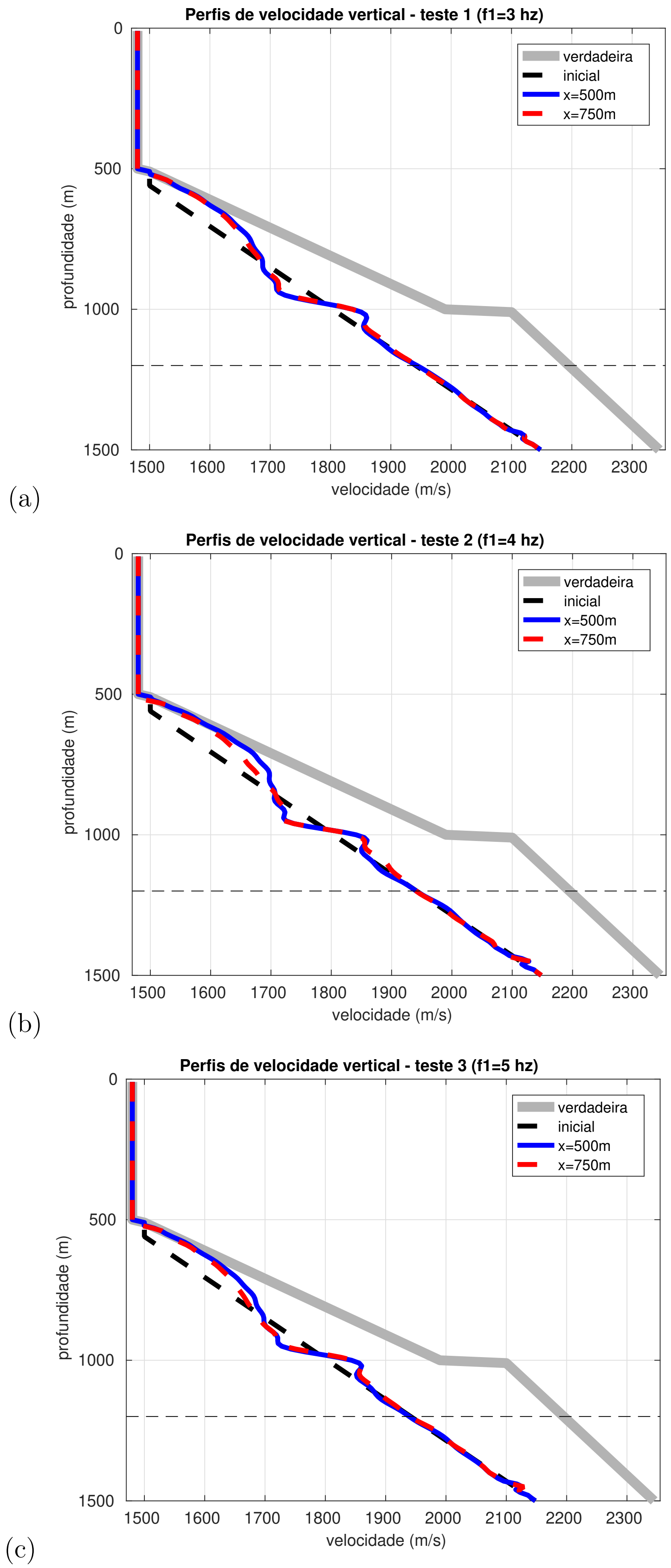


Figura 3.10: Resultados do FWI para a avaliação da primeira banda de frequência. Resultados dos três testes: (a) $f_{max}(1) = 3$ hz, (b) $f_{max}(1) = 4$ hz, (c) $f_{max}(1) = 5$ hz.

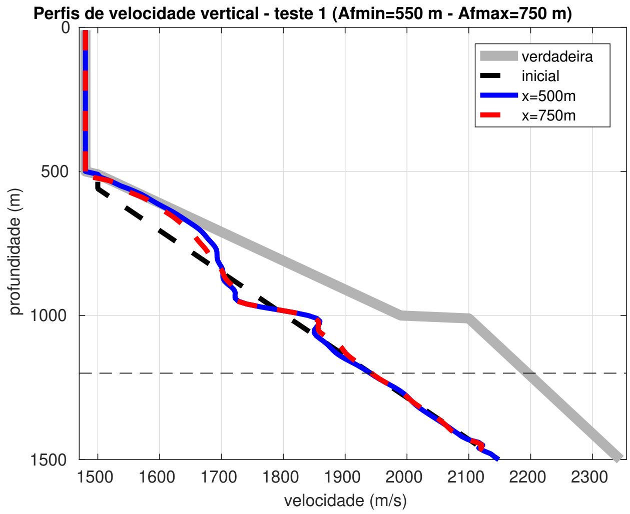
3.3.1.2 Janelamento de afastamentos

A inversão multiescala está diretamente relacionada ao critério de convergência de meio comprimento de onda (Eq. 2.35), pois o valor de desajuste do tempo de trânsito δt entre o dado observado e calculado é modificado com a filtragem do número de comprimento de onda N_λ, o que significa filtrar também bandas de frequência. Para amenizar o critério de convergência é necessário aumentar o valor limite de δt para fugir do efeito de salto de ciclo. Isso pode ser feito diminuindo o valor de N_λ associado à afastamentos longos do dado observado, o que significa eliminar as baixas frequências do dado.

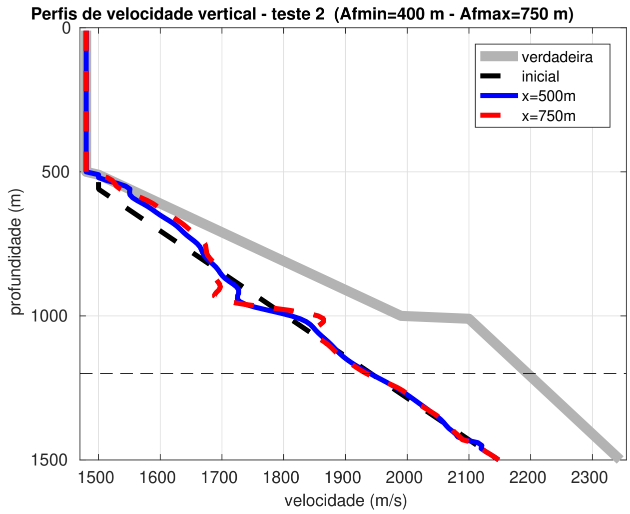
É importante destacar que existe um paradoxo associado ao conteúdo de afastamentos longos no processo de inversão que deve ser avaliado na etapa de janelamento de afastamentos. Os dados de afastamento longo são necessários para delinear estruturas de longo comprimento de onda do modelo. Entretanto, informações de gandes afastamentos também correspondem a altos valores de N_λ que tornam o critério de convergencia do método mais rígido, o que conduz à convergência de um mínimo local indesejado.

A fim de melhorar a reconstrução do modelo de velocidades da onda P, foram realizados três experimentos adicionais com diferentes intervalos de afastamento com um valor máximo e mínimo. O cálculo do afastamento é realizado através da diferença entre a posição da primeira fonte $x_s = 350$ m e a posição do receptor x_r escolhido no teste. O afastamento máximo da geometria de aquisição é 750 m referente à posição do último receptor $x_r = 1100$ e foi escolhido nos três testes como o valor máximo do intervalo de afastamento. Assim, em cada teste é alterado somente os valores mínimos do intervalo de afastamento fonte-receptor. Foram mantidas as configurações da inversão multiescala do teste 2 anterior (Figura 3.10b).

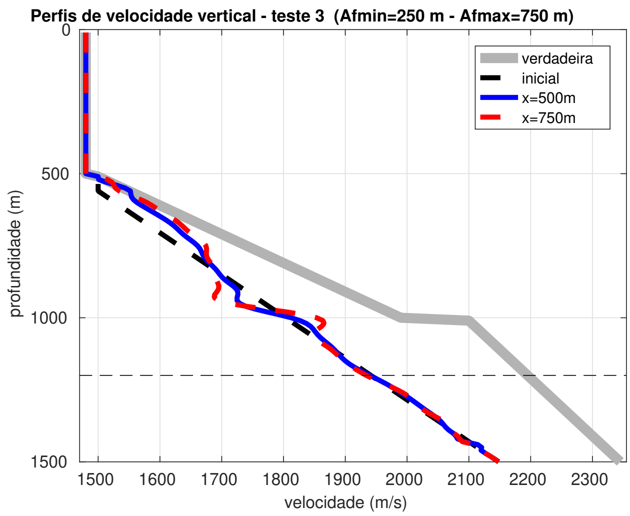
Nota-se que o teste 1 (Figura 3.11a) que utilizou o valor de afastamento (Afmin) igual a 550 m proporcionou a melhor reconstrução do modelo de velocidades da onda P quando comparado aos outros dois testes, isso significa que a eliminação do intervalo de afastamento entre Afmin=550 m e Afmax=750 m do dado observado consegue ajustar o critério de convergência (Eq. 2.35) sem eliminar informações necessárias para a reconstrução do modelo. Enquanto que os testes 2 e 3 (Figura 3.11b,c) mostram que o aumento do intervalor de afastamento eliminado pode ser prejudicial para o processo reconstrução do modelo devido à eliminação de dados associados à reconstrução das partes profundas do modelo.



(a)



(b)



(c)

Figura 3.11: Resultados do FWI para a avaliação da eliminação de afastamentos do dado observado. Resultados dos três testes: (a) intervalo eliminado do afastamento: mínimo = 550 m - máximo = 750 m, (b) intervalo eliminado do afastamento: mínimo = 400 m - máximo = 750 m, (c) intervalo eliminado do afastamento: mínimo = 250 m - máximo = 750 m.

3.3.1.3 Escolha do número de bandas de frequência

A inversão de forma de onda no domínio do tempo está sujeita à convergência para mínimos locais devido à alta não-linearidade do problema, como já observado nos experimentos anteriores. Na tentativa de diminuir a não-linearidade, foram aplicados três experimentos aumentando o número de bandas de frequência utilizadas na inversão, além disso, alterei o valor da profundidade utilizado no cálculo do incremento de frequências (Eq.(3.6)) para z = 1000 m, resultando em $\alpha_{min} = 0.9$.

As configurações dos testes anteriores que produziram a melhor reconstrução do modelo de velocidade da onda P foram mantidas nessa etapa de escolha do número de bandas de frequência, ou seja, a primeira banda de frequência rutilizada foi 4 hz (Figura 3.10b) e o janelamento do afastamento fonte-receptor eliminado foi 550 m - 750 m (Figura 3.11a). A alteração no valor da profundidade k alterou os valores da escala de k_z (Eq. (3.3)), assim como o cálculo do incremento de frequências (Eq. (3.6)). Os novos valores da escala k_z e as bandas de de frequências utilizadas nos testes são mostrados nas tabelas 3.7 e 3.8, respectivamente.

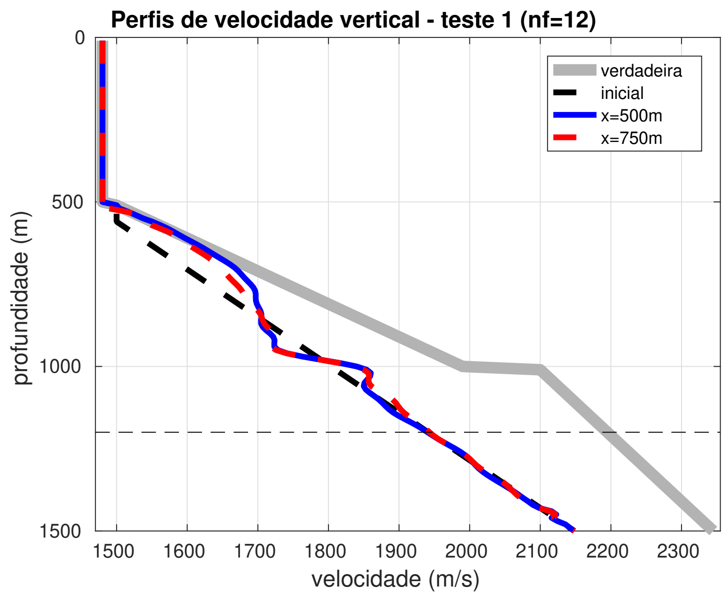
Nota-se que o aumento do número de bandas de frequência melhorou a reconstrução do modelo de velocidade da onda P até aproximadamente a profundidade de 700 m. Esse resultado indica que o dado observado não apresenta informação da subsuperfície o suficiente , ou seja, há a necessidade de mais dados de afastamento longo para a reconstrução das partes mais profundas do modelo. resultados dos testes 2 e 3 (Figura 3.12b,c) estão muito próximos, o que indica que 16 é o número de bandas de frequência que abrange todo o dado disponível, para nf = 20 não houve melhoras na reconstrução do modelo, além do aumento no custo computacional.

Tabela 3.7: Valores máximos e mínimos dos números de onda verticais k_z e comprimentos de onda λ para a banda de frequência de 4 Hz utilizada nos testes para escolha do número de bandas de frequência.

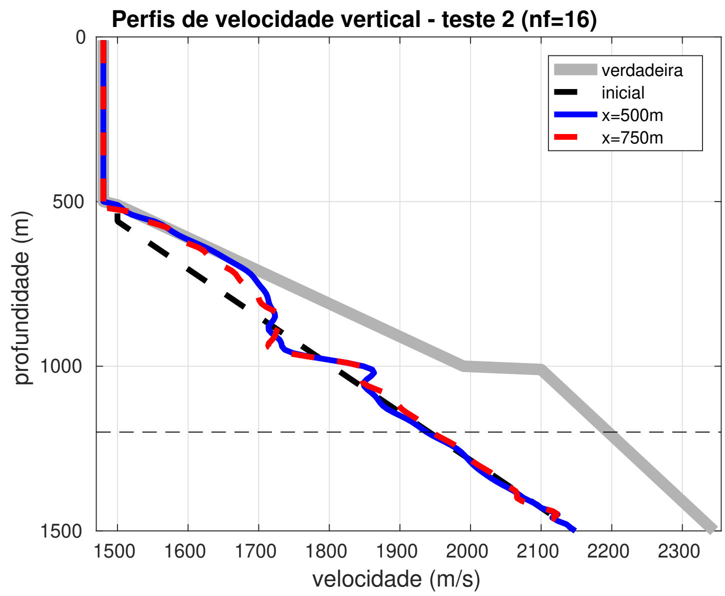
k_{zmin} $(rad.m^{-1})$	0,024
k_{zmax} $(rad.m^{-1})$	0,027
λ_{min} (m)	232
λ_{max} (m)	261

Tabela 3.8: Valores das bandas de frequências utilizadas nos três testes de inversão para escolha do número de bandas de frequência.

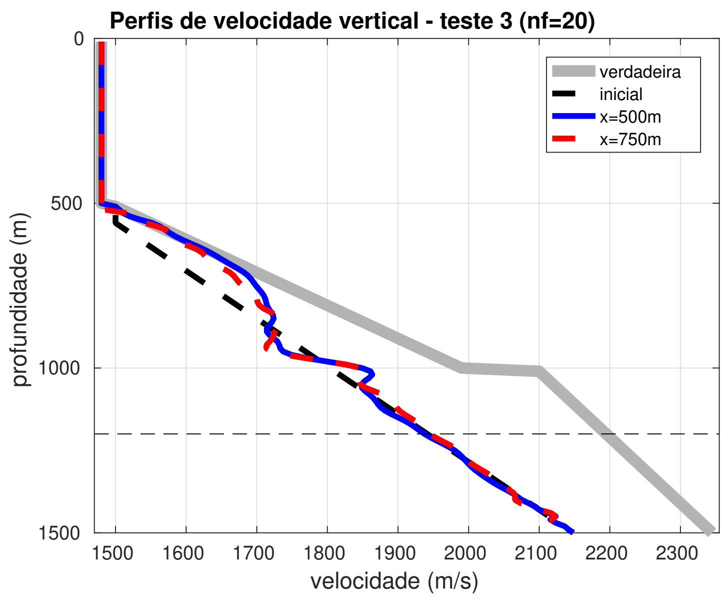
$Teste1$	$(nf = 12)$	4, 4,4, 4,9, 5,4, 6, 6,6, 7,4, 8,2, 9,1, 10,12, 11,2, 12,4
$Teste2$	$(nf = 16)$	4, 4,4, 4,9, 5,4, 6, 6,6, 7,4, 8,2, 9,1, 10,12, 11,2, 12,4 13,7, 15,3, 17, 18,8
$Teste3$	$(nf = 20)$	4, 4,4, 4,9, 5,4, 6, 6,6, 7,4, 8,2, 9,1, 10,12, 11,2, 12,4 13,7, 15,3, 17, 18,8, 20,8, 23,2, 29, 32,2



(a)



(b)



(c)

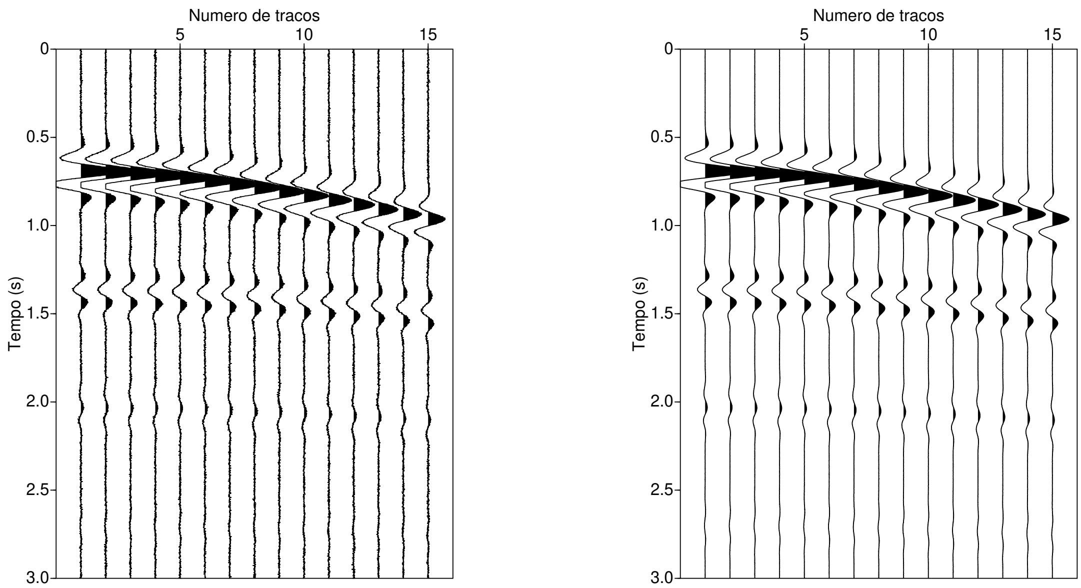
Figura 3.12: Resultados do FWI para a avaliação do número de bandas de frequência (nf). Resultados dos três testes: (a) nf = 12, (b) nf = 16, (c) nf = 20.

3.4 AVALIAÇÃO DO RUÍDO

A fim de avaliar a estabilidade da inversão quando o dado observado contém ruído, foi adicionado ruído aletório ao dado sintético (Figura 3.3) utilizado nos testes anteriores através da equação:

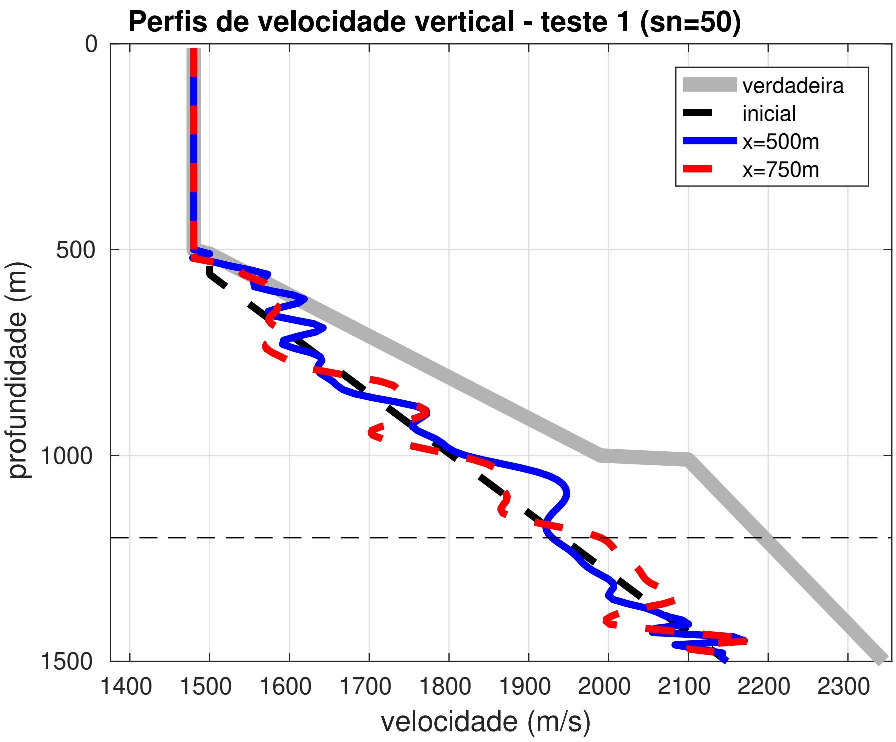
$$sn = 10\log_{10}\left[\frac{\sum_n (d_n(t))^2}{\sum_n (d_n(t) - r_n(t))^2}\right] \tag{3.8}$$

Onde sn é a razão sinal-ruído(dB), d é o vetor do sinal original, r é o vetor do ruído e n é o comprimento do vetor do sinal. No primeiro teste utilizei $sn = 50$ e em seguida $sn = 500$. A Figura 3.14 mostra os resultados da inversão para os dois testes de sn. Através do teste 1 (Figura 3.14a) é possível perceber que o ruído produz valores de desajuste entre o dado observado e calculado δt (Eq. (2.2)) no domínio do dado que causam instabilidade no cálculo do modelo de velocidade da onda P. Enquanto que o teste 2 (Figura 3.14b) que utilizou o valore de $sn = 500$ se mostrou mais estável devido a diminuição da influência do ruído na inversão. No primeiro experimento (Figura 3.14a) foi utilizado $sn = 50$ e no segundo experimento (Figura 3.14b) foi utilizado $sn = 500$.

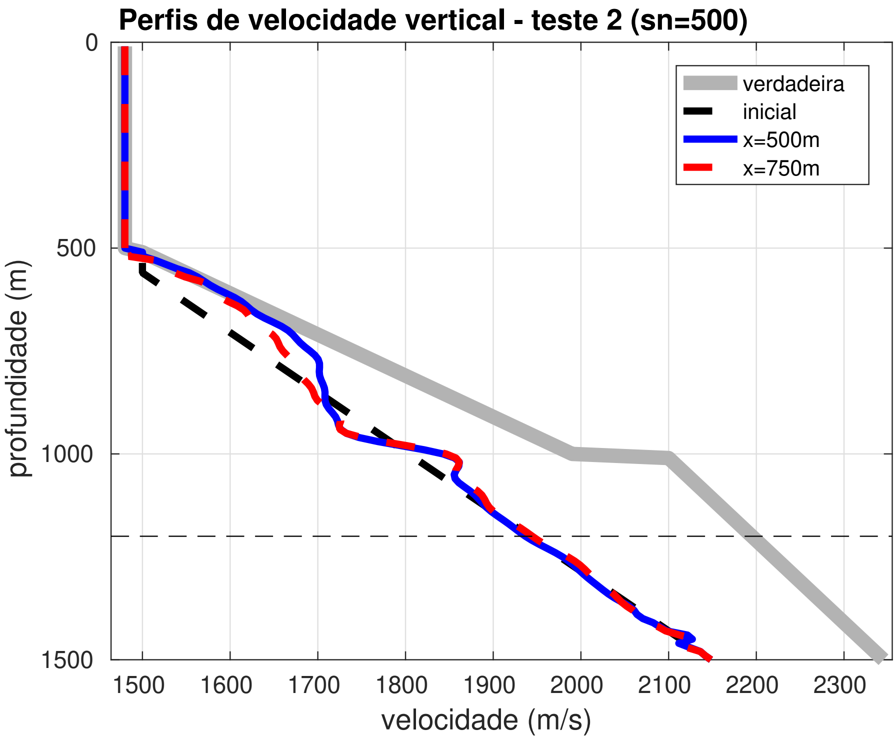


(a) (b)

Figura 3.13: Dados de famílias de tiro comum com ruído aletório. (a) Dado com razão sinal-ruído (sn) = 50 utilizado no teste 1, (b) Dado com razão sinal-ruído (sn) = 500 utilizado no teste 2.



(a)



(b)

Figura 3.14: Resultados do FWI para a avaliação do ruido (sn). Resultados dos dois testes: (a) $sn = 50$, (b) $sn = 500$.

4 CONCLUSÃO

Neste trabalho foi aplicado o método de inversão da forma de onda completa para o caso 1-D acústico. A aplicação foi avaliada sobre dados sintéticos com foco na reconstrução dos modelos de velocidade da onda p. Foram estudados aspectos como o efeito da escolha do modelo inicial na convergência do método e o efeito da parametrização da geometria e frequência na precisão e acurácia do problema inverso. Também foi avaliada a estabilidade do método a partir da contaminação do dados observado por ruído aleatório. O modelo final de velocidade da onda P obtido pela inversão se ajustou ao modelo verdadeiro até um limite de profundidade, como ja esperado devido a alta não-linearidade.

O estudo do efeito da aproximação inicial na convergência do método FWI mostrou que a aproximação inicial do modelo de velocidade deve ser o mais próximo possível do valor desejado para obter a convergência mais rápida. Entretanto, os testes também mostraram que o método exige a utilização de estratégias de inversão para mitigar o problema de mínimos locais, pois o modelo de aproximação inicial escolhido apenas pelo critério de convergência mais rápida não foi capaz de minimizar a função objeto até o mínimo global.

Dessa forma, a abordagem da inversão multiescala foi aplicada e se mostrou uma estratégia eficiente para evitar o problema da alta não-linearidade do problema inverso. Foram avaliados três aspectos sobre a abordagem: escolha da primeira banda de frequência, janelamento de afastamentos e escolha do número de bandas de frequência. Em todos os testes a inversão iniciou-se com baixas frequências e o conteúdo de frequências maiores foi gradualmente adicionado para fugir dos mínimos locais.

Os testes para selecionar a primeira banda de frequência (f1) mostraram que essa escolha deve conter as menores freqüências disponíveis para aproveitar ao máximo a abordagem multiescala, ou seja, para evitar o problema de salto de ciclo.

A partir dos testes com janelamento de afastamentos nota-se de forma acentuada que eliminar dados de afastamento longo na intenção de adicionar estabilidade ao problema, pode prejudicar consideravelmente o resultado da inversão devido sua importância na reconstrução de estruturas de longo comprimento de onda do modelo.

Os resultados dos testes para escolher o número de bandas de frequência (nf) mostraram que aumentar o número de bandas pode produzir melhoras na reconstrução do modelo, mas essa melhora é limitada e deve ser avaliada para não gerar custos computacionais desnecessários. Além disso, a diminuição no valor da profundidade para calcular α_{min} melhorou a reconstrução do modelo vp, o que corrobora com a limitação do método em reconstruir as partes mais profundas do modelo.

Através dos testes de inversão utilizando o dado observado com ruído foi possível concluir que a estabilidade do método FWI é dependente da qualidade dos dados utilizados, comprovando a necessidade do pré-processamento de dados a fim de atenuar o ruído sísmico.

Os resultados obtidos indicam que a aplicação do FWI incluindo a estratégia de inversão adequada, como por exemplo, a multiescala são promissores na reconstrução dos modelos de velocidade. O problema da forte não-linearidade com o aumento da profundidade pode ser superado melhorando a aproximação inicial e aumentando o volume de dados de afastamentos longos a fim de aumentar a capacidade da abordagem multiescala.

REFERÊNCIAS

Berenger, J.-P., 1994, A perfectly matched layer for the absorption of electromagnetic waves: Journal of computational physics, **114**, 185–200.

Beydoun, W. B., and A. Tarantola, 1988, First born and rytov approximations: Modeling and inversion conditions in a canonical example: The Journal of the Acoustical Society of America, **83**, 1045–1055.

Bleibinhaus, F., R. W. Lester, and J. A. Hole, 2009, Applying waveform inversion to wide-angle seismic surveys: Tectonophysics, **472**, 238–248.

Bohlen, T., 2002, Parallel 3-d viscoelastic finite difference seismic modelling: Computers & Geosciences, **28**, 887–899.

Bohlen, T., A. Kurzmann, D. Koehn, A. Przebindowska, and N. Nguyen, 2009, 2d acoustic full waveform tomography–performance and optimization: Presented at the 71st EAGE Conference and Exhibition incorporating SPE EUROPEC 2009.

Bohlen, T., and E. H. Saenger, 2006, Accuracy of heterogeneous staggered-grid finite-difference modeling of rayleigh waves: Geophysics, **71**, T109–T115.

Boonyasiriwat, C., P. Valasek, P. Routh, W. Cao, G. T. Schuster, and B. Macy, 2009, An efficient multiscale method for time-domain waveform tomography: Geophysics, **74**, WCC59–WCC68.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Courant, R., K. Friedrichs, and H. Lewy, 1928, Über die partiellen differenzengleichungen der mathematischen physik: Mathematische annalen, **100**, 32–74.

Koehene, V., 2018, Fwi multiescala: uma implementação em gpu.

Köhn, D., 2011, Time domain 2d elastic full waveform tomography: PhD thesis, Christian-Albrechts Universität Kiel.

Kurzmann, A., 2012, Applications of 2d and 3d full waveform tomography in acoustic and viscoacoustic complex media: PhD thesis, Verlag nicht ermittelbar.

Levander, A. R., 1988, Fourth-order finite-difference p-sv seismograms: Geophysics, **53**, 1425–1436.

Mesquita, M. J. L., J. C. R. Cruz, and G. G. Callapino, 2019, Velocity inversion by global optimization using finite-offset common-reflection-surface stacking applied to synthetic and tacutu basin seismic data: Geophysics, **84**, R165–R174.

Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52**, 1211–1228.

Nocedal, J., and S. Wright, 2006, Numerical optimization: Springer Science & Business Media.

Oppenheim, A. V., and A. S. Willsky, 1992, Signale und systeme: Lehrbuch: VCH Weinheim etc.

Pica, A., J. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: Geophysics, **55**, 284–292.

Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography. part 2: Elastic wave-equation method: Geophysical Prospecting, **38**, 311–329.

———, 2008, Waveform tomography—successes, cautionary tales, and future directions: 70th EAGE Conference & Exhibition, 1–5.

Pratt, R. G., and M. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography. part 1: Acoustic wave-equation method: Geophysical prospecting, **38**, 287–310.

Prieux, V., G. Lambaré, S. Operto, and J. Virieux, 2013, Building starting models for full waveform inversion from wide-aperture data by stereotomography: Geophysical Prospecting, **61**, 109–137.

Przebindowska, A., 2013, Acoustic full waveform inversion of marine reflection seismic data: PhD thesis, KIT-Bibliothek.

Romdhane, A., G. Grandjean, R. Brossier, F. Réjiba, S. Operto, and J. Virieux, 2011, Shallow-structure characterization by 2d elastic full-waveform inversion: Geophysics, **76**, R81–R93.

Shah, N., M. Warner, T. Nangoo, A. Umpleby, I. Stekl, J. Morgan, and L. Guasch, 2012, Quality assured full-waveform inversion: Ensuring starting model adequacy, *in* SEG Technical Program Expanded Abstracts 2012: Society of Exploration Geophysicists, 1–5.

Shin, C., and Y. Ho Cha, 2009, Waveform inversion in the laplace—fourier domain: Geophysical Journal International, **177**, 1067–1079.

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Spadini, A. S., and L. A. Diogo, 2018, Inversão da forma de onda completa de dados de sísmica de reflexão rasa.

Sun, R., and G. A. McMechan, 1992, 2-d full-wavefield inversion for wide-aperture, elastic, seismic data: Geophysical Journal International, **111**, 1–10.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

———, 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893–1903.

———, 1988, Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *in* Scattering and Attenuations of Seismic Waves, Part I: Springer, 365–399.

Virieux, J., 1986, P-sv wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, 889–901.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Warner, M., 2012, Eage short course full-wavefield tomography/ full-waveforme inversion: a game changing technology.

Yang, P., J. Gao, and B. Wang, 2015, A graphics processing unit implementation of time-domain full-waveform inversion: Geophysics, **80**, F31–F39.

Zahradník, J. í., P. Moczo, and F. e. Hron, 1993, Testing four elastic finite-difference schemes for behavior at discontinuities: Bulletin of the Seismological Society of America, **83**, 107–129.

ANEXOS

A– EQUAÇÃO DA ONDA ACÚSTICA COM FORMULAÇÃO PML

A simulação da propagação da onda em um meio ilimitado necessita ser estável e precisa. A abordagem da camada perfeitamente combinada (PML) provou ser um método flexível e preciso para a simulação de ondas. Trata-se de cercar o domínio computacional por uma camada absorvente, que não gera reflexões em sua interface. Assim, um terno de amortecimento é adicionado à equação da onda, que atua apenas na direção perpendicular à camada. Neste anexo é apresentada a equação da onda acústica com formulação PML segunda Kurzmann (2012).

A equação de onda 3-D homogênea é dada por:

$$\frac{1}{\kappa(\mathbf{x})}\ddot{p}(\mathbf{x},t) = \bigtriangledown \cdot \Big(\frac{1}{\rho(\mathbf{x})} \bigtriangledown p(\mathbf{x},t)\Big), \tag{A–1}$$

e com o módulo de bulk:

$$K(\mathbf{x}) = \rho(\mathbf{x})v_p^2, \tag{A–2}$$

sendo a pressão p, a velocidade da onda P, v_p, densidade ρ, o vetor posição $\mathbf{x}$ e o tempo t. O campo de pressão e sua segunda derivada no tempo podem ser transformados para o domínio de Laplace através da equação:

$$\underline{p}(\mathbf{x},s) = \int_0^\infty e^{st}p(\mathbf{x},t)dt, \qquad s \in \mathbb{C}, \tag{A–3}$$

onde $\mathbb{C}$ produzindo o campo de onda complexo $\underline{p}$:

$$\begin{aligned} \frac{1}{\rho(\mathbf{x})v_p^2(\mathbf{x})}s^2\underline{p}(\mathbf{x},s) &= \bigtriangledown \cdot \Big(\frac{1}{\rho(\mathbf{x})} \bigtriangledown \underline{p}(\mathbf{x},s)\Big) \\ \frac{1}{v_p^2}s^2\underline{p} = [\partial_\mathbf{x}^2 + \partial_y^2 + \partial_z^2]\underline{p} &- \frac{1}{\rho}[\partial_\mathbf{x}\rho\partial_\mathbf{x} + \partial_y\rho\partial_y + \partial_z\rho\partial_z]\underline{p} \end{aligned} \tag{A–4}$$

A próxima etapa é o alongamento de coordenadas que tem efeito dentro do limite de absorção, substituindo os operadores diferenciais:

$$\partial_{x,y,z} \mapsto \frac{1}{1+\frac{\sigma_{x,y,z}}{s}}\partial_{x,y,z} = \frac{1}{\gamma_{x,y,z}}\partial_{x,y,z} \tag{A–5}$$

Onde $\gamma_{x,y,z}$ são variáveis auxiliares e $\sigma_{x,y,z}$ são os coeficientes de atenuação nas direções x, y e z:

$$\sigma_x = \sigma_y = \sigma_z = 0 \qquad\qquad \text{(domínio interior)},$$

$$\sigma_x > 0, \sigma_y > 0, \sigma_z > 0 \qquad\qquad \text{(limite PML)}.$$

Após a aplicação do novo operador diferencial na equação (A–4)

$$\begin{aligned} \frac{1}{v_p^2}s^2\underline{p} = \frac{1}{\gamma_x}\partial_x(\frac{1}{\gamma_x}\partial_x\underline{p}) + \frac{1}{\gamma_y}\partial_y(\frac{1}{\gamma_y}\partial_y\underline{p}) + \frac{1}{\gamma_z}\partial_z(\frac{1}{\gamma_z}\partial_z\underline{p}) \\ -\frac{1}{\rho}[\frac{1}{\gamma_x^2}\partial_x\rho\cdot\partial_x + \frac{1}{\gamma_y^2}\partial_y\rho\cdot\partial_y + \frac{1}{\gamma_z^2}\partial_z\rho\cdot\partial_z]\underline{p} \end{aligned} \tag{A–6}$$

Multiplicação de ambos os lados com $\gamma_x\gamma_y\gamma_z$,

$$\begin{aligned} \frac{1}{vp_p^2}\gamma_x\gamma_y\gamma_z s^2\underline{p} = \partial_x(\frac{\gamma_y\gamma_z}{\gamma_x}\partial_x\underline{p}) + \partial_y(\frac{\gamma_x\gamma_z}{\gamma_y}\partial_y\underline{p}) + \partial_z(\frac{\gamma_y\gamma_x}{\gamma_z}\partial_z\underline{p}) \\ -\frac{1}{\rho}[\frac{\gamma_y\gamma_z}{\gamma_x}\partial_x\rho\cdot\partial_x\underline{p} + \frac{\gamma_x\gamma_z}{\gamma_y}\partial_y\rho\cdot\partial_y\underline{p} + \frac{\gamma_y\gamma_x}{\gamma_z}\partial_z\rho\cdot\partial_z\underline{p}], \end{aligned} \tag{A–7}$$

Avaliando os termos

$$\begin{aligned} \frac{\gamma_y\gamma_z}{\gamma_x} &= 1 + \frac{s(\sigma_y+\sigma_z-\sigma_x)+\sigma_y\sigma_z}{s(s+\sigma_x)}, \\ \frac{\gamma_x\gamma_z}{\gamma_y} &= 1 + \frac{s(\sigma_x+\sigma_z-\sigma_y)+\sigma_x\sigma_z}{s(s+\sigma_y)}, \\ \frac{\gamma_x\gamma_y}{\gamma_z} &= 1 + \frac{s(\sigma_x+\sigma_y-\sigma_z)+\sigma_y\sigma_x}{s(s+\sigma_z)}, \\ \gamma_x\gamma_y\gamma_z s^2 &= s^2 + s(\sigma_x+\sigma_y+\sigma_z) + \sigma_y\sigma_z + \sigma_x\sigma_y + \sigma_x\sigma_z + \frac{\sigma_x\sigma_y\sigma_z}{s} \end{aligned} \tag{A–8}$$

Pode-se obter a seguinte equação:

$$\begin{aligned} \frac{1}{v_p^2}[s^2 + s(\sigma_x+\sigma_y+\sigma_z) + \sigma_y\sigma_z + \sigma_x\sigma_y + \sigma_x\sigma_z + \frac{\sigma_x\sigma_y\sigma_z}{s}]\underline{p} = \\ [\partial_x^2\partial_y^2\partial_z^2]\underline{p} - \frac{1}{\rho}\partial_x\rho\cdot\partial_x\underline{p} - \frac{1}{\rho}\partial_y\rho\cdot\partial_y\underline{p} - \frac{1}{\rho}\partial_z\rho\cdot\partial_z\underline{p} \\ +\partial_x[\frac{s(\sigma_y+\sigma_z-\sigma_x+\sigma_y\sigma_z)}{s(s+\sigma_x)}\partial_x\underline{p}] + \partial_y[\frac{s(\sigma_x+\sigma_z-\sigma_y+\sigma_x\sigma_z)}{s(s+\sigma_y)}\partial_y\underline{p}] \\ \partial_z[\frac{s(\sigma_x+\sigma_y-\sigma_z+\sigma_y\sigma_x)}{s(s+\sigma_z)}\partial_z\underline{p}] \\ -\frac{s(\sigma_y+\sigma_z-\sigma_x)+\sigma_y\sigma_z}{s(s+\sigma_x)}\frac{1}{\rho}\partial_x\rho\cdot\partial_x\underline{p} - \frac{s(\sigma_x+\sigma_z-\sigma_y)+\sigma_x\sigma_z}{s(s+\sigma_y)}\frac{1}{\rho}\partial_y\rho\cdot\partial_y\underline{p} \\ -\frac{s(\sigma_y+\sigma_x-\sigma_z)+\sigma_y\sigma_x}{s(s+\sigma_z)}\frac{1}{\rho}\partial_z\rho\cdot\partial_z\underline{p} \end{aligned} \tag{A–9}$$

A substituição dos termos

$$\begin{aligned}\underline{w}_x &:= \frac{s(\sigma_y+\sigma_z-\sigma_x)+\sigma_y\sigma_z}{s(s+\sigma_x)}\partial_x\underline{p},\\ \underline{w}_y &:= \frac{s(\sigma_x+\sigma_z-\sigma_y)+\sigma_x\sigma_z}{s(s+\sigma_y)}\partial_y\underline{p},\\ \underline{w}_z &:= \frac{s(\sigma_x+\sigma_y-\sigma_z)+\sigma_x\sigma_y}{s(s+\sigma_z)}\partial_z\underline{p},\\ \underline{u} &:= \frac{\underline{p}}{s}.\end{aligned} \quad \text{(A–10)}$$

Pelo campos de ondas auxiliares $\underline{\mathbf{w}}$ e $\underline{\mathrm{u}}$ simplificam a equação (A–9) e a introdução de quatro equações diferenciis parciais. As equações (A–9) e (A–10) são transformadas de volta para o domínio do tempo e o sistema pode ser escrito como:

$$\begin{aligned}\ddot{p}+(\sigma_x+\sigma_y+\sigma_z)\dot{p}+(\sigma_y\sigma_z+\sigma_x\sigma_y+\sigma_x\sigma_z)p &= v_p^2\Big[\Delta p+\nabla.\,\mathbf{w}-\frac{1}{\rho}\nabla\rho\cdot(\nabla p+\mathbf{w})\Big]-\sigma_x\sigma_y\sigma_z u,\\ \dot{w}_x &= (\sigma_y+\sigma_z-\sigma_x)\partial_x p+\sigma_y\sigma_z\partial_x u-\sigma_x w_x,\\ \dot{w}_y &= (\sigma_x+\sigma_z-\sigma_y)\partial_y p+\sigma_x\sigma_z\partial_y u-\sigma_y w_y,\\ \dot{w}_z &= (\sigma_x+\sigma_y-\sigma_z)\partial_z p+\sigma_x\sigma_y\partial_z u-\sigma_z w_z,\\ \dot{u} &:= p.\end{aligned} \quad \text{(A–11)}$$

A Equação (A–11) para o caso da onda acústica 2D é reduzida em:

$$\begin{aligned}\ddot{p}+(\sigma_x+\sigma_y)\ddot{p}+\sigma_x\sigma_y p &= v_p^2[\Delta p+\nabla.\,\mathbf{w}-\frac{1}{\rho}\nabla\rho\cdot(\nabla p+\mathbf{w})],\\ \dot{w}_x &= (\sigma_y-\sigma_x)\partial_x p-\sigma_x w_x,\\ \dot{w}_y &= (\sigma_x-\sigma_y)\partial_y p-\sigma_y w_y.\end{aligned} \quad \text{(A–12)}$$

A hipótese de densidade homogênea produz equações de onda modificadas em 3D,

$$\begin{aligned}\ddot{p}+(\sigma_x+\sigma_y+\sigma_z)\dot{p}+(\sigma_y\sigma_z+\sigma_x\sigma_y+\sigma_x\sigma_z)p &= v_p^2[\Delta p+\nabla\cdot\mathbf{w}]-\sigma_x\sigma_y\sigma_z u,\\ \dot{w}_x &= (\sigma_y+\sigma_z-\sigma_x)\partial_x p+\sigma_y\sigma_z\partial_x u-\sigma_x w_x,\\ \dot{w}_y &= (\sigma_x+\sigma_z-\sigma_y)\partial_y p+\sigma_x\sigma_z\partial_y u-\sigma_y w_y,\\ \dot{w}_z &= (\sigma_x+\sigma_y-\sigma_z)\partial_z p+\sigma_x\sigma_y\partial_z u-\sigma_z w_z,\\ \dot{u} &= p,\end{aligned} \quad \text{(A–13)}$$

E em 2D,

$$\begin{aligned}\ddot{p}+(\sigma_x+\sigma_y)\ddot{p}+\sigma_x\sigma_y p &= v_p^2[\Delta p+\nabla\cdot\mathbf{w}],\\ \dot{w}_x &= (\sigma_y-\sigma_x)\partial_x p-\sigma_x w_x,\\ \dot{w}_y &= (\sigma_x-\sigma_y)\partial_y p-\sigma_y w_y\end{aligned} \quad \text{(A–14)}$$

Os parâmetros de atenuação $\sigma_{x,y,z}$ podem ser calculados a partir de funções matemáticas $f_{x,y,z}$ como funções quadráticas, exponenciais ou cossenos. O que garante uma transição suave do interior do domínio do modelo para dentro do limite da PML. Os coeficientes são definidos da seguinte forma:

$$\sigma_{x,y,z} = \sigma_0 f_{x,y,z}, \quad com \quad 0 \leq f_{x,y,z} \leq 1 \quad e \quad \sigma_0 = -\frac{\tilde{v}_p ln(R)}{L}, \tag{A–15}$$

A constante σ_0 depende da discretização e da espessura da camada, onde $\tilde{v}_p$ é a velocidade média da onda P, L e R correspondem à espessura e reflexão relativa do limite absorvente, respectivamente.

A solução por diferenças finitas (FD) da equação de onda acústica completa requer a aproximação de derivadas parciais usando operadores lineares discretos e discretização FD (2.11). As seguintes considerações limitam-se aos operadores FD de segunda ordem no tempo e espaço. Para o componente de pressão p, isso inclui as derivadas centrais dentro do interior do domínio:

$$\partial_x^2 p \approx \frac{p_{i+1} + p_{i-1} - 2p_i}{(\Delta h)^2}, \qquad \ddot{p} \approx \frac{p^{n+1} + p^{n-1} - 2p^n}{(\Delta t)^2}. \tag{A–16}$$

O uso das condições de contorno da PML requer equações adicionais de primeira ordem e variáveis auxiliares correspondentes, como $w = (w_x, w_y, w_z)^T$. Aproximações FD exemplares são

$$\partial_x w_x \approx \frac{w_{x|i} - w_{x|i-1}}{\Delta h}, \qquad e \qquad \dot{w}_x \approx \frac{w_x^n - w_x^{n-1}}{\Delta t}. \tag{A–17}$$

A mistura de equações de primeira e de segunda ordem provoca os cálculos de p em pontos da malha completos e intervalos de tempo, enquanto os componentes de $\mathbf{w}$ são deslocados por meio ponto da malha e meio passo de tempo. Assim, a média adicional de campo de onda é necessária dentro do limite da PML.

A derivação produz as seguintes relações gerais para o campo de pressão p, campo escalar auxiliar u bem como os componentes do campo de vetores auxiliares w_x, w_y e w_z. O uso das variáveis auxiliares é limitado ao PML. A atualização do campo de pressão no momento $n + 1$ e a posição da malha $(k,\, j,\, i)$ é calculada por

$$
\begin{aligned}
p^{n+1}_{k,j,i} = s_{k,j,i}\Big[& v^2_{p|k,j,i}\frac{(\Delta t)^2}{(\Delta h)^2}\Big(p^n_{k,j,i-1} + p^n_{k,j,i+1} + p^n_{k,j-1,i}p^n_{k,j+1,i} + p^n_{k-1,j,i} + p^n_{k+1,j,i} - 6p^n_{k,j,i}\Big) \\
& v^2_{p|k,j,i}\frac{(\Delta t)^2}{(\Delta h)^2}\Big(\tilde{w}_{x,k,j,i+\frac{1}{2}} - \tilde{w}_{x,k,j,i-\frac{1}{2}} + \tilde{w}_{y,k,j+\frac{1}{2},i} - \tilde{w}_{y,k,j-\frac{1}{2},i} + \tilde{w}_{z,k+\frac{1}{2},j,i} - \tilde{w}_{z,k-\frac{1}{2},j,i}\Big) \\
& +\Big(2 - (\Delta t)^2\sigma_{y|j}\sigma_{z|k} + \sigma_{x|i}\sigma_{y|j} + \sigma_{x|i}\sigma_{z|k}\Big)p^n_{k,j,i} \\
& +\Big(\frac{\Delta t}{2}(\sigma_{x|i} + \sigma_{y|j} + \sigma_{z|k}) - 1\Big)p^{n-1}_{k,j,i} \\
& -\sigma_{x|i}\sigma_{y|j}\sigma_{z|k}\frac{(\Delta t)^2}{2}\Big(u^{n+\frac{1}{2}}_{k,j,i} - u^{n-\frac{1}{2}}_{k,j,i}\Big) \\
& +r_{x|k,j,i}\Big(\frac{1}{2\Delta h}(p^n_{k,j,i+1} - p^n_{k,j,i-1}) + \frac{1}{2}(\tilde{w}^n_{x|k,j,i-\frac{1}{2}} + \tilde{w}^n_{x|k,j,i+\frac{1}{2}})\Big) \\
& +r_{y|k,j,i}\Big(\frac{1}{2\Delta h}(p^n_{k,j+1,i} - p^n_{k,j-1,i}) + \frac{1}{2}(\tilde{w}^n_{y|k,j-\frac{1}{2},i} + \tilde{w}^n_{y|k,j+\frac{1}{2},i})\Big) \\
& +r_{z|k,j,i}\Big(\frac{1}{2\Delta h}(p^n_{k+1,j,i} - p^n_{k-1,j,i}) + \frac{1}{2}(\tilde{w}^n_{z|k-\frac{1}{2},j,i} + \tilde{w}^n_{z|k+\frac{1}{2},j,i})\Big)\Big]
\end{aligned}
\tag{A–18}
$$

com o termo PML

$$
s_{k,j,i} = \frac{1}{1 + \frac{\Delta t}{2}(\sigma_{x|i} + \sigma_{y|j} + \sigma_{z|k})},
\tag{A–19}
$$

os termos relacionados à densidade

$$
\begin{aligned}
r_{x|k,j,i} &= -v^2_{P|k,j,i}\frac{(\Delta t)^2}{2\Delta h}\frac{\rho_{k,j,i+1} - \rho_{k,j,i-1}}{\rho_{k,j,i}}, \\
r_{y|k,j,i} &= -v^2_{P|k,j,i}\frac{(\Delta t)^2}{2\Delta h}\frac{\rho_{k,j+1,i} - \rho_{k,j-1,i}}{\rho_{k,j,i}}, \\
r_{x|k,j,i} &= -v^2_{P|k,j,i}\frac{(\Delta t)^2}{2\Delta h}\frac{\rho_{k+1,j,i} - \rho_{k-1,j,i}}{\rho_{k,j,i}}
\end{aligned}
\tag{A–20}
$$

e as médias dos campos de onda auxiliares

$$
\begin{aligned}
\tilde{w}^n_{x|k,j,i+\frac{1}{2}} &= \frac{1}{4}(\tilde{w}^n_{x|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{x|k-\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{x|k-\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{x|k+\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}}),\\
\tilde{w}^n_{x|k,j,i-\frac{1}{2}} &= \frac{1}{4}(\tilde{w}^n_{x|k+\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}} + \tilde{w}^n_{x|k-\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}} + \tilde{w}^n_{x|k-\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}} + \tilde{w}^n_{x|k+\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}}),\\
\tilde{w}^n_{y|k,j+\frac{1}{2},i} &= \frac{1}{4}(\tilde{w}^n_{y|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{y|k-\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{y|k-\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}} + \tilde{w}^n_{y|k+\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}}),\\
\tilde{w}^n_{y|k,j-\frac{1}{2},i} &= \frac{1}{4}(\tilde{w}^n_{y|k+\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{y|k-\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{y|k-\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}} + \tilde{w}^n_{y|k+\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}}),\\
\tilde{w}^n_{z|k+\frac{1}{2},j,i} &= \frac{1}{4}(\tilde{w}^n_{z|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{z|k+\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{z|k+\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}} + \tilde{w}^n_{z|k+\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}}),\\
\tilde{w}^n_{z|k-\frac{1}{2},j,i} &= \frac{1}{4}(\tilde{w}^n_{z|k-\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{z|k-\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}} + \tilde{w}^n_{z|k-\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}} + \tilde{w}^n_{z|k-\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}}),\\
u^n_{k,j,i} &= \frac{1}{2}(u^{n+\frac{1}{2}}_{k,j,i} - u^{n-\frac{1}{2}}_{k,j,i}).
\end{aligned} \tag{A–21}
$$

Além disso, o campo de onda do vetor auxiliar é calculado por:

$$
\begin{aligned}
w^{n+1}_{x|k+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}} &= \frac{1}{1+\frac{\Delta t}{2}\sigma_{x|i+\frac{1}{2}}}\Big[\Delta t(\sigma_{y|j+\frac{1}{2}} + \sigma_{z|k+\frac{1}{2}} + \sigma_{x|i+\frac{1}{2}})D_x p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}\\
&\quad +\Delta t\sigma_{y|j+\frac{1}{2}}\sigma_{z|k+\frac{1}{2}}D_x u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \Big(1-\frac{\Delta t}{2}\sigma_{x|i+\frac{1}{2}}\Big)w^n_{x|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}\Big],\\
w^{n+1}_{y|k+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}} &= \frac{1}{1+\frac{\Delta t}{2}\sigma_{x|i+\frac{1}{2}}}\Big[\Delta t(\sigma_{y|j+\frac{1}{2}} + \sigma_{z|k+\frac{1}{2}} + \sigma_{x|i+\frac{1}{2}})D_x p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}\\
&\quad +\Delta t\sigma_{y|j+\frac{1}{2}}\sigma_{z|k+\frac{1}{2}}D_x u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \Big(1-\frac{\Delta t}{2}\sigma_{y|i+\frac{1}{2}}\Big)w^n_{y|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}\Big],\\
w^{n+1}_{z|k+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}} &= \frac{1}{1+\frac{\Delta t}{2}\sigma_{x|i+\frac{1}{2}}}\Big[\Delta t(\sigma_{y|j+\frac{1}{2}} + \sigma_{z|k+\frac{1}{2}} + \sigma_{x|i+\frac{1}{2}})D_x p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}\\
&\quad +\Delta t\sigma_{y|j+\frac{1}{2}}\sigma_{z|k+\frac{1}{2}}D_x u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \Big(1-\frac{\Delta t}{2}\sigma_{z|i+\frac{1}{2}}\Big)w^n_{z|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}\Big],
\end{aligned} \tag{A–22}
$$

com as médias dos coeficientes de atenuação

$$
\sigma_{x|i+\frac{1}{2}} = \frac{1}{2}(\sigma_{x|i} + \sigma_{x|i+1}), \qquad \sigma_{y|j+\frac{1}{2}} = \frac{1}{2}(\sigma_{y|i} + \sigma_{y|j+1}), \qquad \sigma_{z|k+\frac{1}{2}} = \frac{1}{2}(\sigma_{z|k} + \sigma_{z|k+1}),
$$

e os operadores de gradiente aplicados às médias espaciais e temporais do campo de pressão

$$
\begin{aligned}
D_x p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} = \frac{1}{8\Delta h}\Big[&(p^n_{k,j+1,i} + p^n_{k+1,j+1,i} + p^n_{k+1,j+1,i+1} + p^n_{k,j+1,i+1}) \\
&-(p^n_{k,j,i} + p^n_{k+1,j,i} + p^n_{k+1,j,i+1} + p^n_{k,j,i+1}) \\
&+(p^{n+1}_{k,j+1,i} + p^{n+1}_{k+1,j+1,i} + p^{n+1}_{k+1,j+1,i+1} + p^{n+1}_{k,j+1,i+1}) \\
&-(p^n_{k,j,i} + p^n_{k+1,j,i} + p^n_{k+1,j,i+1} + p^n_{k,j,i+1})\Big], \\
D_y p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} = \frac{1}{8\Delta h}\Big[&(p^n_{k,j,i+1} + p^n_{k+1,j,i+1} + p^n_{k,j+1,i+1} + p^n_{k+1,j+1,i+1}) \\
&-(p^n_{k,j,i} + p^n_{k+1,j,i} + p^n_{k,j+1,i} + p^n_{k+1,j+1,i}) \\
&+(p^{n+1}_{k,j,i+1} + p^{n+1}_{k+1,j,i+1} + p^{n+1}_{k,j+1,i+1} + p^{n+1}_{k,j+1,i+1}) \\
&-(p^n_{k,j,i} + p^n_{k+1,j,i} + p^n_{k,j+1,i} + p^n_{k+1,j+1,i})\Big], \\
D_z p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} = \frac{1}{8\Delta h}\Big[&(p^n_{k+1,j,i} + p^n_{k,j+1,i} + p^n_{k+1,j+1,i+1} + p^n_{k+1,j,i+1}) \\
&-(p^n_{k,j,i} + p^n_{k,j+1,i} + p^n_{k,j+1,i+1} + p^n_{k,j,i+1}) \\
&+(p^{n+1}_{k+1,j,i} + p^{n+1}_{k+1,j+1,i} + p^{n+1}_{k+1,j+1,i+1} + p^{n+1}_{k+1,j,i+1}) \\
&-(p^n_{k,j,i} + p^n_{k,j+1,i} + p^n_{k,j+1,i+1} + p^n_{k,j,i+1})\Big].
\end{aligned}
$$

O cálculo de operadores de gradiente espacial $D_x u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}$, $D_y u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}$ e $D_z u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}$ é análogo. O campo de onda escalar auxiliar é atualizado por

$$
u^{n+\frac{1}{2}}_{k,j,i} = \Delta t p^n_{k,j,i} + u^{n-\frac{1}{2}}_{k,j,i}. \tag{A–23}
$$

Dentro do domínio principal, a atualização inteira do campo de onda simplifica para uma equação de segunda ordem:

$$
\begin{aligned}
p^{n+1}_{k,j,i} = v^2_{P|k,j,i}\frac{(\Delta t)^2}{(\Delta h)^2}&(p^n_{k,j,i-1} + p^n_{k,j,i+1} + p^n_{k,j-1,i} + p^n_{k,j+1,i} + p^n_{k-1,j,i} + p^n_{k+1,j,i} - 6p^n_{k,j,i}) \\
&+2p^n_{k,j,i} - p^{n-1}_{k,j,i} \\
+&\frac{r_{x|k,j,i}(p^n_{k,j,i+1} - p^n_{k,j,i-1}) + r_{y|k,j,i}(p^n_{k,j+1,i} - p^n_{k,j-1,i}) + r_{z|k,j,i}(p^n_{k+1,j,i} - p^n_{k-1,j,i})}{2\Delta h}
\end{aligned} \tag{A–24}
$$

B– MÉTODOS DE OTIMIZAÇÃO

Os métodos de otimização do gradiente conjugado, Newton e quasi-Newton são métodos do tipo linha, neste anexo eles são apresentados segundo a nomeclatura de Nocedal and Wright (2006).

Os métodos de otimização do tipo linha calculam em cada iteração uma direção de busca p_k e, em seguida, decide até que ponto se moverá ao longo dessa direção. A iteração é dada por

$$x_{k+1} = x_k + \alpha_k p_k, \tag{B–1}$$

onde o escalar positivo α_k é chamado de comprimento de passo. O sucesso de um método de busca de linha depende das escolhas efetivas da direção p_k e do comprimento do passo α_k. A maioria dos algoritmos de busca de linha requer que v_k seja uma direção de descida, no qual $p_k^T \nabla f_k < 0$, porque essa propriedade garante que f seja reduzido ao longo dessa direção. A direção de busca é dada:

$$p_k = -B_k^{-1} \nabla f_k, \tag{B–2}$$

onde $\mathbf{B}_k$ é uma matriz simétrica e não-singular. No método de descida mais íngreme, B_k é simplesmente a matriz de identidade I, enquanto no método de Newton, $\mathbf{B}_k$ é a Hessiana exata $\nabla^2 f(x_k)$. Nos métodos quasi-Newton, $\mathbf{B}_k$ é uma aproximação da Hessiana. Quando p_k é definido por (B–2) e $\mathbf{B}_k$ é positivo definido, temos

$$p_k^T = -\nabla f_k^T \mathbf{B}_k^{-1} \nabla f_k < 0, \tag{B–3}$$

e portanto, p_k é uma direção de descida. Para escolher α_k para dar uma redução substancial de f sem gastar muito tempo fazendo a escolha. A escolha ideal seria o minimizador global da função univariada $\phi(\cdot)$ definida por

$$\phi(\alpha) = f(x_k + \alpha p_k), \qquad \alpha > 0. \tag{B–4}$$

O comprimento do passo deve satisfazer as condições de "wolf". A primeira condição, também chamada de condição de diminuição suficiente, requer que o desajuste para um comprimento de passo escolhido seja diminuído pelo menos linearmente com o comprimento do passo. Considerando uma função genérica f(x) para representar a função objeto, x o modelo atual, α_0 o comprimento do passo e δx a atualização do modelo desejado, de modo que a primeira condição "wolf"seja satisfeita (Nocedal and Wright, 2006)

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_k^T p_k. \tag{B–5}$$

Na prática, c_1 é escolhido para ser bem pequeno, por exemplo, $c_1 = 10^{-7}$. A condição de diminuição suficiente não é capaz por si só de garantir que o algoritmo faça um progresso razoável, porque ele é satisfeito para todos os valores suficientemente pequenos de α_0. Para descartar passos inaceitavelmente curtos, é necessário um segundo requisito, chamado de condição de curvatura.

A condição de curvatura requer que a inclinação da função desajuste seja maior que a inclinação inicial. Isso garante que o próximo desajuste mínimo seja alcançado com rapidez suficiente. O segundo critério é definido como (Nocedal and Wright, 2006)

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k \tag{B–6}$$

para alguma constante $c_2 \in (c_1, 1)$, onde c_1 é a constante da condição de diminuição suficiente (B–5). Para verificar esta condição, é calculado o gradiente de todos os tiros contidos nos dados para o modelo atualizado. Então, a verificação desta condição é muito cara em termos de tempo computacional. Para economizar tempo, é possível escolher o comprimento do passo da iteração anterior o qual satisfaz a condição de "wolf".

B–0.1 convergência dos métodos de pesquisa de linha

Para obter uma convergência global, deve-se não apenas ter comprimentos de passos bem escolhidos, mas também direções de busca p_k bem escolhidas. Essa escolha deve satisfazer a relação entre o ângulo θ_k, p_k e a direção de descida mais íngreme $-\nabla f_k$, definida por

$$cos\theta_k = \frac{-\nabla f_k^T v_k}{\|\nabla f_k\| \|p_k\|}. \tag{B–7}$$

A partir da Condição de Zoutendijk, definida pela inequação:

$$\sum_k^{\infty} cos^2\theta_k \|\nabla f_k\|^2 < \infty, \tag{B–8}$$

que implica na relação:

$$cos^2\theta_k \|\nabla f_k\|^2 \|\nabla f_k\|^2 \longrightarrow 0. \tag{B–9}$$

Este limite pode ser usado, por sua vez, para obter resultados de convergência global para algoritmos de busca de linha. Se o método para escolher a direção de pesquisa v_k na iteração B–1 garantir que o ângulo θ_k definido por B–7 é limitado a partir de 90, existe uma constante positiva δ tal que

$$cos\theta_k \geq \delta > 0, \qquad Para \;\; todo \;\; k. \tag{B–10}$$

A partir de (B–9):

$$\lim_{k\to\infty} \|\nabla f_k\| = 0 \tag{B–11}$$

O que significa que as normas do gradiente $\|\nabla f_k\|$ convergem para zero, desde que as direções de busca nunca estejam muito próximas da ortogonalidade com o gradiente. Em particular, o método de descida mais íngreme, para o qual a direção de busca v_k faz um ângulo de zero graus com o gradiente negativo, produz uma seqüência de gradientes que convergem para zero, desde que use uma busca de linha satisfazendo as condições de Wolfe.

Assim, os métodos de Newton e quase-Newton são globalmente convergentes se as matrizes Bk tiverem um número de condição limitado e forem definidas positivas, o que é necessário para garantir que v_k seja uma direção de descida, e se os comprimentos de etapas passos satisfizerem as condições de Wolfe.

Para alguns algoritmos, como métodos de gradiente conjugado, não pode-se provar o limite (B–12):

$$\lim_{k\to\infty} inf\|\nabla f_k\| = 0. \tag{B–12}$$

Significa que apenas uma subsequência das normas de gradiente $\|\nabla f_k\|$ converge para zero, ao invés de toda a sequência.

B–1 MÉTODO DO GRADIENTE CONJUGADO

O método do gradiente conjugado linear é um método iterativo para a resolução de sistemas lineares com matrizes de coeficientes positivos definidos. É uma alternativa à eliminação gaussiana que é muito adequada para resolver grandes problemas. O desempenho do método do gradiente conjugado linear está ligado à distribuição dos autovalores da matriz dos coeficientes. Ao transformar, ou pré-condicionar, o sistema linear, podemos tornar essa distribuição mais favorável e melhorar a convergência do método de forma significativa. O pré-condicionamento desempenha um papel crucial no desenho de estratégias práticas de gradiente conjugado. Trata-se de um método iterativo para resolver um sistema linear de equações

$$Ax = b \tag{B–13}$$

Onde A é uma matriz $n \times n$ simétrica e positiva. O problema (B–13) pode ser declarado de forma equivalente ao problema de minimização:

$$\phi(x) = \frac{1}{2}x^T Ax - b^T x \tag{B–14}$$

O dois problemas nas Equações (B–13) e (B–14) possuem a mesma solução única. Essa equivalência permite interpretar o método do gradiente conjugado como um algoritmo para resolver sistemas lineares ou minimizar funções quadráticas convexas. Nota-se que o gradiente de ϕ é igual ao residudal do sistema linear

$$\nabla\phi(x) = Ax - b = r(x) \tag{B–15}$$

Uma das propriedades notáveis do método do gradiente conjugado é sua capacidade de gerar, de maneira muito econômica, um conjunto de vetores com uma propriedade conhecida como conjugação. Um conjunto de vetores diferentes de zero $v_0, v_1, .., v_l$ é dito ser conjugado em relação a matriz A definida positiva se

$$p_i^T A v_j = 0, \qquad \textit{para} \quad \textit{todo} \qquad i \neq j. \tag{B–16}$$

A importância da conjugação está no fato de que podemos minimizar $\phi(x)$ em n passos, minimizando-o sucessivamente ao longo das direções individuais em um conjunto de conjugados. Para Verificar esta alegação, consideramos o seguinte método de direção conjugada. Dado um ponto de partida x_0 pertencente aos reiais e um conjunto de direções conjugadas $p_0, p_1, .., p_{n1}$,

vamos gerar a sequência x_k definida:

$$x_{k+1} = x_k + \alpha_k p_k, \tag{B–17}$$

onde onde α_k é o minimizador unidimensional da função quadrática $\phi(x)$ ao longo de $x_k + \alpha_k v_k$, dado explicitamente por

$$\alpha_k = -\frac{r_k^T p_k}{p_k^T A v_k} \tag{B–18}$$

A partir das equações (B–15) e (B–17) obtemos:

$$r_{k+1} = r_k + \alpha_k A p_k \tag{B–19}$$

B–1.1 Propriedades do método do gradiente conjugado

O método do gradiente conjugado é um método de direção conjugada com uma propriedade muito especial: Ao gerar seu conjunto de vetores conjugados, ele pode calcular um novo vetor v_k usando somente o vetor anterior p_{k1}. Não precisa saber todos os elementos anteriores $v_0, v_1, .., v_{k2}$ do conjunto conjugado; v_k é automaticamente conjugado a esses vetores. Esta propriedade notável implica que o método requer pouco armazenamento e computação.

Agora, para os detalhes do método de gradiente conjugado. Cada direção v_k é escolhida para ser uma combinação linear da direção de descida mais íngreme $\nabla\phi(x_k)$ (que é o mesmo que o residual negativo r_k, por (B–15)) e a direção anterior v_{k1}. Nós escrevemos

$$p_k = -r_k + \beta_k p_{k-1}, \tag{B–20}$$

onde β_k deve ser determinado pela exigência de p_{k-1} e p_k devem ser conjugados com relação a A. Pré-multiplicando (B–20) por p_{k-1} e impondo a condição de $p_{k-1}Ap_k = 0$, achamos

$$\beta_k = \frac{r_k^T A p_{k-1}}{v_{k-1}^T A p_{k-1}}. \tag{B–21}$$

Faz sentido intuitivo escolher a primeira direção de busca p_0 para ser a direção de descida mais íngreme no ponto inicial x_0. Como no método geral da direção do conjugado, realizamos minimizações unidimensionais sucessivas ao longo de cada uma das direções de busca.

B–1.2 Precondicionamento

O precondicionamento tem por objetivo acelerar o método do gradiente conjugado através da transformação da variável x em $\hat{x}$ utilizando uma matriz não singular C, ou seja,

$$\hat{x} = Cx \tag{B–22}$$

O quadrático ϕ definido na Equação (B–14) é transformado em

$$\hat{\phi} = \frac{1}{2}\hat{x}^T(C^{-T}AC^{-1})\hat{x} - (C^{-T}b)^T\hat{x}. \tag{B–23}$$

A minimização da Equação (B–23) é realizada através de um algoritmo encontrado em Nocedal and Wright (2006) que não utiliza C expliciamente, mas sim a matriz $M = C^TC$, que é simétrica e positiva. Em termos de esforço computacional, a principal diferença entre os métodos de CG pré-condicionados e não-condicionados é a necessidade de resolver sistemas da forma $My = r$.

B–2 MÉTODOS DE NEWTON

Vamos agora considerar a iteração de Newton onde a direção de busca é dada por

$$v_k^N = -\nabla^2 f_k^{-1} \nabla^2 f_k. \tag{B–24}$$

Sabendo que para todo x na vizinhança de um ponto de solução x^* tal que $\nabla^2 f(x^*)$ é positivo definido, a Hessiana $\nabla^2 f_k$ também será positiva e definida. O método de Newton será bem definido nesta região e convergirá quadraticamente, desde que os comprimentos de passos α_k sejam sempre 1.

O passo básico de Newton v_k^N é obtido resolvendo o sistema linear $N \times N$ simétrico

$$\nabla^2 f(x_k) v_k^N = -\nabla f(x_k). \tag{B–25}$$

Para obter convergência global, é necessário que a direção de busca v_k^N seja uma direção de descida, o que será verdade se a Hessiana $\nabla^2 f(x_k)$ for positiva e definida. Se a Hessiana não obdecer essas características, ou estiver perto de ser singular, v_k^N pode ser uma direção de subida ou pode ser excessivamente longo.

B–3 MÉTODOS QUASI-NEWTON

As direções de busca Quasi-Newton fornecem uma alternativa atraente na medida em que não exigem o cálculo explícito da Hessiana e ainda assim atingem uma taxa superlinear de convergência. No lugar da Hessiana verdadeira $\nabla^2 f_k$, é utilizada uma aproximação $\mathbf{B}_k$, que é atualizada após cada passo para levar em consideração o conhecimento adicional obtido durante o passo. As atualizações fazem uso do fato de que mudanças no gradiente g fornecem informações sobre a segunda derivada de f ao longo da direção de busca. Assim, a direção da busca tenha a forma:

$$p_k = -\mathbf{B}_k^{-1} \nabla f_k, \tag{B–26}$$

onde a matriz simétrica e positiva definida B_k é atualizada a cada iteração. Considerando f a função objeto, os métodos quasi-Newton axigem apenas o gradiente da função objeto fornecido em cada iteração. Ao medir as mudanças nos gradientes, eles constroem um modelo da função objeto que é bom o suficiente para produzir a convergência. Como as derivadas secundárias não são necessárias, os métodos quasi-Newton são algumas vezes mais eficientes do que o método de Newton.

B–3.1 Método BFGS

Considerando o modelo quadrático da função objetivo na iteração atual x_k:

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T \mathbf{B}_k p. \tag{B–27}$$

Aqui $\mathbf{B}_k$ é uma matriz definida positiva simétrica $N \times N$ que será revisada ou atualizada a cada iteração. Note que o valor e o gradiente deste modelo em p_0 correspondem f_k e ∇f_k, respectivamente. O minimizador $\mathbf{p}_k$ deste modelo quadrático convexo, na forma

$$\mathbf{p}_k = -\mathbf{B}_k^{-1} \nabla f_k, \tag{B–28}$$

é usado como direção de procura, e a nova iretação é

$$x_{k+1} = x_k + \alpha_k v_k, \tag{B–29}$$

onde o comprimento de passo é escolhido satisfazendo a condição de Wolf. Esta iteração é bastante semelhante ao método Newton de pesquisa de linha; a principal diferença é que o Hessiana aproximada $\mathbf{B}_k$ é usado no lugar da verdadeira Hessiana. Em vez de calcular $\mathbf{B}_k$ novamente a cada iteração,é proposto atualizá-lo de uma maneira simples para explicar a curvatura medida durante a etapa mais recente. Suponha a nova iteração x_{k+1} e deseja-se construir um novo modelo quadrático, da forma

$$m_{k+1}(p) = f_{k+1} + \nabla f_{k+1}^T p + \frac{1}{2} p^T \mathbf{B}_{k+1} p. \tag{B–30}$$

Impondo a condição:

$$\nabla \mathbf{m}_{k+1}(-\alpha_k p_k) = \nabla f_{k+1} - \alpha_k \mathbf{B}_{k+1} p_k = \nabla f_k. \tag{B–31}$$

Ao rearranjar, obtemos:

$$\mathbf{B}_{k+1} \alpha_k p_k = \nabla f_{k+1} - \nabla f_k. \tag{B–32}$$

Para simplificar a notação, defini-se os vetores:

$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k, \qquad \mathbf{y}_k = \nabla f_{k+1} - \nabla f_k, \tag{B–33}$$

tornando a Equação (B–32):

$$\mathbf{B}_{k+1} s_k = y_k. \tag{B–34}$$

Dado o deslocamento s_k e a mudança de gradientes y_k, a Equação (B–34) requer que a matriz definida positiva simétrica $\mathbf{B}_{k+1}$ mapeie s_k em y_k. Isso só será possível se s_k e y_k satisfizerem a condição de curvatura

$$s_k^T y_k > 0. \tag{B–35}$$

Para determinar $\mathbf{B}_{k+1}$ unicamente, então, utiliza-se a condição adicional de que entre todas as matrizes simétricas que satisfazem a equação (B–34), $\mathbf{B}_{k+1}$ é o mais próximo da matriz atual $\mathbf{B}_k$. Em outras palavras, trata-se do problema:

$$\begin{aligned} &min \parallel \mathbf{B} - \mathbf{B}_k \parallel \\ sujeito \quad a \quad \mathbf{B} = \mathbf{B}^T, \qquad &\mathbf{B} s_k = y_k \end{aligned} \tag{B–36}$$

onde s_k e y_k satisfazem a condição de curvatira e $\mathbf{B}_k$ é simétrica e positiva definida. Muitas normas matriciais podem ser usadas em (B–36), e cada norma dá origem a um método quase-Newton diferente. A atualização do BFGS pode ser obtida impondo condições ao inverso de $\mathbf{B}_k$, ou seja, $\mathbf{H}_k$. A aproximação atualizada $\mathbf{H}_{k+1}$ deve ser simétrica definida positiva, e deve satisfazer a equação (B–34), agora escrita como

$$\mathbf{H}_{k+1} y_k = s_k \tag{B–37}$$

Aplicando a condição análoga a (B–36):

$$\begin{aligned} & min \parallel \mathbf{H} - \mathbf{H}_k \parallel \\ sujeito \quad a \quad & \mathbf{H} = \mathbf{H}^T, \qquad \mathbf{H} s_k = y_k \end{aligned} \tag{B–38}$$

A solução única $\mathbf{H}_{k+1}$ para (B–38) é dada por:

$$\mathbf{H}_{k+1} = (Id_k s_k y_k^T)\mathbf{H}_k(Id_k y_k s_k^T) + d_k s_k s_k^T, \tag{B–39}$$

onde

$$d_k = \frac{1}{y_k^T s_k}. \tag{B–40}$$

A versão do método BFGS que trabalha com a aproximação da Hessiana $\mathbf{B}_k$ ao invés de $\mathbf{H}_k$ é dada:

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k s_k s_k^T \mathbf{B}_k}{s_k^T \mathbf{B}_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \tag{B–41}$$

B–3.2 Método L-BFGS

Os métodos quasi-Newton de memória limitada são úteis para resolver grandes problemas cujas matrizes da Hessiana não podem ser calculadas a um custo razoável ou são muito densas para serem facilmente manipuladas. Esses métodos mantêm aproximações simples e compactas de matrizes Hessianas: em vez de armazenar aproximações totalmente densas $N \times N$, elas salvam apenas alguns vetores de comprimento N que representam as aproximações implicitamente. Apesar desses requisitos modestos de armazenamento, eles geralmente produzem uma taxa de convergência aceitável. A idéia principal deste método é usar informações de curvatura somente das iterações mais recentes para construir a aproximação da Hessiana. A informação de curvatura de iterações anteriores, que é menos provável de ser relevante para o comportamento real da Hessiana na iteração atual, é descartada no interesse de economizar armazenamento. Para explicar o método L-BFGS, consideramos inicialmente o método BFGS com cada passo na forma

$$x_{k+1} = x_k - \alpha_k \mathbf{H}_k \nabla f_k, \qquad k = 0, 1, 2, .., \tag{B–42}$$

onde α_k é o comprimento do passo, e $\mathbf{H}_k$ é atualizado a cada iteração por meio da fórmula

$$\mathbf{H}_{k+1} = V_k^T \mathbf{H}_k V_k + d_k s_k s_k^T, \tag{B–43}$$

onde

$$d_k = \frac{1}{y_k^T s_k}, \qquad V_k = I - d_k y_k s_k^T, \tag{B–44}$$

e

$$s_k = x_{k+1} - x_k \qquad y_k = \nabla f_{k+1} - \nabla f_k. \tag{B–45}$$

Dizemos que a matriz $\mathbf{H}_{k+1}$ é obtida atualizando $\mathbf{H}_k$ usando o par s_k, y_k.

A aproximação Hessiana inversa geralmente será densa, de modo que o custo de armazená-la e manipulá-la é limitado quando o número de variáveis for grande. Para contornar este problema, uma versão modificada de H_k é armazenada implicitamente, armazenando um certo número (digamos m) dos pares de vetores s i, y i que são usados nas fórmulas (B–43) e (B–45). O produto $\mathbf{H}_k \nabla f_k$ pode ser obtido através da execução de uma sequência de produtos internos e de somatórios de vetores envolvendo ∇f_k e os pares s_i, y_i. Depois que a nova iteração é calculada, o par vetorial mais antigo no conjunto de pares s_i, y_i é deletado e substituído pelo novo par s_k, y_k obtido da etapa atual (B–45). Dessa forma, o conjunto de pares de vetores inclui informações de curvatura das mais recentes iterações.