



DIVERGÊNCIA DE KULLBACK-LEIBLER APLICADA A FULL WAVEFORM INVERSION(FWI)

Gilberto Barbosa Neto Carvalho

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Produção, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Produção.

Orientadores: Juan Pablo

Virgílio José Martins Ferreira
Filho

Rio de Janeiro
Setembro de 2023

DIVERGÊNCIA DE KULLBACK-LEIBLER APLICADA A FULL WAVEFORM
INVERSION(FWI)

Gilberto Barbosa Neto Carvalho

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE
ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO
PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU
DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE PRODUÇÃO.

Orientadores: Juan Pablo

Virgílio José Martins Ferreira Filho

Aprovada por: Prof. Nome do Primeiro Examinador Sobrenome

Prof. Nome do Segundo Examinador Sobrenome

RIO DE JANEIRO, RJ – BRASIL

SETEMBRO DE 2023

Barbosa Neto Carvalho, Gilberto

Divergência de Kullback-Leibler aplicada a Full
Waveform Inversion(FWI) /Gilberto Barbosa Neto
Carvalho. – Rio de Janeiro: UFRJ/COPPE, 2023.

XI, 64 p.: il.; 29, 7cm.

Orientadores: Juan Pablo

Virgílio José Martins Ferreira Filho

Dissertação (mestrado) – UFRJ/COPPE/Programa de
Engenharia de Produção, 2023.

Referências Bibliográficas: p. 59 – 64.

1. Divergência de Kullback-Leibler. 2. FWI. 3.
Regularização. I. Pablo, Juan *et al.* II. Universidade
Federal do Rio de Janeiro, COPPE, Programa de
Engenharia de Produção. III. Título.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

DIVERGÊNCIA DE KULLBACK-LEIBLER APLICADA A FULL WAVEFORM INVERSION(FWI)

Gilberto Barbosa Neto Carvalho

Setembro/2023

Orientadores: Juan Pablo

Virgílio José Martins Ferreira Filho

Programa: Engenharia de Produção

A Inversão Completa de Forma de Onda (FWI, do inglês Full-Waveform Inversion) é um dos métodos mais robustos para extrair informações sísmicas. No entanto, a utilização da norma L2 para medir a discrepância entre os dados sísmicos nem sempre se mostra como a melhor opção, devido à sua alta sensibilidade a erros nos dados. Desta forma, este trabalho propõe a utilização da Divergência de Kullback-Leibler como uma alternativa à norma L2 para medir a diferença entre os dados observados e os dados modelados. Para atingir esse objetivo, foi desenvolvida uma metodologia para adaptar a função de divergência Kullback-Leibler à FWI, visto que inicialmente ela foi concebida para resolver problemas em um contexto estatístico. Para tanto, foi aplicada a parte positiva da função para lidar com possíveis valores negativos, aliada a métodos de suavização para tornar a função parte positiva diferenciável. Com o intuito de comparar os resultados, foram realizados cinco testes em diferentes contextos. A função de divergência de Kullback-Leibler apresentou um desempenho notável na maioria dos testes, indicando que pode ser uma ferramenta valiosa na resolução do problema de FWI. Assim, este trabalho oferece uma contribuição significativa ao desenvolver uma metodologia que integra a divergência de Kullback-Leibler à técnica de Inversão Completa de Forma de Onda (FWI). **Palavras-Chaves:** Divergência de Kullback-Leibler, FWI, Regularização

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

KULLBACK-LEIBLER DIVERGENCE APPLIED TO FULL WAVEFORM INVERSION (FWI)

Gilberto Barbosa Neto Carvalho

September/2023

Advisors: Juan Pablo

Virgílio José Martins Ferreira Filho

Department: Production Engineering

Full-Waveform Inversion (FWI) is one of the most robust methods for extracting seismic information. However, the use of the L2 norm to measure the discrepancy between seismic data is not always the best choice due to its high sensitivity to data errors. Therefore, this study proposes the use of Kullback-Leibler Divergence as an alternative to the L2 norm to measure the difference between observed and modeled data. To achieve this goal, a methodology was developed to adapt the Kullback-Leibler Divergence function to FWI, as it was initially designed to solve problems in a statistical context. The positive part of the function was applied to handle potential negative values, along with smoothing methods to make the positive part of the function differentiable. To compare the results, five tests were conducted in different contexts. The Kullback-Leibler Divergence function showed remarkable performance in most of the tests, indicating that it can be a valuable tool in solving the FWI problem. Thus, this work provides a significant contribution by developing a methodology that integrates the Kullback-Leibler divergence into Full Waveform Inversion (FWI) technique.

Key-Words: Kullback-Leibler, Divergence, FWI, Regularization

Sumário

Lista de Figuras	viii
Lista de Tabelas	xi
1 Introdução	1
1.1 Divergência de Kullback-Leibler	3
2 Inversão da forma completa da onda	5
2.1 O método sísmico	5
2.2 Modelagem Matemática	7
2.3 O problema Direto	9
2.3.1 Discretização Espacial	10
2.3.2 Discretização Temporal	11
2.4 O Problema Inverso	12
2.5 Método do Estado-adjunto	14
2.5.1 Condição de Imagem	14
2.5.2 Cálculo de gradiente da função de discrepância	17
2.6 Algoritmos	18
2.6.1 LBFGS	19
2.6.2 Busca Linear	20
3 Divergência	23
3.1 Convergência	24
3.2 Divergência de Kullback Leibler	25
3.2.1 Propriedades	27
3.3 Suavização da Função Parte Positiva	28
3.3.1 Propriedades da função parte positiva	29
4 Divergência aplicado ao FWI	32
4.1 O problema de desbalanceamento e valores negativos nos dados sísmicos	32
4.2 Discrepância entre dados sísmicos	34
4.2.1 Divergência Kullback Leibler	35

5 Experimentos	36
5.1 Modelo da bola pequeno	37
5.1.1 Resultados	37
5.2 Modelo da bola grande	41
5.2.1 Resultados	41
5.3 Modelo Petro	44
5.3.1 Resultados	46
5.4 Modelo Marmousi suavizado	48
5.4.1 Resultados	50
5.5 Fontes simultâneas	52
5.5.1 Resultados	52
5.6 Discussão	56
6 Conclusão	57
Referências Bibliográficas	59

Lista de Figuras

2.1	Processo de coleta de dados em ambientes marinhos. Fonte: Geology Learn Blog	6
2.2	impedância acústica . Fonte: maxwell	7
2.3	Esquema do problema inverso resolvido como um processo de otimização Fonte: Elaborada pelo autor	13
2.4	Condição de diminuição suficiente.	22
3.1	gráfico da função $f_c = x \log x + 1 - x$ Fonte: Elaborada pelo autor . . .	26
3.2	gráfico da função $\tilde{f}_c = \log \frac{1}{x} + x - 1$ Fonte: Elaborada pelo autor . . .	27
5.1	função Chen-Harker-Kanzow-Samle Smoothing	36
5.2	Modelo inicial	37
5.3	Modelo Final	37
5.4	Resultado com a norma L2	38
5.5	Resultado com Kl	38
5.6	Perfil de velocidade na posição $x = 100$ metros (KL)	38
5.7	Perfil de velocidade na posição $x = 100$ metros (L2)	38
5.8	Perfil de velocidade na posição $x = 125$ metros (KL)	38
5.9	Perfil de velocidade na posição $x = 125$ metros (L2)	38
5.10	Perfil de velocidade na posição $x = 150$ metros (KL)	39
5.11	Perfil de velocidade na posição $x = 150$ metros (L2)	39
5.12	Erro relativo $x = 100$ metros	39
5.13	Erro relativo na posição $x = 125$ metro	39
5.14	Erro relativo na posição $x = 150$ metros	39
5.15	Evolução da função objetiva L2	40
5.16	Evolução da função objetiva Kl	40
5.17	Modelo inicial	41
5.18	Modelo Final	41
5.19	Resultado com a norma L2	42

5.20	Resultado com Kl	42
5.21	Perfil de velocidade na posição x=110 metros (KL)	43
5.22	Perfil de velocidade na posição x=110 metros (L2)	43
5.23	Perfil de velocidade na posição x=125 metros (KL)	43
5.24	Perfil de velocidade na posição x=125 metros (L2)	43
5.25	Perfil de velocidade na posição x=150 metros (KL)	43
5.26	Perfil de velocidade na posição x=150 metros (L2)	43
5.27	Erro relativo x=110 metros	44
5.28	Erro relativo na posição x=125 metros	44
5.29	Erro relativo na posição x=150 metros	44
5.30	Evolução da função objetiva L2	44
5.31	Evolução da função objetiva Kl	44
5.32	Modelo inicial	45
5.33	Modelo Final	45
5.34	Resultado com a norma L2	45
5.35	Resultado com Kl	45
5.36	Perfil de velocidade na posição x=87 metros (Kl)	46
5.37	Perfil de velocidade na posição x=87 metros (L2)	46
5.38	Perfil de velocidade na posição x=118 metros (Kl)	46
5.39	Perfil de velocidade na posição x=118 metros (L2)	46
5.40	Perfil de velocidade na posição x=241 metros (Kl)	46
5.41	Perfil de velocidade na posição x=241 metros (L2)	46
5.42	Erro relativo na posição x=87 metros	47
5.43	Erro relativo na posição x=118 metros	47
5.44	Erro relativo na posição x=241 metros	47
5.45	Evolução da função objetiva L2	47
5.46	Evolução da função objetiva Kl	47
5.47	Modelo inicial	49
5.48	Modelo Final	49
5.49	Resultado com a norma L2	49
5.50	Resultado com a função KL	49
5.51	Perfil de velocidade na posição x =153 metros (L2)	50
5.52	Perfil de velocidade na posição x =153 metros (Kl)	50
5.53	Perfil de velocidade na posição x =209 metros (L2)	50
5.54	Perfil de velocidade na posição x =209 metros (kl)	50
5.55	Perfil de velocidade na posição x =426 metros (L2)	50
5.56	Perfil de velocidade na posição x =426 metros (kl)	50
5.57	Erro relativo x = 153 metros	51
5.58	Erro relativo na posição x = 209 metros	51

5.59	Erro relativo na posição $x = 426$ metros	51
5.60	Evolução da função objetiva L2	51
5.61	Evolução da função objetiva Kl	51
5.62	Modelo inicial	53
5.63	Modelo Final	53
5.64	Resultado com a norma L2	53
5.65	Resultado com Kl	53
5.66	Perfil de velocidade na posição $x = 100$ metros (L2)	53
5.67	Perfil de velocidade na posição $x = 100$ metros (Kl)	53
5.68	Perfil de velocidade na posição $x = 125$ metros (L2)	54
5.69	Perfil de velocidade na posição $x = 125$ metros (kl)	54
5.70	Perfil de velocidade na posição $x=150$ metros (L2)	54
5.71	Perfil de velocidade na posição $x=150$ metros (kl)	54
5.72	Erro relativo $x = 100$ metros	55
5.73	Erro relativo na posição $x = 125$ metros	55
5.74	Erro relativo na posição $x = 150$ metros	55
5.75	Evolução da função objetiva L2	55
5.76	Evolução da função objetiva Kl	55

Lista de Tabelas

5.1	Tabela de erro relativo dos resultados	41
5.2	Tabela de erro relativo dos resultados	45
5.3	Tabela de erro relativo dos resultados	48
5.4	Tabela de erro relativo dos resultados	49
5.5	Tabela de erro relativo dos resultados	55

Capítulo 1

Introdução

Ao longo dos anos, a ciência da terra tem desenvolvido ferramentas mais precisas para estudar as propriedades da subsuperfície da Terra, impulsionadas pelo avanço da computação de alto desempenho, sofisticação dos sensores que medem ondas sísmicas e inovações em técnicas matemáticas, incluindo otimização de equações diferenciais parciais (EDP) e processamento avançado de sinais (YANG, 2018) .

Uma das principais áreas da geofísica é o estudo das ondas sísmicas, que fornecem informações sobre a subsuperfície por onde se propagam. Essa ferramenta é fundamental na geofísica de exploração para obter informações sobre estruturas geofísicas, como reservatórios de petróleo e gás (KEAREY *et al.*, 2002). Após a obtenção desses dados, é realizada uma modelagem matemática com base em propriedades físicas para obter os dados observados (KEAREY *et al.*, 2002). Dentre os métodos de processamento sísmico, a Migração Reversa no Tempo (RTM - Reverse Time Migration) e a Inversão da Forma de Onda Completa (FWI - Full Waveform Inversion) são os mais utilizados (YANG, 2018).

As técnicas de migração permitem identificar detalhes das subsuperfícies refletoras, como falhas e camadas sedimentares, a partir dos dados medidos e de uma estimativa aproximada das propriedades geofísicas (KEAREY *et al.*, 2002) . Essa aplicação pode ser realizada no domínio do tempo ou da frequência. Por outro lado, a FWI é uma técnica de alta resolução amplamente utilizada na indústria e na academia. Seu objetivo é identificar os componentes de pequena e grande escala que descrevem as propriedades físicas das formas de onda medidas nos receptores sísmicos, também conhecidos como traços sísmicos (YANG, 2018). O diferencial da FWI é sua capacidade de inverter ondas de diferentes comprimentos simultaneamente, preenchendo a lacuna entre regimes de transmissão e refletividade, superando as limitações do fluxo de trabalho de imageamento sísmico convencional, que envolve etapas separadas de tomografia e migração.

O processo de inversão de ondas começa com um modelo inicial obtido por modelagem direta. A partir desse modelo, os campos de onda modelados são previstos

na posição dos receptores. Os resíduos dos dados são então calculados pela diferença entre os dados calculados e observados, usando uma função de custo. Esse problema é formulado como uma otimização de EDP, onde a função de custo convencional é a norma L2 (VIRIEUX *et al.*, 2017):

$$F_D = \min J(d_v, d_0) = \frac{1}{2} \|d_v - d_0\|^2 \quad (1.1)$$

Onde d_v e d_0 são dados modelados e observado, respectivamente. O próximo passo envolve calcular o gradiente da função objetivo, que representa a derivada da função em relação aos parâmetros do modelo. Esse gradiente é usado para atualizar iterativamente o modelo de velocidade, usando um método de busca na direção descendente apropriada. Esse processo de atualização é repetido até que um critério de parada seja atingido.

Espera-se que à medida que a discrepância entre d_v e d_0 diminua, o modelo inicial se aproximará do modelo verdadeiro. Para resolver esse problema de otimização, são utilizadas diversas áreas da matemática. O primeiro passo é discretizar o problema contínuo, frequentemente usando métodos numéricos como diferenças finitas e elementos finitos. No entanto, essa discretização e a solução iterativa levam a um problema de grande escala, exigindo considerável poder computacional (YANG, 2018). Lidar com inversões de problemas de grande escala requer métodos de otimização numérica robustos e eficientes. Além disso, a modelagem direta deve ser eficiente para ser uma representação precisa das complexas ondas físicas (CARNEIRO *et al.*, 2018).

A norma L2, reconhecida por sua robustez em uma variedade de casos práticos, apresenta qualidades distintas, incluindo sensibilidade às amplitudes dos dados, preservação eficaz de informações espaciais, robustez diante de outliers e facilidade tanto no cálculo quanto na interpretação. Entretanto, no contexto específico de medição de discrepâncias em dados sísmicos no FWI, a norma L2 revela uma sensibilidade ao ruído. Esta característica pode resultar em soluções que tendem a convergir para pontos locais, especialmente em situações onde há presença de ruído nos dados ou quando o modelo inicial está significativamente distante do modelo real (VIRIEUX e OPERTO, 2009).

Para mitigar problemas como mínimos locais e não-linearidade, foram desenvolvidas várias estratégias multiescala hierárquicas para suavizar a não-linearidade e a má convergência da FWI. Essa abordagem começa a minimização a partir de ondas de maior duração no espaço de parâmetros e avança progressivamente. Isso ocorre porque frequências mais baixas contêm mais informações sobre a estrutura e são menos sensíveis. Essa estratégia multiescala resolve subgrupos de dados com crescente

poder para incluir mais ondas no modelo tomográfico, auxiliando na convergência em direção ao mínimo global (BUNKS *et al.*, 1995).

Diante da inconsistência da norma L2, várias abordagens alternativas surgiram, como a adoção da norma L1 (CRASE *et al.*, 1990), a norma L2 suavizada (ASNA-ASHARI *et al.*, 2013), e até mesmo modelos mais sofisticados, como o transporte ótimo ((CHEN e PETER, 2018), (YANG *et al.*, 2018), (CHIZAT *et al.*, 2018)) e o método bayesiano (ZHU *et al.*, 2016). Neste trabalho, está sendo proposto o uso da divergência Kullback-Leibler como função de discrepância para medir o resíduo entre os dados sísmicos. Uma das vantagens em utilizar a divergência de Kullback-Leibler em vez da norma L2 é que a operação logarítmica atenua as variações de baixa intensidade. Assim, em alguns cenários, o método da divergência KL pode proporcionar imagens de melhor qualidade (MUNIZ *et al.*, 2000).

1.1 Divergência de Kullback-Leibler

O princípio da entropia relativa mínima (PERM) foi inicialmente proposto por (KULLBACK, 1959) como um método de inferência estatística. A entropia relativa também é conhecida como a divergência de Kullback-Leibler em homenagem ao Solomon Kullback e Richard Leibler pelo trabalho seminal em (KULLBACK e LEIBLER, 1951). Anteriormente, (JAYNES, 1957) havia introduzido o princípio da entropia máxima, que incorporou o trabalho de (SHANNON, 1948) na teoria da informação e o aplicou a problemas na mecânica estatística. Desde a sua introdução, esses princípios têm encontrado aplicações em uma ampla variedade de campos.

No âmbito da geofísica, técnicas conhecidas como análise espectral de alta resolução (HAYKIN, 2006), modelagem autorregressiva [(ULRYCH e BISHOP, 1975) e convolução preditiva (ROBINSON e TREITEL, 1980)] estão intrinsecamente relacionadas ao PME. Esse princípio também foi empregado por (DONG *et al.*, 1984) para tratar da relação entre magnitude e frequência em terremotos, bem como por (RIETSCH, 1988) na inversão de sismogramas unidimensionais. As aplicações na área de reconstrução de imagens têm obtido resultados notáveis (GULL *et al.*, 1984).

Em termos de função objetivo para quantificar a discrepância entre os dados observados e os dados calculados, a divergência de Kullback-Leibler tem sido utilizada principalmente como uma função auxiliar para regularizar os dados e incorporar informações adicionais. Nos artigos ((CRUZ *et al.*, 2019) e (CRUZ *et al.*, 2021)), os autores exploraram algumas características da entropia relativa e propuseram três abordagens distintas para incorporar informações por meio dela no problema inverso. Os pesquisadores chegaram a conclusão de que a informação prévia incorporada pela adição da entropia relativa ao FWI, pode resultar em estimativas com

melhor resolução.

No estudo de (DA SILVA *et al.*, 2019), seguiram uma abordagem semelhante, com a diferença de que os autores utilizaram a entropia de Shannon. A entropia relativa também é amplamente empregada para regularizar os problemas de transporte ótimo aplicados a FWI ((CHEN e PETER, 2018),(LI *et al.*, 2022), (YU, 2021)).

Este trabalho se distingue dos demais mencionados acima pelo fato de adotar a função de divergência de Kullback-Leibler como elemento central na formulação da função objetivo para calcular a discrepância entre os dados observados e os dados calculados. Normalmente, o uso de entropia ocorre em um contexto de distribuição de probabilidade ou métodos bayesianos, o que implica em etapas adicionais, como a definição de distribuições posteriores e anteriores (LANTÉRI, 2020).

Ao contrário dos casos em contextos estatísticos, a divergência não está sendo aplicada para quantificar a diferença entre dois conjuntos de dados, mas sim para minimizar a discrepância entre os dados observados e os dados calculados em relação aos parâmetros desconhecidos. Para contornar a limitação da função de divergência, que não aceita valores negativos, foi implementada a função parte positiva. Essa abordagem garante que todos os valores sejam não negativos, possibilitando uma análise precisa e coerente. No entanto, é importante notar que a função parte positiva não é diferenciável. Diante disso, foi necessário incorporar técnicas de suavização para torná-la diferenciável.

Esta dissertação está dividida em seis partes, além desta introdução. A segunda parte oferece uma revisão abrangente da FWI, abordando tanto os métodos diretos quanto os indiretos, bem como o algoritmo subjacente. A terceira parte introduz a função Kullback-Leibler, explora suas propriedades e detalha o método de suavização adotado neste estudo. Na quinta parte, mostra as técnicas utilizadas para incorporar a função Kullback-Leibler ao FWI. Os resultados numéricos são apresentados na sexta parte, seguidos da conclusão e das considerações finais no encerramento deste trabalho.

Capítulo 2

Inversão da forma completa da onda

O petróleo, cujo termo provém do latim "petra oleum"(pedra e óleo), era conhecido desde os tempos da antiga Babilônia e passou pelas mãos de fenícios e egípcios. No entanto, foi somente em 1859, com a descoberta de petróleo pelo coronel Drake, a uma profundidade de apenas 21 metros da superfície em sua propriedade no estado da Pensilvânia, que se deu início à exploração comercial desse recurso (THOMAS, 2001). Naquela época, os hidrocarbonetos demonstraram ser um substituto vantajoso para carvão, querosene e óleo de baleia. Contudo, a verdadeira revolução ocorreu após a invenção dos motores a gasolina e diesel.

Apesar da crescente busca por fontes de energia alternativa, o petróleo ainda representa cerca de 31% da matriz energética global (IEA). A indústria petrolífera tem impulsionado a pesquisa e o desenvolvimento de técnicas de exploração sísmica. Além disso, os avanços em técnicas e materiais, principalmente o aço, têm possibilitado a perfuração de poços com profundidades de até 10.000 metros (VASCONCELOS, 2018).

2.1 O método sísmico

A geofísica de exploração tem como objetivo revelar as estruturas ocultas sob a superfície terrestre, o que pode ser alcançado tanto de maneira direta quanto indireta. A abordagem direta envolve acessar a subsuperfície por meio de perfurações em rochas. No entanto, muitas vezes esse método não é viável devido ao alto custo ou à falta de acesso ao local. Por outro lado, o método indireto busca inferir as propriedades da subsuperfície por meio da interpretação de medições indiretas. Além de ser mais econômico, o uso de métodos sísmicos é uma maneira poderosa de explorar a subsuperfície em profundidades maiores, empregando técnicas avançadas, como a geração artificial de campos de ondas (KEAREY *et al.*, 2002).

A exploração das estruturas da subsuperfície por meio de métodos sísmicos começa pela geração de ondas a partir de fontes artificiais. Essas ondas se propagam

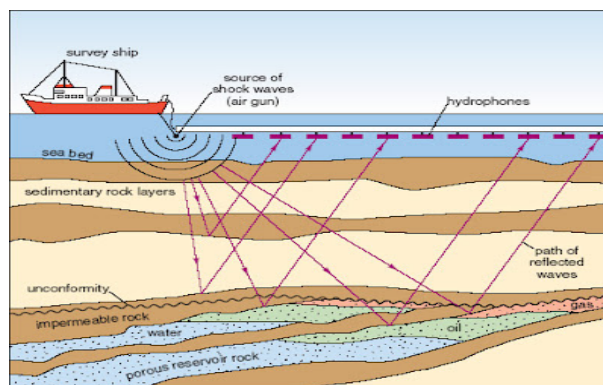


Figura 2.1: Processo de coleta de dados em ambientes marinhos.

Fonte: Geology Learn Blog

pelo interior da Terra, refletindo nas interfaces entre rochas com diferentes características petrofísicas. Essas ondas refletidas retornam à superfície e são captadas por sensores. Esses métodos sísmicos são usados para mapear sequências de camadas sedimentares (THOMAS, 2001). O objetivo final é obter imagens das estruturas, camadas geológicas na subsuperfície de várias formas e algumas propriedades físicas (VASCONCELOS, 2018). O resumo do processo de aquisição de dados no alto-mar pode ser visto na figura 2.1.

Os métodos sísmicos podem ser divididos em duas categorias: refração e reflexão. O primeiro registra apenas as ondas refratadas com ângulos críticos, tendo sido pioneiro no estudo das profundezas terrestres a partir dos anos 50 (THOMAS, 2001). Já o método de reflexão baseia-se no registro das ondas refletidas ao chegarem ao receptor. Esse método de reflexão é particularmente vantajoso, pois oferece alta definição das características geológicas subterrâneas, sendo útil para identificar acúmulos de hidrocarbonetos a um custo relativamente baixo (KEAREY *et al.*, 2002).

As ondas sísmicas são criadas a partir de uma fonte externa de energia e se propagam através da subsuperfície. Ao longo da propagação no meio, essas ondas passam por reflexão, difração, transmissão e conversão, o que provoca perturbações mecânicas ao retornarem à superfície (KEAREY *et al.*, 2002). Esses movimentos são capturados por sensores espalhados pelo campo, que registram o tempo de trânsito em diferentes alcances e a amplitude das ondas refletidas e difratadas. Os registros de tempo podem ser convertidos em valores de profundidade, possibilitando o mapeamento sistemático da distribuição das interfaces geológicas internas da subsuperfície (THOMAS, 2001). Após o processamento desses dados, o resultado consiste em seções ou volumes migrados, em profundidade, que representam um mapa da distribuição do contraste de impedância do meio ou da refletividade.

A impedância acústica é o produto entre a densidade e a velocidade do meio, enquanto a refletividade é a função de contraste da rocha. À medida que essas ondas se propagam e refratam no meio, captam informações valiosas da estrutura que retornam à subsuperfície (THOMAS, 2001). Para a geração das ondas, as fontes usuais incluem dinamites, vibradores e canhões de ar comprimido em levantamentos marítimos. Quanto aos receptores, os dois tipos mais comuns para registrar reflexões são os eletromagnéticos e geofones e hidrofones (VASCONCELOS, 2018). A figura 2.2 ilustra o processo de cálculo de impedância e refletividade,

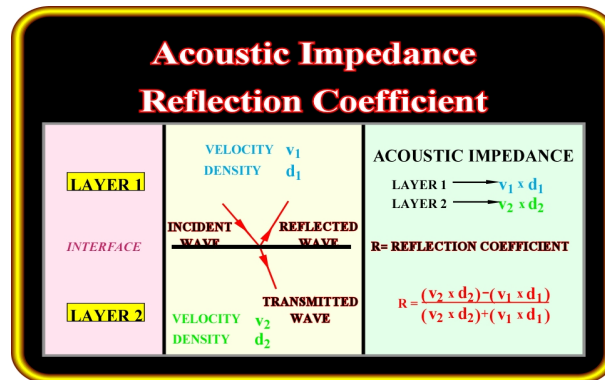


Figura 2.2: impedância acústica .
Fonte: maxwell

Todos os parâmetros, como profundidade e posição dos receptores, são configurados para obter a melhor imagem possível, respeitando as restrições orçamentárias (THOMAS, 2001).

2.2 Modelagem Matemática

A etapa de modelagem desempenha um papel crucial na obtenção de imagens sísmicas. Após a coleta dos dados, é necessário encontrar uma maneira de estabelecer relações entre as variáveis envolvidas. O desenvolvimento do modelo e a definição dos parâmetros dependem das características do meio, da fonte de energia e de outros fatores físicos relevantes (KEAREY *et al.*, 2002).

Existem diversas abordagens para simular a propagação das ondas sísmicas, que abrangem desde a consideração de parâmetros anisotrópicos e densidade até fatores de atenuação. Estas equações descrevem a propagação e o comportamento das ondas sísmicas tanto em meios viscosos quanto elásticos (CAMARGO, 2019).

Por outro lado, a modelagem acústica é amplamente utilizada para atualizar as ondas de compressão (onda-P), deixando de lado os parâmetros elásticos. Essa abordagem tem se mostrado altamente eficaz em termos de relação custo-benefício,

considerando tanto a qualidade das informações obtidas quanto os recursos computacionais demandados (VIRIEUX e OPERTO, 2009). Ela é frequentemente escolhida para testar novos modelos, o que justifica sua aplicação neste trabalho.

Considerando um campos de ondas $u : \Omega \times [0, t] \rightarrow R$, no qual $\Omega \subset R^2$ é domínio espacial definido através de um campo de velocidade $v : \Omega \rightarrow R$, ao longo de um período de tempo $[0, t]$, proveniente de uma função fonte $S : \Omega \times [0, t] \rightarrow R$. A equação da onda acústica, em um meio homogêneo, considerado densidade constante, é dada por:

$$\frac{1}{v^2} \frac{\partial^2 u(x, t)}{\partial t^2} = \Delta u(x, t) + S(x, t),$$

junto com as condições de contorno:

$$\begin{aligned} u(x, 0) &= 0, \quad \forall x \in \Omega \\ u_t(x, 0) &= 0, \quad \forall x \in \Omega \\ u(x, t) &= 0, \quad \forall x \in \partial\Omega, \forall t \in [0, t] \end{aligned} \tag{2.1}$$

onde :

$$\Delta = \left(\frac{\partial^2}{\partial x_1^2}, \frac{\partial^2}{\partial x_2^2}, \dots, \frac{\partial^2}{\partial x_n^2} \right)$$

Nessa equação e nas condições de contorno, todas as derivadas estão sendo expressas em relação às variáveis x e t . A equação da onda descreve a propagação da onda acústica em um meio físico, onde a velocidade de propagação v é influenciada pela função fonte S que introduz a perturbação no meio. As condições de contorno estabelecem comportamentos específicos nas bordas do domínio espacial.

A equação 2.1 possui a solução analítica para o caso homogêneo. No entanto, em situações mais complexas, como ambientes não homogêneos, são necessários métodos de simulação de propagação de ondas. Esses métodos podem ser agrupados em três categorias principais: métodos diretos, métodos integrais e métodos assintóticos (CARCIONE *et al.*, 2002).

Nos métodos diretos, o modelo geológico é aproximado por uma malha numérica, ou seja, o modelo é discretizado em um número finito de pontos. Isso permite a solução da equação da onda para cada ponto da malha, considerando as propriedades do meio em cada posição.

Os métodos integrais são baseados em representações integrais do campo de ondas a partir de fontes pontuais. Isso significa que o campo de ondas é calculado

a partir das contribuições de várias fontes pontuais, o que pode ser mais eficiente em certos casos.

Por sua vez, os métodos assintóticos, como por exemplo o traçado de raios, podem ser úteis para resolver problemas em que é possível fazer aproximações em certas direções de propagação. No traçado de raios, o campo de ondas é considerado como um conjunto de eventos, decompondo-o em duas equações: a equação iconal e a equação de transporte (CAMARGO, 2019). Esses métodos são úteis em situações onde é possível considerar aproximações em determinadas direções de propagação. No contexto deste trabalho, optou-se pela utilização do método direto para resolver a equação da onda.

2.3 O problema Direto

O problema direto tem como objetivo determinar o campo de onda sísmico quando os parâmetros do modelo são conhecidos. No contexto da FWI, o problema direto envolve a simulação computacional da resposta física de um meio, cujas propriedades desejam ser determinadas. Em outras palavras, envolve a definição sistemática dos parâmetros (VASCONCELOS, 2018), que podem ser expressos da seguinte forma:

$$G(v) = d_v \tag{2.2}$$

v representa diversas propriedades da estrutura interna, como densidade, impedância, anisotropia e parâmetros elásticos e inelásticos. Aqui, G é um operador que representa a modelagem direta, ou seja, a simulação das ondas sísmicas através do meio para calcular os dados sísmicos sintéticos (d_v) correspondentes aos parâmetros do modelo v . Esse operador G simula as mesmas condições presentes nos dados reais coletados em campo durante a aquisição sísmica (KEAREY *et al.*, 2002).

A base da modelagem sísmica provém do conhecimento da estrutura do meio. Essas informações são obtidas por meio da simulação da resposta física resultante da emissão de sinais acústicos, associados às configurações das fontes sísmicas S . Esses sinais percorrem o espaço Ω , causando reflexões de acordo com a estrutura do subterrâneo. Posteriormente, esses dados são coletados, contendo informações do espaço associadas aos receptores sísmicos (KEAREY *et al.*, 2002).

Portanto, o valor do campo de ondas u pode ser obtido a partir de um subconjunto de receptores $\{x^{i1}, x^{i2}, \dots, x^{ir}\} \subset M(\Omega)$ que captam e registram os sinais do campo de onda ao longo do tempo $[0, T]$. Esse registro é comumente conhecido como

sismograma e pode ser descrito como uma função $d : [0, T] \rightarrow \mathbb{R}^r$,

$$\begin{aligned} d(t) &= (d_1(t), d_2(t), \dots, d_r(t)), \text{ onde} \\ d_k(t) &= u_k(t) \simeq u(x^k, t) \\ &\text{e } k = 1, 2, \dots, r \end{aligned} \quad (2.3)$$

Dessa forma, no caso acústico, o problema direto simula o processo de obtenção de um sismograma a partir de um campo de velocidade v e um conjunto de fontes sísmicas. O campo de onda u_v é calculado com base no modelo $v \in \mathbb{R}$, que resolve o sistema descrito pela equação 2.3, estabelecendo a relação com o sismograma d_v .

Os métodos diretos, também conhecidos como métodos de equação de onda completa, são aplicados ao discretizar o modelo do meio em um número finito de pontos. Exemplos desses métodos incluem o método dos elementos finitos, o método pseudo-espectral e o método de diferenças finitas (TARANTOLA, 1984). Neste trabalho, a abordagem de diferenças finitas foi selecionada devido à sua facilidade de implementação e baixo custo computacional em relação a outros métodos.

2.3.1 Discretização Espacial

Dado que a solução analítica da equação de onda é aplicável apenas a casos simples, métodos numéricos são empregados para simular a propagação de ondas sísmicas. Para isso, é necessário discretizar o campo de onda contínuo $u(x, t)$, aproximando-o por um número finito de coeficientes $\bar{u}_1(t), \bar{u}_2(t), \dots, \bar{u}_n(t)$, os quais podem ser organizados em um vetor $\bar{u}(t)$ de dimensão N .

A equação da onda 2.1 pode ser reescrita da seguinte forma:

$$M(v) \frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + S(x, t), \quad x \in [0, 1] \quad (2.4)$$

onde $M(v)$ é a matriz de massa. A discretização espacial da equação 2.4 é representada por (FICHTNER, 2010):

$$M(\bar{v})\tilde{u}(t) + K\bar{u}(t) = \bar{S}(t) \quad (2.5)$$

Aqui, a notação $(\bar{\cdot})$ denota a discretização aproximada de:

$$\bar{u}_i(t) \approx u(x^i, t) \quad (2.6)$$

$$K_i \bar{u} = \Delta(x^i, t) \quad (2.7)$$

$$\bar{v}_i = v(x^i) \quad (2.8)$$

$$\bar{S}_i(t) = S(t), \quad i = 1, 2, 3, \dots, N, \quad t \in [0, T] \quad (2.9)$$

K é uma matriz de rigidez definida em $\mathbf{R}^{M \times M}$, $M(\bar{v})$ é a matriz de massa.

2.3.2 Discretização Temporal

Conforme o princípio da causalidade, não há propagação de ondas sem um estímulo do termo fonte (CAMARGO, 2019). Suponhamos que o estado do campo de onda no início siga esse princípio, ou seja, que a condição de pressão nula seja atendida. Nesse caso, as condições iniciais são definidas como:

$$u(x, t) = 0, \quad (2.10)$$

$$u_t(x, t) = 0. \quad (2.11)$$

O processo de discretização é realizado para $N + 1$ pontos, a fim de evitar problemas de reflexão nos contornos, onde $x_i = i\Delta x$. As condições de contorno são as seguintes (DEMANET, 2016):

$$\bar{u}_1 = \bar{u}_2 = \dots = \bar{u}_N = 0 \quad (2.12)$$

A fórmula para a segunda derivada espacial usando diferenças finitas centradas é expressa da seguinte maneira:

$$\frac{\partial^2 u}{\partial x^2}(x_i, t) = \frac{\bar{u}_{i+1} - 2\bar{u}_i + \bar{u}_{i-1}}{(\Delta x)^2} + O((\Delta x)^2) \quad (2.13)$$

desde que a função u seja suficientemente suave. A notação $O(\cdot)$ representa uma constante multiplicativa proporcional a $\frac{\partial^4 u}{\partial x^4}$. A multiplicação por $M(v)$ é efetuada multiplicando cada componente pelo valor $m(v_i)$. Reunindo todos esses componentes, a discretização espacial pode ser expressa como (FICHTNER, 2010):

$$M(\bar{v})\tilde{u} - \frac{\bar{u}_{i+1} - 2\bar{u}_i + \bar{u}_{i-1}}{(\Delta x)^2} = 0 \quad (2.14)$$

Reescrevendo a equação 2.14 em termos das matrizes de massa M e de rigidez K :

$$M(\bar{v})\tilde{u}(t) + K\bar{u}(t) = 0 \quad (2.15)$$

O vetor \bar{u} consiste em $N + 1$ coeficientes \bar{u}_i . Como M é uma matriz diagonal, ela pode ser invertida. Para a discretização temporal, partimos de:

$$\tilde{u}(t) = M^{-1}[\bar{f}(t) - K\bar{u}(t)] \quad (2.16)$$

Para discretizar a segunda derivada temporal, utilizamos a aproximação de dife-

rença finita:

$$\tilde{u} \approx \frac{1}{\Delta t^2} [\bar{u}(t + \Delta t) - 2\bar{u}(t) + \bar{u}(t - \Delta t)] \quad (2.17)$$

Combinando a equação 2.17 com a equação 2.15, obtemos a versão discretizada da equação de onda:

$$\bar{u}(t + \Delta t) = 2\bar{u}(t) - \bar{u}(t - \Delta t) - \Delta t^2 M^{-1} [\bar{f}(t) - K\bar{u}(t)] \quad (2.18)$$

A equação 2.18 sugere uma abordagem iterativa para a solução de diferenças finitas da equação de onda. Começando com $\bar{u}(0)$ e $\bar{u}(-\Delta t)$ definidos pelas condições iniciais, calculamos $\bar{u}(\Delta t)$. Em seguida, usando $\bar{u}(0)$ e $\bar{u}(\Delta t)$, encontramos $\bar{u}(2\Delta t)$. Esse processo é repetido conforme necessário (FICHTNER, 2010).

2.4 O Problema Inverso

O problema inverso tem como objetivo recuperar as características quantitativas da estrutura geológica da região estudada, ou seja, encontrar um modelo que descreva o meio de propagação das ondas sísmicas. Nesse contexto, partimos de um ponto inicial com um vetor de dados calculados (d_v) a partir da estimativa atual dos parâmetros do modelo v . O propósito é minimizar a discrepância entre os dados observados e os dados calculados ($\delta d = d_0 - d_v$). A resolução desse problema ocorre de maneira iterativa, em que a função de discrepância é utilizada para quantificar a diferença entre os dois conjuntos de dados (THOMASSEN, 2008). A solução esperada é obter a distribuição de velocidade na subsuperfície.

O cálculo da função de discrepância é expresso por:

$$F_D = D(d_v, d_0) \quad (2.19)$$

onde F_D é a função que avalia a discrepância entre o dado calculado d_v com base no modelo v e o dado observado d_0 obtido a partir de uma superfície de aquisição.

Assim, o problema de estimar o campo de velocidade real pode ser formulado como um problema de otimização não linear:

$$\min_v F_D(v) \quad (2.20)$$

Existem várias funções de discrepância que podem ser utilizadas nesse caso, sendo a mais comum a norma $L2$. Essa abordagem é expressa da seguinte maneira:

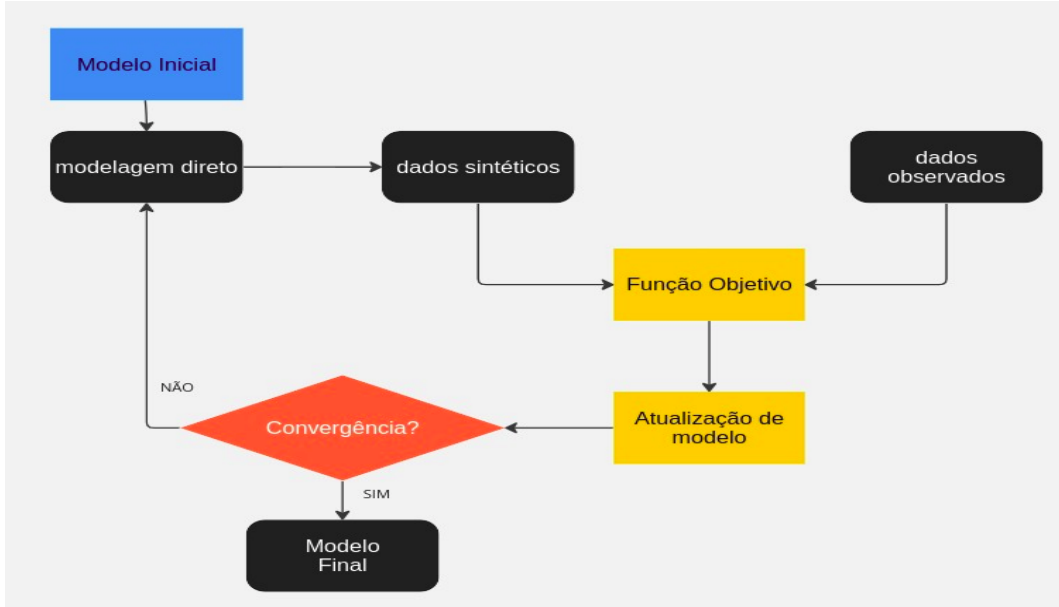


Figura 2.3: Esquema do problema inverso resolvido como um processo de otimização
 Fonte: Elaborada pelo autor

$$F_D = D(d_v, d_0) = \frac{1}{2} \sum_r \int_0^T |d_v(x_r, t; v) - d_0(x_r, t)|^2 dt \quad (2.21)$$

Nessa equação, x_r representa a localização dos receptores, e essa formulação pode ser estendida para incorporar várias fontes. Na prática, a integral de tempo é tratada numericamente como uma soma discreta.

Apesar da norma $L2$ ser amplamente utilizada para resolver o problema inverso 2.20, ela tende a ser sensível a ruídos presentes em dados sísmicos reais (BUBE e LANGAN, 1997). Para contornar esse problema, abordagens de regularização e informações a priori são frequentemente empregadas para auxiliar na convergência, já que informações adicionais podem ajudar a aproximar a solução do mínimo global (MARTINS, 2015). Diversas outras formas de normas foram propostas para mitigar essa sensibilidade, como a norma $L1$ (CRASE *et al.*, 1990; TARANTOLA, 2005), o critério Huber (GUITTON e SYMES, 2003; HA *et al.*, 2009) e uma abordagem híbrida entre as normas $L1$ e $L2$ (BUBE e LANGAN, 1997).

No entanto, nenhuma dessas alternativas proporcionou uma solução geral para o problema. No próximo capítulo, será introduzida a função de discrepância Kullback-Leibler, que será empregada como função de ajuste no escopo deste trabalho.

Os algoritmos mais empregados para resolver o problema 2.20 podem ser agrupados em três classes, métodos tipos gradientes, método tipo Newton e método tipo quasi-Newton. A descrição detalhada desses métodos podem ser encontradas em (NOCEDAL e WRIGHT, 2006). Embora esses algoritmos possuam características próprias, existe algo em comum entre esses grupos, o cálculo do gradiente.

2.5 Método do Estado-adjunto

O cálculo do jacobiano do campo de ondas em relação aos parâmetros usando o método de diferenciação numérica requer um considerável esforço computacional. Isso ocorre porque a abordagem exige a recriação do modelo em cada ponto para cada perturbação (FICHTNER, 2010). Esse método de cálculo do gradiente por diferenciação numérica pode se tornar inviável à medida que o número de parâmetros do modelo aumenta. Para contornar essa limitação, é empregado o método do estado adjunto, que encontra seus fundamentos na teoria variacional e tem sido amplamente utilizado para resolver problemas de controle ótimo regidos por equações diferenciais parciais (PLESSIX e PERKINS, 2010).

2.5.1 Condição de Imagem

Ao se obter a equação da onda através do método adjunto, considera-se uma função $d_r(t)$ que depende do índice do receptor r e do tempo t , v representa o modelo de velocidade e $u(x, t)$ é o campo da onda. O operador F é o mapeamento linear básico do espaço do modelo para o espaço de dados, enquanto F^* é o mapeamento linear básico do espaço de dados para o espaço do modelo. O operador F desempenha função importante na formulação do gradiente em uma função de custo $F_D \frac{\partial F_D}{\partial v}$:

$$\frac{\partial F_D}{\partial v}[v] = F^*(F[v] - s)$$

A relação entre essas três funções é expressa como (DEMANET, 2016):

$$\langle d, Fv \rangle = \langle F^*d, v \rangle \quad (2.22)$$

O produto interno no lado esquerdo refere-se ao espaço dos dados:

$$\langle d, Fv \rangle = \sum_r \int_0^T d_r(t)u(x, t)dt, \quad u = Fv \quad (2.23)$$

Enquanto o produto interno no lado direito refere-se ao espaço do modelo:

$$\langle F^*d, v \rangle = \int_{\mathbb{R}^n} (F^*d)(x)v(x)dx \quad (2.24)$$

A relação $u = Fv$ está implicitamente codificada nas duas equações (DEMANET, 2016):

$$\left(v_0 \frac{\partial^2}{\partial t^2} - \Delta \right) u = -v \frac{\partial^2 u_0}{\partial t^2} \quad \left(v_0 \frac{\partial^2}{\partial t^2} - \Delta \right) v_0 = f \quad (2.25)$$

Ao considerar a equação 2.23, o objetivo é destacar a contribuição de v na expressão $\langle d, Fv \rangle = \sum_r \int_0^T d_r(t)u(x, t)dt$ e isso pode ser feito com técnica de integração por partes. Para resolver essa integral em relação a x , é necessário transformar essa soma discreta em uma integral. Para integrar por partes na variável x , é necessário converter essa soma discreta sobre os receptores em uma integral continua ao longo do eixo x , isso pode ser feito por meio de um conjunto de dados estendidos de distribucional, onde cada medição $d_r(t)$ é acompanhada por um delta de Dirac δ localizado em x_r :

$$d_{ext}(x, t) = \sum d_r(t)\delta(x - x_r) \quad (2.26)$$

Dessa forma, tem-se:

$$\langle s, Fv \rangle = \int_{\mathbb{R}^n} \int_0^T d_{ext}(x, t)u(x, t)dxdt \quad (2.27)$$

Para simplificar o uso da equação da onda para u , introduzimos um campo auxiliar $q(x, t)$ que resolve a mesma equação da onda:

$$\left(v_0 \frac{\partial^2}{\partial t^2} - \Delta \right) q(x, t) = d_{ext}(x, t), \quad x \in \mathbb{R}^n \quad (2.28)$$

Com a condição de contorno ainda não especificadas no tempo. Ao substituir essa expressão para $d_{ext}(x, t)$ e aplicar a técnica de integração por partes tanto no espaço quanto no tempo, os seguintes termos são obtidos:

$$\begin{aligned}
\langle d, F_v \rangle &= \int_V \int_0^T q(x, t) \left(v_0 \frac{\partial^2}{\partial t^2} - \Delta \right) u(x, t) dx dt \\
&\quad + \int_V v_0 \frac{\partial q}{\partial t} u \Big|_T dx - \int_V v_0 q \frac{\partial u}{\partial t} \Big|_T dx \\
&\quad + \int_{\partial V} \int_0^T \frac{\partial q}{\partial n} dS_x dt - \int_{\partial V} \int_0^T q \frac{\partial u}{\partial n} dS_x dt
\end{aligned} \tag{2.29}$$

Onde V é um volume que se estende por todo o \mathbb{R}^n , e ∂V representa o contorno desse volume. Os termos de contorno desaparecem quando o V é grande suficiente. Já os termos de contorno $t = 0$ desaparecem porque $u|_{t=0} = \frac{\partial u}{\partial t}|_{t=0} = 0$. Os termos de contorno $t = T$ desaparecem se forem impostas as seguintes condições:

$$q|_{t=T} = \frac{\partial q}{\partial t} \Big|_{t=T} = 0 \tag{2.30}$$

Nesse caso, essas condições são consideradas condições finais, já que são aplicadas de $t = T$ e a equação 2.28 é resolvida retroativamente no tempo. Dessa forma, a parte esquerda dessa equação pode ser simplificada da seguinte maneira:

$$\begin{aligned}
\langle d, F_v \rangle &= \int_{\mathbb{R}^n} \int_0^T q(x, t) \left(v_0 \frac{\partial^2}{\partial t^2} - \Delta \right) u(x, t) dx dt \\
&= - \int_{\mathbb{R}^n} \int_0^T q(x, t) v(x) \frac{\partial^2 u_0}{\partial t^2} dx dt
\end{aligned} \tag{2.31}$$

O campo de onda q é denominado campo adjunto ou estado adjunto, e a equação 2.28 é chamada de equação adjunta. Observa-se que o termo q foi introduzido por conveniência computacional. Além disso, é importante destacar que o resultado é independente de v .

O termo de condição de imagem é dado por:

$$(F^* d)(x) = - \int_0^T q(x, t) \frac{\partial^2 u_0}{\partial t^2}(x, t) dt \tag{2.32}$$

o campo adjunto q_s relacionado a fonte s é definida por:

$$\left(v_0 \frac{\partial^2}{\partial t^2} - \Delta \right) q_s(x, t) = d_{ext}(x, t) \tag{2.33}$$

Onde $d_{ext}(x, t)$ é a fonte adjunta, e a equação 2.32 é conhecida como condição

de imagem.

2.5.2 Cálculo de gradiente da função de discrepância

Considere uma função de discrepância $F_D(v)$ continuamente diferenciável que mede a diferença entre o sismograma real d_0 e o sismograma calculado d_v a partir do modelo de velocidade v . A equação de propagação da onda pode ser escrita como:

$$F(v)u = S \quad (2.34)$$

A primeira derivada da função de discrepância $F_D(v)$ pode ser expressa como:

$$\frac{\partial F_D(v)}{\partial v} = \left\langle \frac{\partial F_D(v)}{\partial u}, \frac{\partial u}{\partial v} \right\rangle \quad (2.35)$$

O cálculo da derivada $\frac{\partial u}{\partial v}$ pode ser computacionalmente custoso, o que pode dificultar a resolução numérica do problema. No entanto, o método adjunto oferece uma maneira de eliminar esse termo (DEMANET, 2016). Derivando a equação 2.34 em relação a v :

$$\frac{\partial F}{\partial v} u + F \frac{\partial u}{\partial v} \quad (2.36)$$

Onde u é o campo de onda calculado, F é a matriz de impedância e S é o termo fonte.

É notável que as derivadas nas equações 2.35 e 2.36 contêm o mesmo termo $\frac{\partial u}{\partial v}$, com a diferença de que na última equação esse termo é multiplicado por F . Como o termo fonte S não depende do modelo de parâmetros v , a ideia é usar o método adjunto para clonar a matriz F através desse termo, ou seja, usando a equação adjunta:

$$Fq = \frac{\partial F_D(v)}{\partial u} \quad (2.37)$$

Aqui, q é o campo adjunto e $\frac{\partial F_D(v)}{\partial u}$ pode ser visto como a adjunta de F .

Dessa forma, a expressão 2.38 pode ser reescrita como:

$$\frac{\partial F_D(v)}{\partial v} = \left\langle Fq, \frac{\partial u}{\partial v} \right\rangle = \left\langle q, F \frac{\partial u}{\partial v} \right\rangle = - \left\langle q, \frac{\partial F}{\partial v} u \right\rangle \quad (2.38)$$

É interessante notar que essa expressão é a mesma que a condição de imagem 2.32. Portanto, para os propósitos deste trabalho, podemos estender a equação 2.38 para calcular o gradiente de qualquer função de discrepância, desde que seja possível obter a fonte adjunta associada a essa função (YANG, 2018).

Corolário 0.1 *Considerando uma função $f : \mathbb{R}^r \times \mathbb{R}^r \rightarrow \mathbb{R}_+$ uma função continuamente diferenciável e a função de discrepância definida como:*

$$D_f(d_1, d_2) = \int_0^T f(d_1(t), d_2(t)) dt \quad (2.39)$$

Se definirmos:

$$F_D(v) = D_f(d_v, d_0) \quad (2.40)$$

Então a fonte adjunta associada $Z_v(t)$ será:

$$Z_v(t) = \frac{\partial f}{\partial d_1}(d_v(t), d_0(t)) \quad (2.41)$$

Esse corolário proporciona uma fórmula para calcular a fonte adjunta de uma função de discrepância que mede a diferença entre os sismogramas d_v e d_0 no mesmo instante de tempo para todo $t \in [0, T]$.

2.6 Algoritmos

Os métodos empregados na resolução de problemas inversos podem ser categorizados em duas abordagens distintas: local e global. Os métodos globais visam identificar a solução ótima dentre todas as possibilidades, exemplos dos quais incluem Monte Carlo, algoritmos genéticos, Simulated Annealing, Particle Swarm Optimization, entre outros. No entanto, é notável que essas abordagens geralmente incorrem em custos computacionais substanciais, o que limita sua aplicação em contextos que envolvem problemas de grande envergadura (RYU e YIN, 2022).

Por outro lado, os métodos locais, embora não garantam a obtenção da solução global ideal, destacam-se por seu menor consumo computacional. Essas estratégias iterativas ajustam os parâmetros na direção descendente, seguindo tal abordagem até que o critério de parada seja atingido (CARNEIRO, 2017). Consequentemente, o ponto de partida inicial desempenha um papel crucial no sucesso do processo de otimização. Entre os métodos locais mais relevantes estão o método de Newton, os métodos quasi-Newton, gradiente conjugado, gradiente descendente, entre outros,

como detalhadamente abordado por (NOCEDAL e WRIGHT, 2006).

Dentre os métodos locais, destaca-se o método quasi-Newton como uma abordagem frequentemente empregada na resolução de problemas de *Full-Waveform Inversion* (FWI). Este método apresenta a vantagem de combinar a rapidez de convergência característica do método de Newton com um custo computacional mais acessível, proporcionando assim um equilíbrio relevante para a busca da solução ótima (YANG e ENGQUIST, 2018).

2.6.1 LBFGS

O método L-BFGS (Limited-memory Broyden-Fletcher-Goldfarb-Shanno) é classificado como um algoritmo de otimização de segunda ordem, sendo uma variante dos métodos Quasi-Newton. Este método se destina a situações em que a segunda derivada não pode ser calculada diretamente. Diferentemente do método de Newton, que se baseia na matriz Hessiana (uma vez que é um método de segunda ordem), o L-BFGS supera a limitação associada à computação do inverso da Hessiana, que pode ser computacionalmente dispendioso. Em vez disso, o método Quasi-Newton estima o inverso da Hessiana utilizando informações do gradiente, o que o torna uma abordagem computacionalmente eficiente. O L-BFGS aborda essa limitação ao aproximar essa segunda derivada. Enquanto o método de Newton faz uso da matriz Hessiana, sua aplicação é restrita devido à exigência do cálculo do inverso da Hessiana, uma tarefa computacionalmente onerosa. (NOCEDAL e WRIGHT, 2006)

Derivado do método BFGS (Broyden-Fletcher-Goldfarb-Shanno), o L-BFGS representa uma expansão do BFGS, com a característica adicional de atualizar o cálculo da matriz Hessiana a cada iteração, em vez de recalculá-la integralmente. No entanto, a dimensão da Hessiana e seu inverso está intrinsecamente ligada ao número de parâmetros de entrada da função objetivo. Portanto, em problemas de grande escala, a magnitude da Hessiana pode se tornar uma preocupação significativa. O L-BFGS enfrenta esse desafio ao adotar uma simplificação do inverso da Hessiana da iteração anterior. Ao contrário do BFGS, que se baseia no histórico completo de gradientes, o L-BFGS depende dos n gradientes mais recentes, geralmente variando entre 5 a 20 iterações anteriores. Isso resulta em uma demanda substancialmente menor de armazenamento (NOCEDAL e WRIGHT, 2006). A atualização das variáveis de decisão é dada pela equação a seguir:

$$x_{k+1} = x_k - \alpha_k H_k \nabla f_k \quad (2.42)$$

onde x_k é o ponto atual, x_{k+1} é o esquema de atualização do novo ponto, α é o

tamanho do passo, Δf_k gradiente é H_k é a aproximação do inverso da Hessiana e pode ser atualizada da seguinte forma:

$$H_{k+1} = V_k^T H_k V_k + \rho_k S_k S_k^T \quad (2.43)$$

onde

$$\rho_k = \frac{1}{Y_k^T S_k} \quad (2.44)$$

$$V_k = I - \rho_k Y_k S_k^T \quad (2.45)$$

e

$$S_k = x_{k+1} - x_k \quad (2.46)$$

$$Y_k = \nabla f_{k+1} - \nabla f_k \quad (2.47)$$

O algoritmo L-BFGS é apresentado a seguir:

Algoritmo 0.1 Escolha o ponto inicial x_0 , e um inteiro $m > 0$;

$k \leftarrow 0$;

repete

Escolha H_k^0 ;

calcula $p_k \leftarrow -H_k \nabla f_k$;

calcula α_k que satisfaz a condição de Wolfe;

calcula $x_{k+1} \leftarrow x_k + \alpha_k p_k$

se $k > m$

descarta o vetor $\{s_{k-m}, y_{k-m}\}$ do armazenamento;

calcula e salva $s_k \leftarrow x_{k+1} - x_k, y_k = \nabla f_{k+1} - \nabla f_k$;

calcula e salva s_k e y_k

calcula $k \leftarrow k + 1$;

Até convergir

Esse algoritmo visa otimizar a convergência para a solução desejada. Ele é projetado para resolver problemas de otimização com aproximações da matriz Hessiana de forma iterativa.

2.6.2 Busca Linear

Em atividades que envolvem movimento, como em otimização, após a seleção de uma direção, o próximo passo envolve calcular a quantidade de esforço necessária para atingir o objetivo, ou seja, determinar a distância que se deseja percorrer

nessa direção. Para esse fim, são empregados algoritmos de busca linear, que têm o propósito de encontrar um passo que leve à convergência para o mínimo da função em cada iteração (NOCEDAL e WRIGHT, 2006)

$$x_{k+1} = x_k + \alpha_k p_k \quad (2.48)$$

onde α_k denota o tamanho do passo, e p_k representa a direção de descida. O sucesso do processo de otimização é influenciado por ambos os parâmetros. No entanto, a seleção adequada do tamanho do passo pode ser desafiadora, envolvendo um equilíbrio delicado entre a escolha de α_k , que reduz o valor da função, e o custo computacional de avaliar todas as opções possíveis (NOCEDAL e WRIGHT, 2006). Idealmente, a escolha ótima do tamanho do passo para alcançar o mínimo global é dada pela função

$$\phi(\alpha) = f(x_k + \alpha p_k), \alpha > 0. \quad (2.49)$$

No entanto, a avaliação da função e do gradiente para encontrar o mínimo global pode ser computacionalmente custosa. Como alternativa, são adotadas condições que os valores de α devem satisfazer, e os passos escolhidos são aqueles que atendem a essas condições. O processo de seleção de α pode ser dividido em duas etapas. Primeiro, são identificados intervalos de valores possíveis para α . Em seguida, dentro desses intervalos, é escolhido o melhor passo (RYU e YIN, 2022).

Uma das condições fundamentais é a chamada condição de decrescimento suficiente, que estabelece que o tamanho do passo escolhido, α_k , deve resultar em uma redução na função f (NOCEDAL e WRIGHT, 2006). Isso é expresso como:

$$f(x_k + \alpha_k) \leq f(x_k) + c\alpha_k \nabla f_k^T p_k \quad (2.50)$$

O último termo contém uma derivada direcional, onde $0 < c < 1$. A inequação 2.50 é conhecida como a condição de Armijo.

Para a busca linear utilizando a estratégia de *backtracking*, se o tamanho do passo candidato for apropriado, a condição de decrescimento é suficiente para encerrar o processo de busca linear. O algoritmo 0.2 ilustra esse processo.

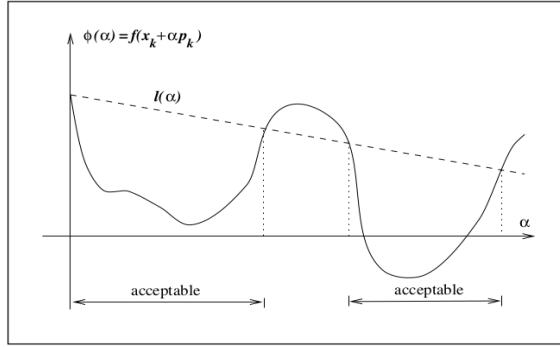


Figura 2.4: Condição de diminuição suficiente.

Algoritmo 0.2

Escolha $\bar{\alpha} > 0, \rho \in (0, 1), c \in (0, 1)$; *Defina* $\alpha \leftarrow \bar{\alpha}$
repete até $f(x_k + \alpha p_k) \leq f(x_k) + c\alpha \nabla f_k^T p_k$
 $\alpha \leftarrow \rho\alpha$;
end (repete)
termina com $\alpha_k = \alpha$

O tamanho adequado do passo é determinado após algumas iterações, pois α_k é reduzido até um ponto onde a condição de decrescimento é satisfeita.

Capítulo 3

Divergência

A função de divergência é uma ferramenta amplamente empregada na solução de uma variedade de problemas em áreas como engenharia, matemática aplicada, visão computacional, estatística, economia e outros (NOUNAGNON, 2016). Tipicamente, essa função é utilizada para quantificar a diferença entre dois conjuntos de dados. A divergência entre dois elementos p e q pertencentes ao conjunto dos números reais positivos \mathbb{R}^+ é convencionalmente representada por $D(p||q)$ mas, por questões de simplificação, será usada a notação $D(p, q)$.

As categorias de funções de divergência clássicas, tais como as divergências de Csiszár, de Bregman (BREGMAN, 1967) ou de Jensen (JENSEN, 1906), normalmente se baseiam em funções diferenciáveis e estritamente convexas, e sua caracterização depende das propriedades dessas funções (LANTÉRI, 2020). Vale destacar que essas categorias são formalmente distintas, embora, na prática, uma divergência possa pertencer a mais de uma delas, como é o caso da divergência de Kullback-Leibler.

Ainda que as aplicações das funções de divergência abranjam diversas áreas, a maior parte dos fundamentos teóricos sobre esse tema foi desenvolvida no âmbito da teoria da informação ((KULLBACK e LEIBLER, 1951),(CSISZÁR, 1967),(JENSEN, 1906),(BASSEVILLE, 1989)). Esses trabalhos foram concebidos para abordar a discrepância entre duas distribuições de probabilidade, implicando que os dados em estudo são não negativos e satisfazem $\sum_i p_i = 1$ e $\sum_i q_i = 1$.

Nesse contexto, se $D(p, q)$ for zero, isso indica que as distribuições p e q são idênticas. Quanto maior for o valor de $D(p, q)$ maior é a discrepância entre as distribuições e, por conseguinte, mais informação extra é necessária para descrever corretamente os dados usando q em vez de p (COVER, 1999).

A relação com a entropia relativa é fundamental aqui. A entropia relativa mede o grau de incerteza associado a uma distribuição em relação a outra. A KL pode ser considerada como uma forma específica de entropia relativa entre as distribuições p

e q (COVER, 1999).

Se os conjuntos de dados for interpretado como representações de distribuições de probabilidade (onde cada dado é uma possível observação de uma variável aleatória), então a KL fornece uma medida de quanta informação do conjunto de dados representado por p está presente no conjunto representado por q . Quanto menor a KL entre p e q , mais próximo está q de representar corretamente a distribuição p e, portanto, mais informação de p está contida em q . Quando a KL é maior, implica que q não está próximo o suficiente de p e mais informações são necessárias para representar p com precisão usando q (COVER, 1999).

No entanto, existem muitos problemas que não se enquadram nesse tipo de dados, tal como é o caso do presente trabalho. Como resultado, a função de divergência utilizada aqui leva em consideração essas particularidades, seguindo as diretrizes apresentadas em (LANTÉRI, 2020).

3.1 Convergência

Definição 0.1 *Seja C_f uma divergência Csiszár ou (f -divergência), onde $f(x)$ é uma função estritamente convexa, a divergência Csiszár entre p, q é definido como*

$$C_f(p, q) = \sum_i c_f(p_i, q_i) = \sum_i q_i f\left(\frac{p_i}{q_i}\right) \quad (3.1)$$

onde $p = (p_1, p_2, \dots, p_n)$ e $q = (q_1, q_2, \dots, q_n)$.

No contexto clássico, $p, q > 0$ e $\sum_i p_i = 1$, $\sum_i q_i = 1$, uma vez que p, q são distribuições de probabilidade e $f(1) = 0$ (ÖSTERREICHER, 2002). O gradiente dessa função em relação a q_j é dado por:

$$\frac{\partial C_f(p, q)}{\partial q_j} = f\left(\frac{p_j}{q_j}\right) - \frac{p_j}{q_j} f'\left(\frac{p_j}{q_j}\right) \quad (3.2)$$

O gradiente se anulará quando $p_j = q_j, \forall j$, o que implica que $f(1) = f'(1)$. No contexto desse trabalho, foi adotada uma função convexa padrão, $f_c : R \rightarrow R$, às particularidades dos dados discretos, a fim de criar uma divergência com propriedades invariáveis (ZHANG, 2004).

$$f_c(x) = f(x) - f(1) - (x - 1)f'(1) \quad (3.3)$$

O gradiente da equação 3.3 é dado por:

$$\frac{\partial C_{f_c}(p, q)}{\partial q_j} = f\left(\frac{p_j}{q_j}\right) - \frac{p_j}{q_j} f'\left(\frac{p_j}{q_j}\right) - f(1) + f'(1) \quad (3.4)$$

O gradiente se anula quando $p_j = q_j, \forall j$. No calculo da hessiana, é considerado a função c_f definida como:

$$c_f(p, q) = qf\left(\frac{p}{q}\right) \quad (3.5)$$

onde $p, q \in R^+$, a hessiana da função 3.5 pode ser expressa da seguinte maneira:

$$\frac{\partial^2 c_f(p, q)}{\partial p^2} = \frac{1}{q} f''\left(\frac{p}{q}\right) > 0 \quad (3.6)$$

$$\frac{\partial^2 c_f(p, q)}{\partial q^2} = \frac{p^2}{q^3} f''\left(\frac{p}{q}\right) > 0 \quad (3.7)$$

$$\frac{\partial^2 c_f(p, q)}{\partial p \partial q} = \frac{p}{q^2} f''\left(\frac{p}{q}\right) > 0 \quad (3.8)$$

A partir das derivadas de segunda ordem 3.6 e 3.7, é possível concluir que a função 3.5 é convexa. A análise da convexidade conjunta é realizada considerando o determinante hessiano:

$$\frac{p^2}{q^4} \left[f''\left(\frac{p}{q}\right) \right]^2 - \left[\frac{p}{q^2} f''\left(\frac{p}{q}\right) \right]^2 = 0 \quad (3.9)$$

Um dos autovalores é igual a zero, enquanto o outro é igual ao traço da matriz hessiana, que é positivo. Portanto, a expressão 3.5 além de ser convexa, a matriz é positiva semi-definida. Dessa forma, a divergência de Csiszár é conjuntamente convexa, visto que é uma soma de termos conjuntamente convexos.

3.2 Divergência de Kullback Leibler

Seja f_c uma função convexa definida como:

$$f_c(x) = x \log x + 1 - x \quad (3.10)$$

A correspondente Divergência de Csiszár, associada à 3.10, é identificada como a Divergência de Kullback-Leibler, conforme discutido em (LANTÉRI, 2020). O

gráfico da equação 3.10 pode ser visto na figura 3.1.

$$KL(p, q) = \sum_i p_i \log \frac{p_i}{q_i} + q_i - p_i \quad (3.11)$$

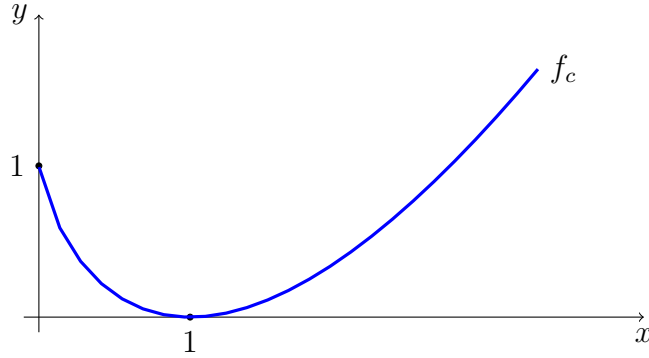


Figura 3.1: gráfico da função $f_c = x \log x + 1 - x$
 Fonte: Elaborada pelo autor

O gradiente da Divergência de Kullback-Leibler em relação a 3.11 pode ser expresso como:

$$\frac{\partial KL(p, q)}{\partial q_j} = -\frac{p_j}{q_j} + 1 \quad (3.12)$$

Este gradiente é igual a zero quando $q_j = p_j, \forall j$. Também é possível considerar a forma dual da função KL, empregando a função espelho de 3.10:

$$\tilde{f}_c(x) = \log \frac{1}{x} + x - 1 \quad (3.13)$$

Nesse contexto, obtém-se:

$$KL(p, q) = \sum_i q_i \log \frac{q_i}{p_i} + p_i - q_i \quad (3.14)$$

O gradiente correspondente a essa função é:

$$\frac{\partial KL(p, q)}{\partial q_j} = \log \frac{q_j}{p_j} \quad (3.15)$$

O gráfico que representa a função 3.13 pode ser visto na figura 3.2.

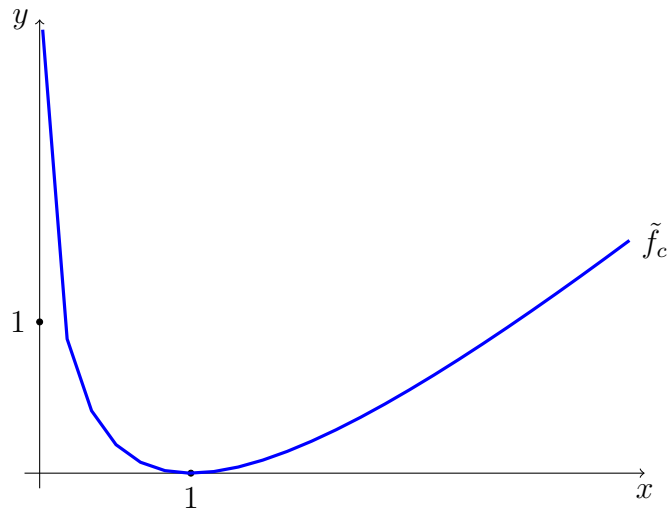


Figura 3.2: gráfico da função $f_c = \log \frac{1}{x} + x - 1$
 Fonte: Elaborada pelo autor

3.2.1 Propriedades

A partir da desigualdade de Shannon-Gibbs (COVER, 1999), é possível concluir que a Divergência de Kullback-Leibler é sempre positiva ou igual a zero. A igualdade ocorre quando $p_i = q_i, \forall i$.

$$0 \leq KL(p, q)$$

Assimetria

A Divergência de Kullback-Leibler não é referida como 'distância' principalmente por não satisfazer a desigualdade triangular:

$$KL(p, q) \neq KL(q, p)$$

Desigualdade triangular

A divergência de Kullback-Leibler não é chamada de 'distância' principalmente porque não satisfaz a desigualdade triangular. Isso é:

$$KL(p, q) + KL(q, h) \not\leq KL(p, h)$$

Como a função de Kullback-Leibler mede a discrepância apenas entre valores positivos, é necessário garantir que todos os valores do sismograma sejam positivos. Isso é alcançado através da aplicação da função parte positiva $x : [x]^+$ onde $x = (x_1, \dots, x_n)$ $x \in \mathbb{R}^n$. No entanto, a função positiva não é suave, o que leva à aplicação de técnicas de suavização (smoothing) da função parte positiva, conforme descrito

na seção subsequente.

3.3 Suavização da Função Parte Positiva

A função parte positiva é um recurso fundamental na matemática, que tem encontrado aplicabilidade em diversas áreas, como engenharia, economia e ciências naturais. Muitos problemas podem ser reconfigurados utilizando essa função, principalmente quando as propriedades dos objetos estudados impõem a restrição de serem não negativos. Isso ocorre com frequência em contextos que envolvem minimização de preços, restauração de imagens, reconstrução de sinais, equilíbrio estocástico, entre outros (CHEN, 2012).

Uma das vantagens significativas da função parte positiva é a capacidade de transformar problemas com restrições de desigualdade, como $f(x) < 0$, onde f é uma função de dimensão n definida no espaço real $R^n \rightarrow R$ em problemas de minimização irrestrita. Essa transformação ocorre da seguinte maneira: $\min_{x \in R^n} \left\| [f(x)]^+, \right\|$, em que $[x]^+$ denota a função parte positiva aplicada a cada componente de x e $\left\| \cdot \right\|$ representa uma norma monotônica em R^n (CHEN, 1995).

Contudo, vale destacar que, apesar de suas propriedades vantajosas, a função parte positiva não é diferenciável, o que a torna não suave. Essa característica impede a utilização direta de métodos numéricos de otimização baseados em gradientes nessa função. Para superar essa limitação, (CHEN, 1995) propõem a utilização de uma função suavizada para aproximar a função parte positiva, tornando-a diferenciável e, portanto, mais adequada para métodos de otimização numérica.

Seja $[x]^+ = \max\{x, 0\}$ a função fundamental de aproximação suavizada, definida da seguinte forma:

$$[x]^+ = \int_{-\infty}^x \rho(t) dt \quad (3.16)$$

onde ρ é a função degrau, definida como:

$$\rho(x) = \begin{cases} 1 & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} \quad (3.17)$$

A função degrau ρ pode ser expressa da seguinte forma:

$$\rho(x) = \int_{-\infty}^x \delta(t) dt \quad (3.18)$$

onde $\delta(x)$ é a função delta de Dirac que atende as seguintes propriedades:

$$\delta(x) \geq 0, \int_{-\infty}^{+\infty} \delta(t) dt = 1 \quad (3.19)$$

A função delta de Dirac pode ser relacionada com funções de densidade de probabilidade, (CHEN, 1995) propôs suavizar a função de delta de Dirac e as suas integrais através de funções de densidade. Dessa forma, seja $\delta_1(x)$ uma função continuamente em partes e que satisfaz critérios de uma função de densidade, então para parametrizar a função de densidade considera-se :

$$\delta_\tau(x) = \frac{1}{\tau} \delta_1\left(\frac{x}{\tau}\right) \quad (3.20)$$

Nessa parametrização, τ controla o grau de aproximação. Dado que 3.20 é satisfeita, $\rho_\tau : \mathbb{R} \rightarrow \mathbb{R}$ pode ser definida como:

$$\rho_\tau(x) = \int_{-\infty}^x \delta_\tau(t) dt \quad (3.21)$$

Vale notar que quando o τ tende a zero, a função ρ_τ converge para ρ . A função de aproximação $\sigma_\tau : \mathbb{R} \rightarrow \mathbb{R}$ é expressa como:

$$\sigma_\tau(x) = \int_{-\infty}^x \rho_\tau(t) dt \quad (3.22)$$

Uma característica importante é que, como a função δ é não negativa, a função σ_τ é bem definida e convexa. Além disso, ela pode se tornar estritamente convexa quando $\delta > 0$.

3.3.1 Propriedades da função parte positiva

A função $\sigma_\tau(x)$ exhibe certas propriedades conforme o parâmetro τ se aproxima de zero. Essas propriedades serão discutidas a seguir.

3.3.1.1 Proposição

Considere a função de densidade de probabilidade δ definida como $\rho_\tau(x) = \frac{1}{\tau}\delta_1(\frac{x}{\tau})$, onde τ é um parâmetro positivo. Suponha que $\delta(x)$ satisfaça as seguintes suposições:

A1 $\delta(x)$ é contínua por partes com um número finito de peças e satisfaz a equação (3.19)

A2 $E[|x|]_{\delta(x)} = \int_{-\infty}^{+\infty} |x|\delta(x)dx < +\infty$, então a definição de $\sigma_\tau(x)$ dada pela equação 3.22 é consistente.

3.3.1.2 Proposição propriedade de $\sigma_\tau(x), \tau > 0$

Seja $\delta(x)$ e $\delta_\tau(x)$ definidas de acordo com a proposição 3.3.1.1, e supondo que $\delta(x)$ satisfaz as condições (A1) e (A2). Então a função $\sigma_\tau(x)$ possui as seguintes propriedades:

(1) $\sigma_\tau(x)$ é continuamente diferenciável, e se, além disso, $\delta(x)$ for k -vezes continuamente diferenciável, então $\sigma_\tau(x)$ é $(k + 2)$ vezes continuamente diferenciável.

(2) $D_2\tau \leq \sigma_\tau(x) - [x]^+ \leq D_1\tau$, onde

$$D_1 = \int_{-\infty}^0 |t|\delta(t)dt$$

e

$$D_2 = \max\left\{\int_{-\infty}^{+\infty} t\delta(t)dt, 0\right\}$$
(3.23)

Dentre os vários exemplos de funções de aproximação suave σ_τ da função $[\cdot]^+$, pode se considerar as seguintes funções (CHEN, 2000):

Neural Networks smoothing Function

Seja

$$\delta_1(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$
(3.24)

$$\sigma(x, \tau) = x + \tau \log(1 + e^{\frac{-x}{\tau}})$$
(3.25)

$$\sigma'_\tau(x, \tau) = \left(\frac{1}{1 + e^{\frac{-x}{\tau}}}\right)$$
(3.26)

Chen-Harker-Kanzow-Samle Smoothing Function

Seja

$$\delta_1(x) = \frac{2}{(x^2 + 4)^{\frac{3}{2}}} \quad (3.27)$$

$$\sigma(x, \tau) = \frac{x + \sqrt{x^2 + 4\tau^2}}{2} \quad (3.28)$$

$$\sigma'_\tau(x, \tau) = \frac{1}{2} \left(1 + \frac{x}{\sqrt{x^2 + 4\tau^2}} \right) \quad (3.29)$$

Em que δ_1 é a função de densidade, σ é a sua versão suavizada e σ' é a derivada da função σ .

Com as ferramentas necessárias à disposição, a próxima etapa é explorar como a função Kullback-Leibler pode ser aplicada no contexto do problema de FWI. No capítulo subsequente, serão apresentados os passos e as abordagens adotadas para incorporar essa técnica na resolução dos desafios associados à *Full-Waveform Inversion*.

Capítulo 4

Divergência aplicado ao FWI

Neste capítulo, é apresentada a metodologia adotada para abordar a discrepância entre dois sismogramas através da aplicação da divergência Kullback-Leibler. Essa escolha é motivada pelo fato de que a divergência Kullback-Leibler possui característica que não permite a presença de valores negativos. Para lidar com essa particularidade, é proposto o uso da função parte positiva. No entanto, essa função não é diferenciável, o que levou à aplicação de um método de suavização para torná-la diferenciável, conhecido como método de aproximações suaves (smoothing).

Um aspecto importante abordado neste capítulo é a questão da assimetria, uma vez que os objetos comparados podem ter massas desiguais. Como solução para essa assimetria, é introduzida uma versão de medidas assimétricas, que levam em consideração o desbalanceamento. Essas medidas assimétricas são formuladas de maneira compatível com a abordagem da divergência Kullback-Leibler.

4.1 O problema de desbalanceamento e valores negativos nos dados sísmicos

Seja a função parte positiva e negativa $[\cdot]^+, [\cdot]^- : \mathbb{R} \rightarrow \mathbb{R}$, definido respectivamente

$$[x]^+ = \max\{x, 0\} \text{ e } [x]^- = \max\{-x, 0\} \quad (4.1)$$

Considerando um sismograma $d : [0, T] \rightarrow R^r$, onde os dados sísmicos $d(t) \in R^r$ representam a variação do sinal ao longo do tempo, possível definir a parte positiva e negativa dos dados sísmicos da seguinte forma:

$$\begin{aligned} [d(t)]^+ &:= ([d_1(t)]^+, [d_2(t)]^+, \dots, [d_n(t)]^+) \\ [d(t)]^- &:= [-d(t)]^+ \end{aligned} \quad (4.2)$$

Considerando um modelo de velocidade v e o sismograma associado d_v , assim

como um sismograma observado e outro calculado d_0 , o objetivo é medir a discrepância entre $d_v(t)$ e $d_0(t)$. Contudo, esses dados podem conter valores negativos, o que impossibilita a aplicação direta da teoria tradicional de Kullback-Leibler, já que essa teoria se aplica apenas a valores não negativos (LANTÉRI, 2020). Para superar essa limitação, é empregada a função parte positiva, garantindo a ausência de valores negativos nos dados utilizados.

Isso é realizado aplicando o artifício de Manini (MAININI, 2012):

$$x = [d_v(t)]^+ + [d_0(t)]^- \text{ e } y = [d_v(t)]^- + [d_0(t)]^+ \quad (4.3)$$

Embora essa abordagem seja eficaz na lida com valores negativos, pode surgir um problema de falta de diferenciabilidade devido à função $[\cdot]^+$. Para superar essa dificuldade, uma solução é aplicar uma técnica de suavização à função parte positiva, obtendo assim $[\cdot]^+ \sigma_\tau(\cdot)$, onde $\sigma_\tau(\cdot)$ é uma função suavizada. Dessa forma, a transformação é realizada da seguinte maneira:

$$\sigma_\tau(d_v(t)) + \sigma_\tau(-d_0(t)) \text{ e } \sigma_\tau(-d_v(t)) + \sigma_\tau(d_0(t)).$$

no lugar de

$$[d_v(t)]^+ + [d_0(t)]^- \text{ e } [d_v(t)]^- + [d_0(t)]^+$$

$$\sigma_\tau(d_v(t)) + \sigma_\tau(-d_0(t)) \text{ e } \sigma_\tau(-d_v(t)) + \sigma_\tau(d_0(t)).$$

de

$$[d_v(t)]^+ + [d_0(t)]^- \text{ e } [d_v(t)]^- + [d_0(t)]^+$$

Isso permite contornar a questão da falta de diferenciabilidade da função parte positiva, possibilitando assim a aplicação da teoria de Kullback-Leibler com as devidas modificações para medir a discrepância entre os sismogramas observado e calculado.

4.2 Discrepância entre dados sísmicos

Considerando uma função de discrepância diferenciável, denotada como:

$$f(a, b) = D_1(a, b) \quad (4.4)$$

aplicando as transformações descritas na seção anterior, resulta na seguinte expressão:

$$f(a, b) = D_1(\sigma_\tau(a) + \sigma_\tau(-b), \sigma_\tau(-a) + \sigma_\tau(b)) \quad (4.5)$$

Aqui, D_1 é uma função que mapeia de $\mathbb{R}^v + \times \mathbb{R}^v +$ para \mathbb{R} , onde a e b pertencem a \mathbb{R}^r . Da equação 2.40, a divergência D entre os sismogramas d_v e d_0 pode ser escrita como:

$$D_1(d_v, d_0) = \int_0^T f(d_v(t), d_0(t)) dt \quad (4.6)$$

Através do corolário 0.1 e aplicando a regra da cadeia, a fonte adjunta $Z_v(t)$ pode ser obtida:

$$Z_v(t) = \frac{\partial D_1}{\partial x}(x_\tau, y_\tau) \odot \sigma'_\tau(d_v(t)) - \frac{\partial D_1}{\partial y}(x_\tau, y_\tau) \odot \sigma'_\tau(-d_v(t)) \quad (4.7)$$

Onde x e y são definidos como:

$$x = \sigma_\tau(d_v(t)) + \sigma_\tau(-d_0(t)) \quad (4.8)$$

$$y = \sigma_\tau(-d_v(t)) + \sigma_\tau(d_0(t)) \quad (4.9)$$

e para $a = (a_1, \dots, a_r), b = (b_1 \dots b_r) \in \mathbb{R}^r, a \odot b := (a_1 b_1 \dots, a_r b_r)$.

Desta forma, o objetivo deste trabalho é resolver o problema de otimização:

$$\min_{(v)} D(d_v, d_0) \quad (4.10)$$

Existem várias abordagens para resolver esse problema de otimização. Neste trabalho, é proposto utilizar a divergência Kullback-Leibler para encontrar o mínimo entre os sismogramas d_0 e d_v .

4.2.1 Divergência Kullback Leibler

A divergência de Kullback-Leibler é introduzida como uma alternativa à função de norma L2. A função de divergência KL (Kullback-Leibler) é definida da seguinte forma:

$$D(x, y) = \text{KL}(x, y) = \sum \phi\left(\frac{x_i}{y_i}\right) y_i \quad (4.11)$$

onde , $\phi(p) = p \log(p) - p + 1, p \geq 0$

No entanto, a divergência KL não aceita valores negativos. Para lidar com isso, aplicamos o método de separação por partes e aproximação suave desenvolvidos na seção 4.1.

$$D = \text{KL}(\sigma_\tau(d_v(t)) + \sigma_\tau(-d_0(t)), \sigma_\tau(-d_v(t)) + \sigma_\tau(d_0(t))) \quad (4.12)$$

O gradiente da função objetivo em relação à v é calculado de acordo com o descrito na seção 2.5 e que a fonte adjunta utilizada no cálculo da variável adjunta é dada pela equação:

$$J = \log\left(\frac{x_\tau}{y_\tau}\right) \odot \sigma'_\tau(d_v(t)) - \left(1 - \frac{x_\tau}{y_\tau}\right) \odot \sigma'_\tau(-d_v(t)) \quad (4.13)$$

A introdução do termo de regularização de entropia suaviza a função, o que abre oportunidade para a aplicação de algoritmos como o método do gradiente do otimizador L-BFGS.

Capítulo 5

Experimentos

Neste capítulo, são apresentados os resultados dos testes numéricos conduzidos, os quais foram baseados em um meio 2D isotrópico acústico no domínio de tempo. Foram realizados testes em três modelos distintos: o modelo Marmousi e o modelo da bola e o modelo Petro. Cada cenário de teste considerou duas funções de discrepância, a norma L2 e a divergência Kullback-Leibler. A comparação entre essas funções em diferentes ambientes auxilia na avaliação da eficácia de cada abordagem em relação à qualidade dos resultados, considerando critérios pré-estabelecidos.

Os testes foram feitos em uma máquina com 16 GB de memória RAM, a linguagem de programação usada foi Python, com auxílio do pacote `geopetro` e do pacote `pysit`.

No que diz respeito à metodologia de otimização empregada, utilizou-se o método quasi-Newton em conjunto com a busca linear por *backtracking* (NOCEDAL e WRIGHT, 2006). Para a suavização da função de discrepância, optou-se pela função Chen-Harker-Kanzow-Samle Smoothing, conforme ilustrado na figura 5.1, utilizando um valor $\tau = 1e - 4$ em ambos os cenários.

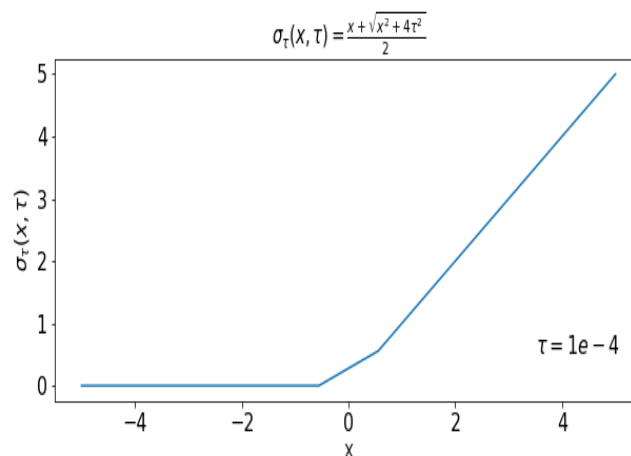


Figura 5.1: função Chen-Harker-Kanzow-Samle Smoothing

Para facilitar a validação dos resultados, o erro relativo normalizado ε foi calculado de acordo com referências ((JI *et al.*, 2019), (SONG *et al.*, 2015)). Esse erro é dado por:

$$\varepsilon = \frac{\|d_v - d_0\|_2}{\|d_0\|_2} \quad (5.1)$$

5.1 Modelo da bola pequeno

O modelo acústico foi discretizado no domínio da frequência utilizando o método de diferenças finitas. A grade de discretização possui dimensões de 241 metros tanto na vertical quanto na horizontal, com espaçamento de 12.5 metros em ambas as direções. Foram utilizados 60 tiros a uma profundidade de 40 metros, e os 241 receptores foram distribuídos em um espaçamento regular de 12.5 metros, alcançando uma profundidade de 1046 metros.

No processo de inversão, foram empregadas frequências discretas variando de 1.5 Hz a 10.5 Hz, com um intervalo de 1 Hz entre elas. Para cada frequência, foram realizadas 30 iterações para resolver o problema de inversão. As velocidades inicial e verdadeira podem ser observadas nas Figuras 5.2 e 5.3, respectivamente.

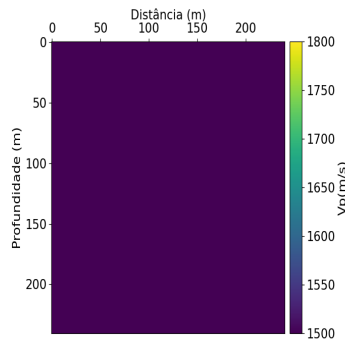


Figura 5.2: Modelo inicial

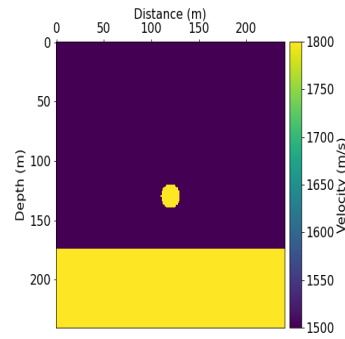


Figura 5.3: Modelo Final

5.1.1 Resultados

Embora as figuras apresentem uma semelhança visual, é notável que existem diferenças significativas entre elas. A Figura 5.5, que representa o resultado utilizando a divergência Kullback-Leibler, exibe um preenchimento melhor correspondente à bola em comparação com a Figura 5.4, obtida a partir da função L2, principalmente no trecho final da bola.

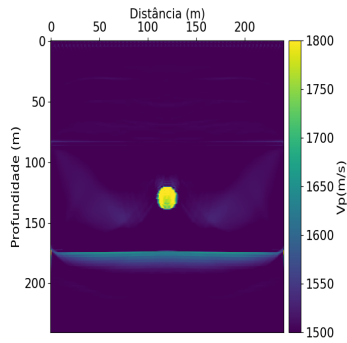


Figura 5.4: Resultado com a norma L2

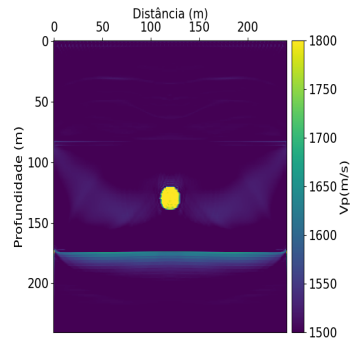


Figura 5.5: Resultado com KL

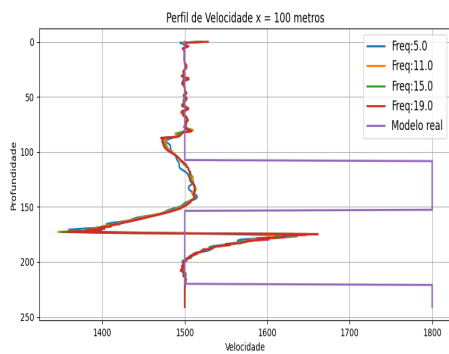


Figura 5.6: Perfil de velocidade na posição $x = 100$ metros (KL)

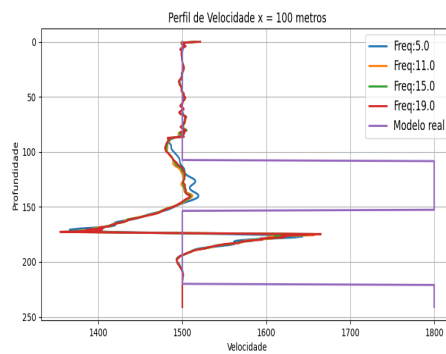


Figura 5.7: Perfil de velocidade na posição $x = 100$ metros (L2)

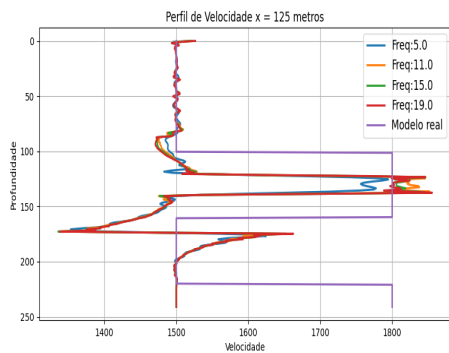


Figura 5.8: Perfil de velocidade na posição $x = 125$ metros (KL)

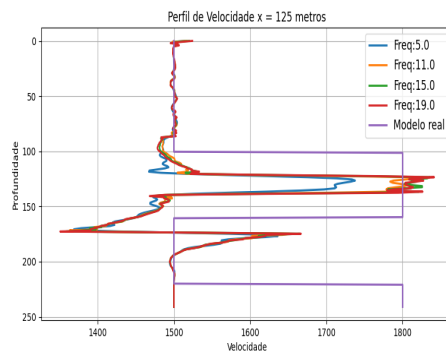


Figura 5.9: Perfil de velocidade na posição $x = 125$ metros (L2)

Para o perfil de velocidades e erro relativo, optamos por analisar a região horizontal compreendida entre 100,125 e 150 metros. Esses valores representam o início, o meio e o fim da região da bola, permitindo focar na qualidade das funções em reproduzir a parte mais crucial do modelo.

No perfil de velocidade em $x=100$ metros, como mostrado nas figuras 5.6 e 5.7,

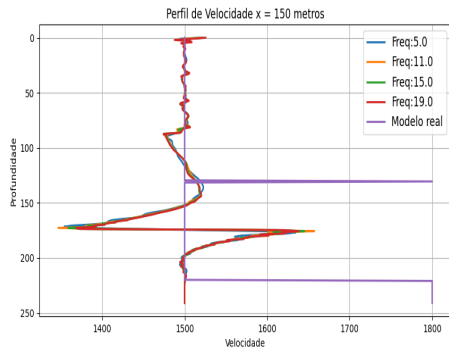


Figura 5.10: Perfil de velocidade na posição $x = 150$ metros (KL)

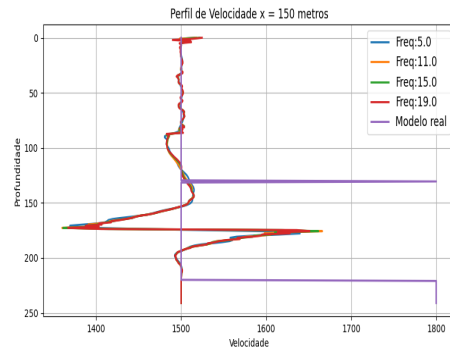


Figura 5.11: Perfil de velocidade na posição $x = 150$ metros (L2)

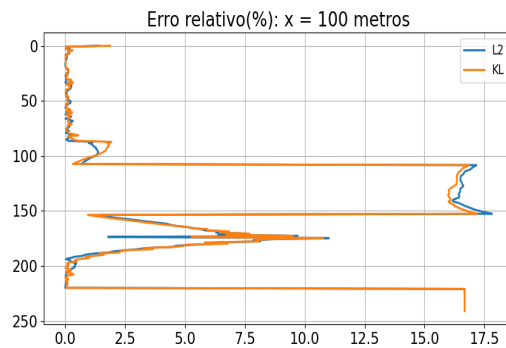


Figura 5.12: Erro relativo $x = 100$ metros

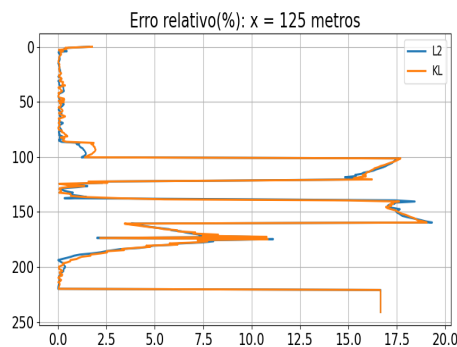


Figura 5.13: Erro relativo na posição $x = 125$ metro

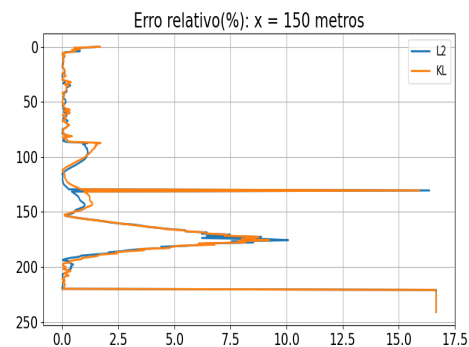


Figura 5.14: Erro relativo na posição $x = 150$ metros

os resultados obtidos não conseguiram reproduzir com precisão a curva do modelo real na faixa de profundidade de 100 a 150 metros, que é onde a bola se encontra. Isso fica evidente na Figura de Erro Relativo 5.12, onde o percentual de erro é maior em toda a profundidade. Em relação aos resultados, as curvas apresentam

comportamentos semelhantes, mas o resultado com a função L2 teve um percentual de erro maior na região de interesse, profundidade de 100 a 150 metros. Na posição $x=125$ metros (ver Figuras 5.8 e 5.9), houve uma melhoria considerável em comparação com a posição $x=100$ metros. Embora os comportamentos das curvas sejam semelhantes, os desvios ocorrem em profundidades distintas. É notável que as curvas dos resultados se aproximam mais do modelo real ao longo das frequências.

Na Figura 5.8, o perfil de velocidade na frequência de 5 Hz está mais próximo da velocidade 1800 do que na Figura 5.9. No entanto, na Figura 5.13, o percentual de erro diminuiu na região em torno da profundidade de 120 a 140 metros. Por outro lado, a profundidade de 100-120 metros ainda mantém um alto percentual de erro, e na faixa de 150 a 170 metros, o percentual aumentou, chegando próximo de 20%.

O resultado obtido com a função L2 apresentou um percentual de erro maior do que o resultado utilizando a função KL. Essa diferença nos resultados destaca a influência das funções empregadas nos modelos e suas respostas em diferentes profundidades. Por último, na região $x=150$ metros 5.10 e 5.11, a diferença entre a curva do modelo real e as outras curvas diminuiu consideravelmente. Apesar de as outras curvas não terem conseguido se aproximar da velocidade de 1700, o percentual de erro também diminuiu, como pode ser visto na Figura 5.14. No entanto, o resultado da função KL ainda apresentou um erro maior na profundidade de 100 a 150 metros, embora na profundidade de 125 a 150 metros, o resultado da função KL tenha tido um desempenho inferior.

Em relação a evolução da função objetiva ao longo das frequências, as figuras 5.15 e 5.16 apresentaram um decaimento semelhante, mas a função L2 5.15 teve maior sucesso ao resolver o problema na frequência 2.5 Hz em relação a função KL

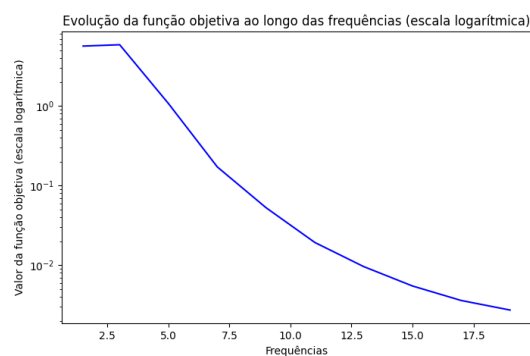
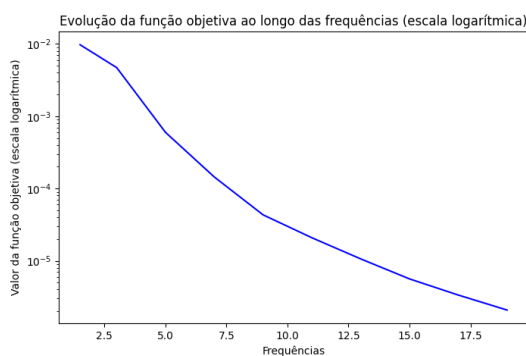


Figura 5.15: Evolução da função objetiva L2
 Figura 5.16: Evolução da função objetiva KL

A Tabela 5.1 revela que o erro relativo entre os resultados obtidos com a divergência Kullback-Leibler e a norma L2 é extremamente baixo, apenas 0,0004. Mesmo que as diferenças nos resultados sejam sutis, a divergência Kullback-Leibler demons-

Modelo	Erro
Norma l2	0.0956
Kullback-Leibler	0.0952

Tabela 5.1: Tabela de erro relativo dos resultados

trou a capacidade de alcançar resultados consistentes, sugerindo ser uma alternativa viável à abordagem da norma L2.

5.2 Modelo da bola grande

O modelo acústico foi discretizado no domínio de frequência utilizando o método de diferenças finitas, com uma grade de 241 metros tanto na direção vertical quanto na horizontal, e um espaçamento de 12.5 metros em ambas as direções. No total, foram realizados 60 tiros a uma profundidade de 40 metros, e os 241 receptores foram posicionados em intervalos regulares de 12.5 metros, abrangendo uma profundidade de até 1046 metros.

Durante o processo de inversão, foram consideradas frequências discretas variando de 1.5 Hz a 15 Hz, com um intervalo de 1.5 Hz entre elas. Em cada frequência, foram executadas 30 iterações para resolver o problema. As figuras 5.62 e 5.63 ilustram a velocidade inicial e verdadeira do modelo, respectivamente.

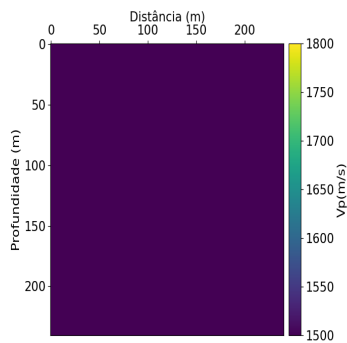


Figura 5.17: Modelo inicial

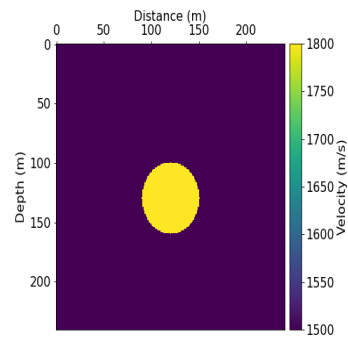


Figura 5.18: Modelo Final

5.2.1 Resultados

As figuras 5.20 e 5.19 apresentam uma semelhança visual. Ambos os resultados não conseguiram preencher totalmente a região da bola. A partir da profundidade 120 metros os resultados tiveram muita dificuldade em reconstruir a bola.

Para uma análise mais detalhada da região de recuperação da bola, foi adotado o mesmo esquema da seção anterior, dividindo a análise em 3 seções: 110, 125 e

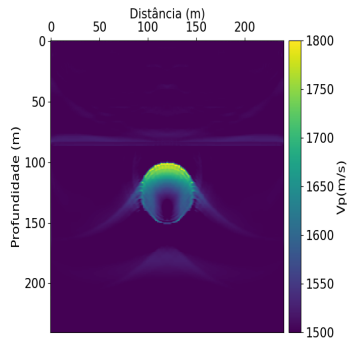


Figura 5.19: Resultado com a norma L2

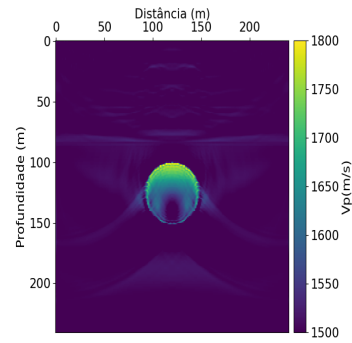


Figura 5.20: Resultado com KL

150 metros na posição horizontal. Vale ressaltar que o modelo inicial em todos os pontos de perfis de velocidades analisados é constante, com um perfil de velocidade igual a 1500. Mesmo partindo de um ponto com pouca informação, os resultados de ambas as funções conseguiram uma taxa de sucesso considerável na reconstrução das imagens.

Nos perfis de velocidade em $x = 100$ metros 5.21 e 5.22, as curvas dos resultados conseguiram se aproximar do valor máximo do perfil de velocidade do modelo real, embora em profundidades diferentes. O percentual de erro relativo é alto e chega próximo de 20% na profundidade em torno de 110 metros 5.27. Nessa região, o resultado com a função KL teve um percentual de erro menor, com uma diferença de aproximadamente 2,5% em relação ao resultado com a função norma L2.

Na região $x = 125$ metros, tanto o perfis de velocidade 5.23 e 5.24 quanto o erro relativo 5.28 tiveram comportamentos semelhantes à distância $x = 110$ metros, possivelmente devido à proximidade das duas regiões. No entanto, pode ser observado que a diferença entre os erros relativos dos resultados da função KL e L2 é mais expressiva, chegando próximo a 4%. Já na posição 150 metros, o cenário de velocidade é totalmente diferente, e ambos os resultados ficaram bem abaixo do valor de velocidade do modelo real. Em relação ao erro relativo, o resultado da função L2 obteve um percentual de erro maior do que o KL em quase toda a região da bola, próxima a 100 a 160 metros.

As figuras 5.30 e 5.31 ilustram variações no desempenho das funções em relação às diferentes frequências. Observa-se que a resolução do problema apresentou flutuações significativas, especialmente em frequências mais baixas, onde foram notadas algumas oscilações no declínio da curva até atingir os 6 Hz.

A tabela 5.2 demonstra que a diferença na soma de erro relativo entre os resultados obtidos com a divergência Kullback-Leibler (KL) e a norma L2 é extremamente pequena, com uma diferença de apenas 0,002. Embora essa diferença seja baixa em termos absolutos, como apresentado anteriormente, é importante observar que, na

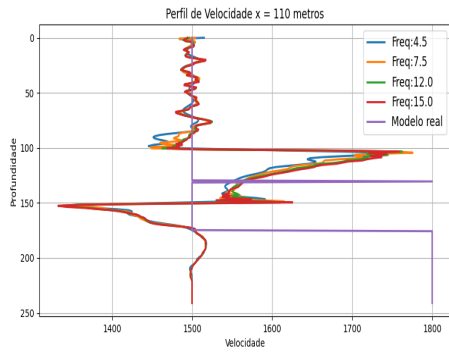


Figura 5.21: Perfil de velocidade na posição $x=110$ metros (KL)

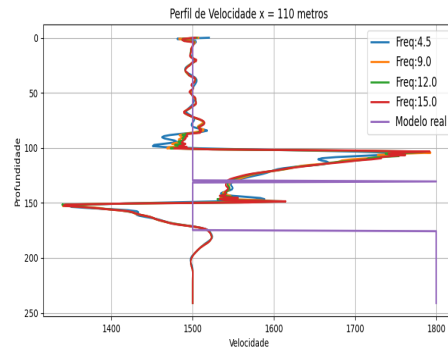


Figura 5.22: Perfil de velocidade na posição $x=110$ metros (L2)

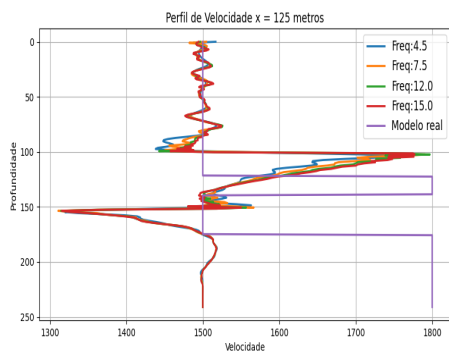


Figura 5.23: Perfil de velocidade na posição $x=125$ metros (KL)

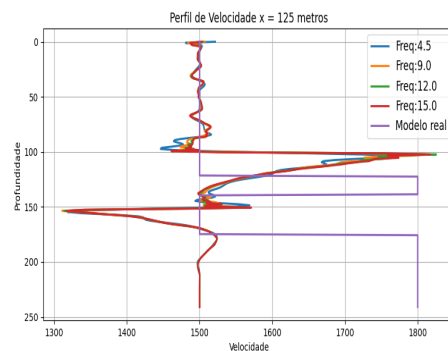


Figura 5.24: Perfil de velocidade na posição $x=125$ metros (L2)

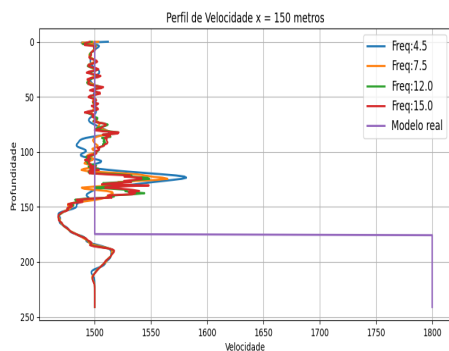


Figura 5.25: Perfil de velocidade na posição $x=150$ metros (KL)

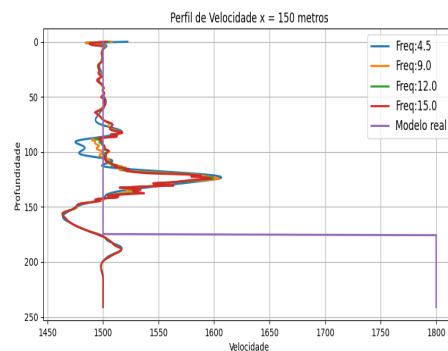


Figura 5.26: Perfil de velocidade na posição $x=150$ metros (L2)

região onde se encontra a bola, o percentual de erro do resultado da função KL é significativamente menor do que o da função norma L2. Isso indica que, apesar da pequena discrepância na soma total de erros, a função KL desempenha melhor na parte crítica da análise, onde a precisão é mais crucial.

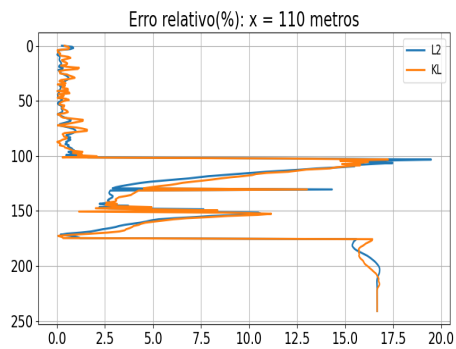


Figura 5.27: Erro relativo x=110 metros

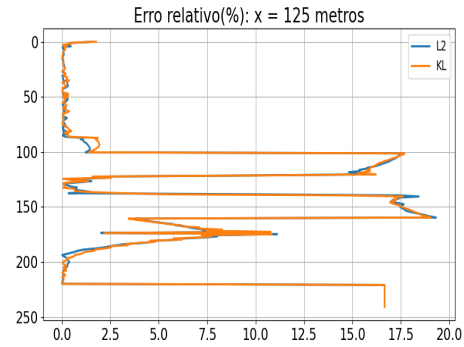


Figura 5.28: Erro relativo na posição x=125 metros

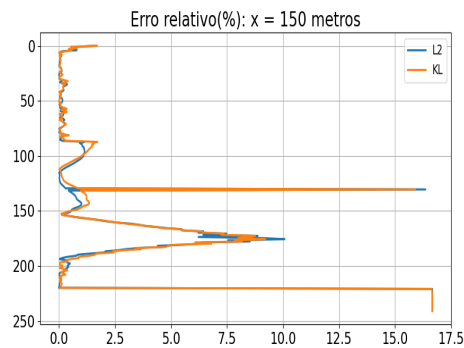


Figura 5.29: Erro relativo na posição x=150 metros

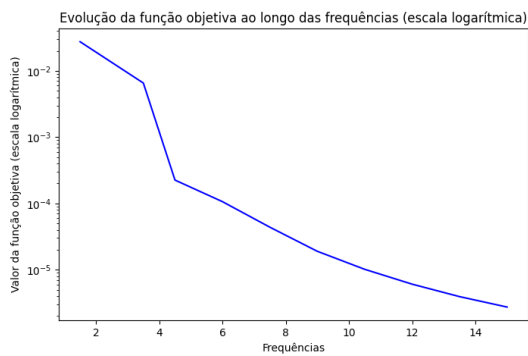


Figura 5.30: Evolução da função objetiva L2

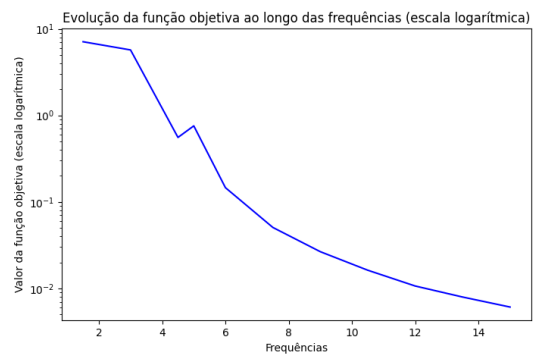


Figura 5.31: Evolução da função objetiva K1

5.3 Modelo Petro

O modelo Petro (DA COSTA *et al.*, 2019) foi discretizado no domínio da frequência utilizando o método de diferenças finitas com uma resolução de 261 metros na

Modelo	Erro
Norma l2	0.037
Kullback-Leibler	0.036

Tabela 5.2: Tabela de erro relativo dos resultados

direção vertical e 101 metros na direção horizontal. Foram utilizados 60 tiros a uma profundidade de 40 metros, com os receptores posicionados a uma profundidade de 1046 metros. Para simplificar o teste, o modelo foi redimensionado para 25% do tamanho do modelo original.

No processo de inversão, foram utilizadas as seguintes frequências discretas: 1.5, 3,5, 7, 9, 13 e 15 Hz. Para cada frequência, foram realizadas 26 iterações para resolver o problema. As figuras 5.32 e 5.33 mostram a velocidade inicial e verdadeira do modelo, respectivamente

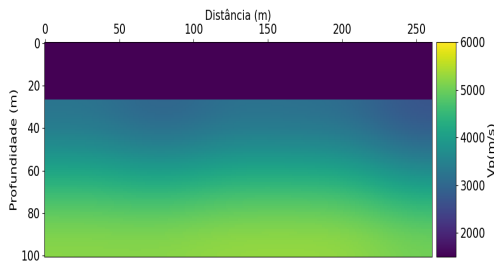


Figura 5.32: Modelo inicial

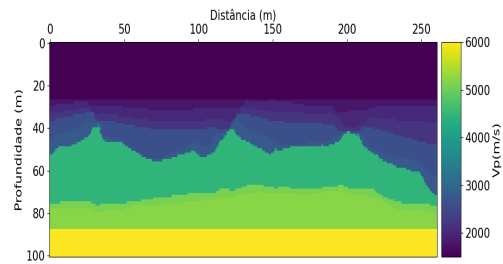


Figura 5.33: Modelo Final

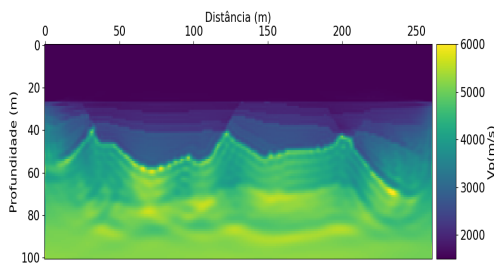


Figura 5.34: Resultado com a norma L2

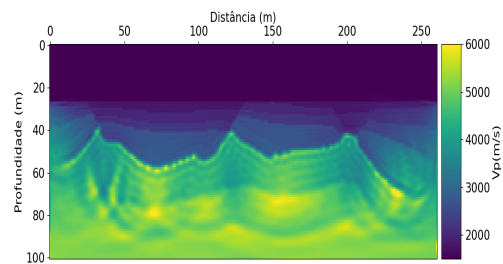


Figura 5.35: Resultado com KI

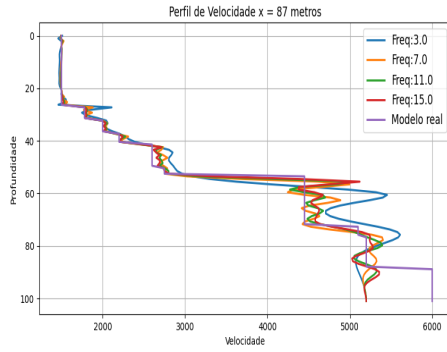


Figura 5.36: Perfil de velocidade na posição x=87 metros (Kl)

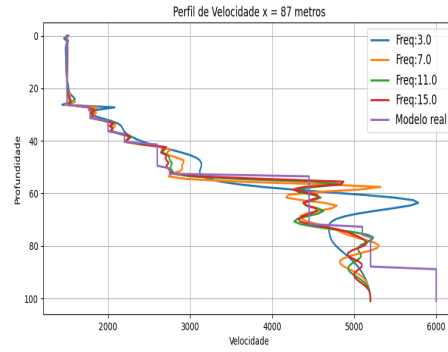


Figura 5.37: Perfil de velocidade na posição x=87 metros (L2)

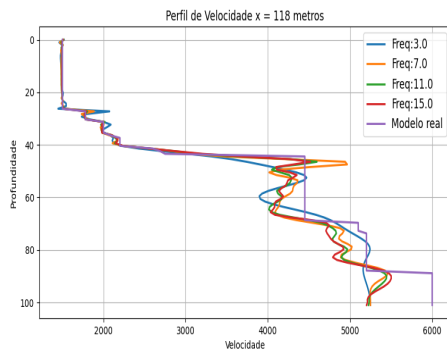


Figura 5.38: Perfil de velocidade na posição x=118 metros (Kl)

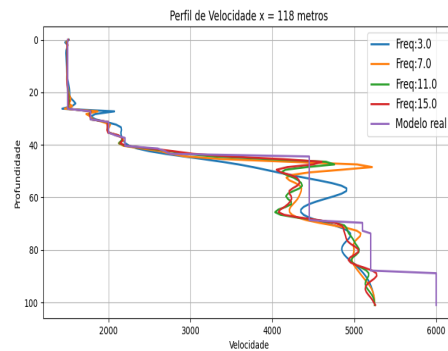


Figura 5.39: Perfil de velocidade na posição x=118 metros (L2)

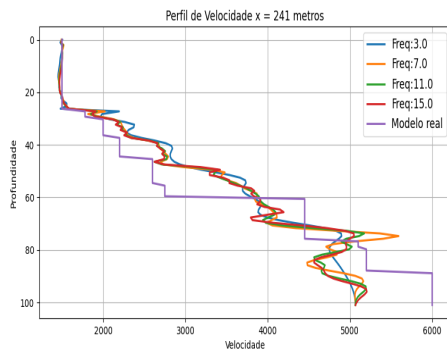


Figura 5.40: Perfil de velocidade na posição x=241 metros (Kl)

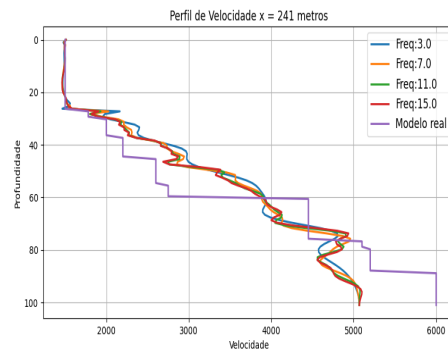


Figura 5.41: Perfil de velocidade na posição x=241 metros (L2)

5.3.1 Resultados

O modelo Petro representou o desafio mais complexo entre todos os modelos que foram resolvidos. Os resultados obtidos utilizando as funções KL (Kullback-Leibler) e norma L2 podem ser visualizados nas figuras 5.35 e 5.34, respectivamente. Entre

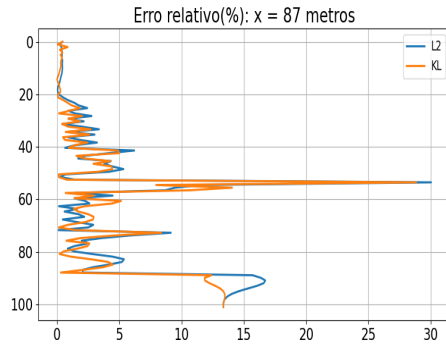


Figura 5.42: Erro relativo na posição $x=87$ metros

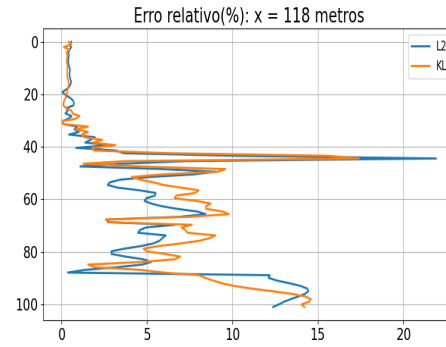


Figura 5.43: Erro relativo na posição $x=118$ metros

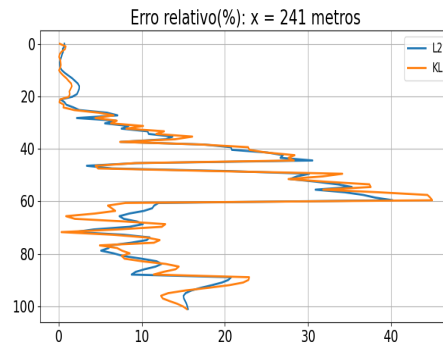


Figura 5.44: Erro relativo na posição $x=241$ metros

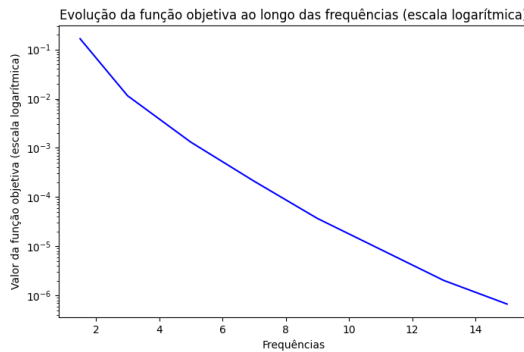


Figura 5.45: Evolução da função objetiva
L2

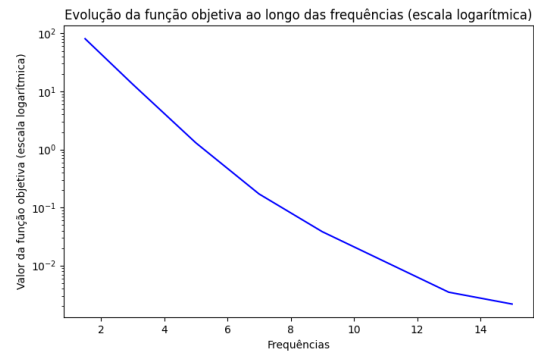


Figura 5.46: Evolução da função objetiva
K1

todos os modelos analisados, as soluções para o modelo Petro revelaram o maior percentual de erro.

Ao examinar os primeiros perfis de velocidade na posição $x = 87$ metros, conforme mostrado nas figuras 5.36 e 5.37, fica evidente que a curva do modelo real começa a se distanciar das outras curvas após a profundidade de 60 metros. O

maior percentual de erro relativo foi observado próximo aos 60 metros de profundidade, que é justamente a região de maior interesse. Nessa área, o erro relativo resultante da função L2 apresentou um percentual superior em relação à função KL, estendendo-se inclusive à zona de maiores profundidades.

No perfil de velocidade na posição $x = 118$ metros, 5.38 teve uma oscilação menor em relação ao perfil de velocidade do modelo real do que a norma L2 5.39, identificam-se duas áreas de desvio mais notáveis em comparação à posição $x=87$ metros. O maior percentual de erro ocorreu em torno dos 50 metros de profundidade, onde o erro relativo resultante da norma L2 ficou significativamente acima do erro obtido com a função KL. Contudo, essa tendência se inverteu logo em seguida, com o erro relativo da função KL tornando-se maior do que o da norma L2 na faixa compreendida entre 57 e 82 metros de profundidade, seguida novamente por um aumento no erro percentual.

Na posição $x = 241$ metros, houve dois momentos notáveis de afastamento entre o modelo real e os resultados obtidos, especificamente nas profundidades entre 30 e 60 metros e entre 78 e 100 metros, conforme ilustrado nas figuras 5.40 e 5.41 . O percentual de erro relativo nessa posição foi mais elevado em comparação com as outras posições analisadas, chegando próximo a 43%.

A tabela 5.2 resume a soma dos resultados das duas funções em análise.

Modelo	Erro
Norma l2	0.113
Kullback-Leibler	0.107

Tabela 5.3: Tabela de erro relativo dos resultados

Comparando com outros modelos, fica evidente que as discrepâncias nos erros relativos dos resultados do modelo Petro são mais acentuadas. A soma dos erros relativos da divergência KL é ligeiramente menor do que os obtidos pela função L2. Vale ressaltar que o percentual de erro relativo pode variar de acordo com a região analisada. A função KL apresentou um erro menor nas posições $x = 87$ e 118 metros, enquanto a função L2 obteve um percentual de erro menor na posição $x = 241$ metros.

5.4 Modelo Marmousi suavizado

Neste experimento, o modelo Marmousi foi suavizado em cinco etapas, utilizando o filtro passa-baixas gaussiano. Esse processo de suavização gradual foi aplicado ao modelo original com uma frequência de corte especificada. Essa abordagem permitiu testar as funções de inversão com diferentes modelos iniciais, garantindo uma melhor

convergência e resultados mais precisos na reconstrução do subsolo. O uso do filtro gaussiano ajuda a reduzir as variações de alta frequência do modelo, proporcionando um modelo com menos ruído (LE, 1994). Neste estudo, foram adotadas as seguintes frequências de corte: $1/300$, $1/250$, $1/225$, $1/200$ e $1/150$. Para cada uma dessas frequências de corte, realizou-se um total de 15 iterações. O algoritmo adotado foi o LBFGS, e a frequência central utilizada foi de 7 hertz.

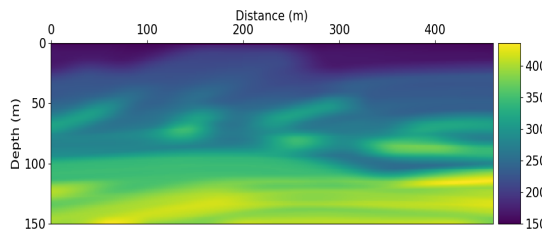


Figura 5.47: Modelo inicial

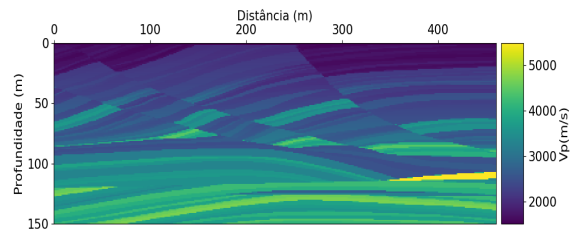


Figura 5.48: Modelo Final

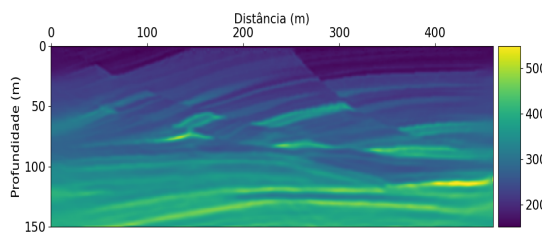


Figura 5.49: Resultado com a norma L2

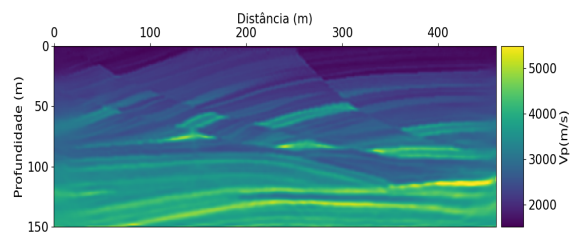


Figura 5.50: Resultado com a função KL

Tabela 5.4: Tabela de erro relativo dos resultados

Frequência	Erro relativo	
	KL	L2
$1/300$	0.1245	0.1214
$1/250$	0.1158	0.1136
$1/225$	0.1119	0.1090
$1/200$	0.1064	0.1042
$1/150$	0.09542	0.09139

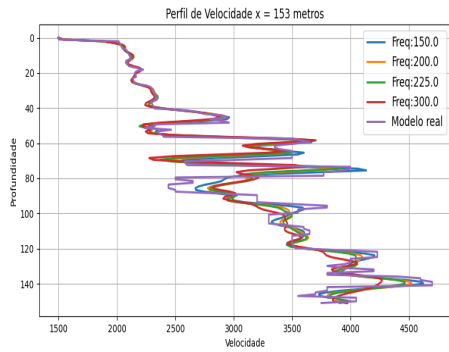


Figura 5.51: Perfil de velocidade na posição $x = 153$ metros (L2)

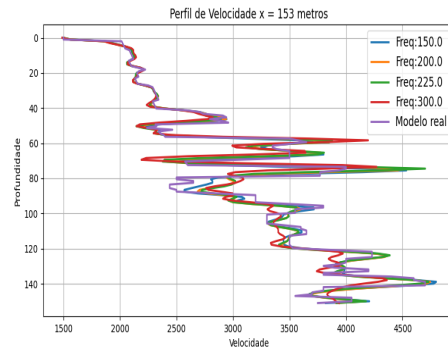


Figura 5.52: Perfil de velocidade na posição $x = 153$ metros (Kl)

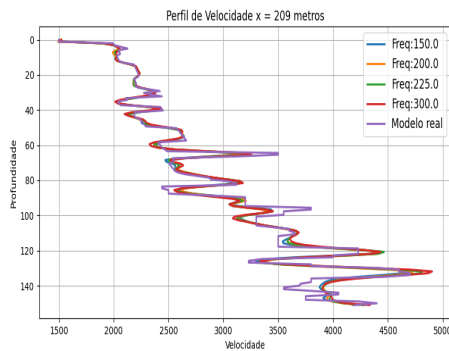


Figura 5.53: Perfil de velocidade na posição $x = 209$ metros (L2)

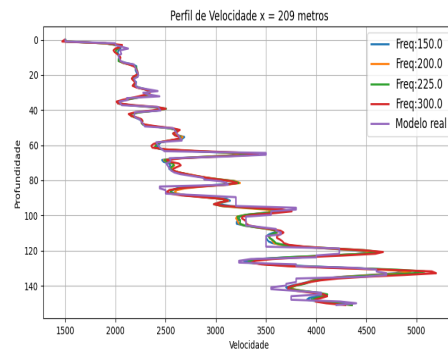


Figura 5.54: Perfil de velocidade na posição $x = 209$ metros (kl)

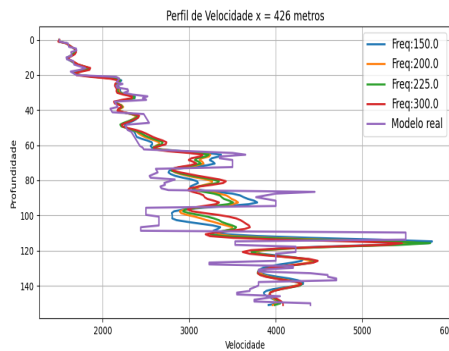


Figura 5.55: Perfil de velocidade na posição $x = 426$ metros (L2)

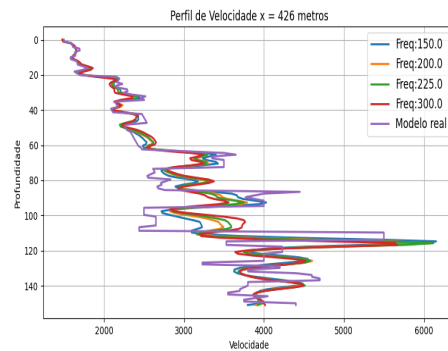


Figura 5.56: Perfil de velocidade na posição $x = 426$ metros (kl)

5.4.1 Resultados

Nas figuras 5.49 e 5.50, é possível observar que ambos os conjuntos de resultados obtiveram sucesso relativo na inversão. No entanto, a Figura 5.49 apresenta uma imagem mais nítida. Ao examinar uma região específica, como ilustrado nas figuras

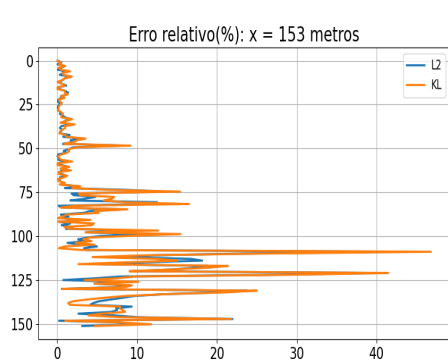


Figura 5.57: Erro relativo $x = 153$ metros

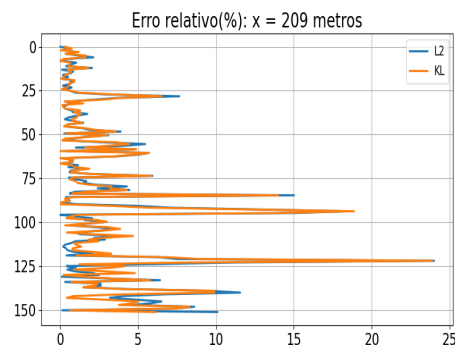


Figura 5.58: Erro relativo na posição $x = 209$ metros

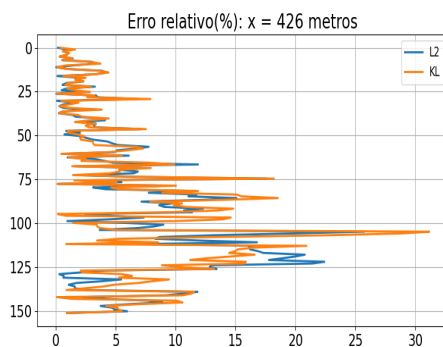


Figura 5.59: Erro relativo na posição $x = 426$ metros

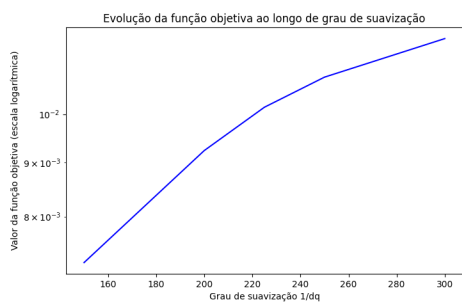


Figura 5.60: Evolução da função objetiva L2

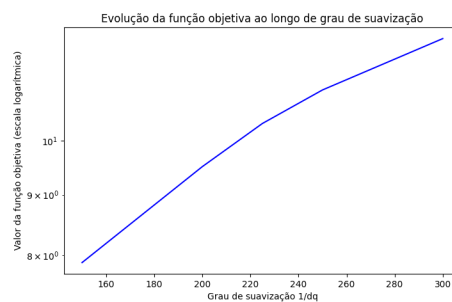


Figura 5.61: Evolução da função objetiva Kl

5.51 e 5.51 , nota-se que ambos os resultados conseguiram acompanhar de perto o perfil de velocidade do modelo original. O fato de ter iniciado com um bom modelo inicial deve ter contribuído significativamente para esse resultado. Além disso, o resultado da função KL obteve um percentual de erro menor do que o resultado da norma L2 em quase todas as profundidades, como demonstrado na figura 5.57.

O perfil de velocidade na posição $x = 209$ metros segue a mesma dinâmica

observada na posição $x = 153\text{metros}$, e o percentual de erro da função KL é menor do que o da norma L2 em quase todas as profundidades (ver a figura 5.58). Nessa região, o percentual de erro máximo é menor do que na posição $x = 153$ metros.

No entanto, os perfis de velocidade na posição $x = 426$ metros, conforme ilustrado nas figuras 5.55 5.56, apresentaram variações mais pronunciadas entre as curvas de perfil de velocidade dos resultados e do modelo real em comparação às outras posições estudadas. Isso se refletiu no percentual de erro relativo máximo, que atingiu mais de 30% com a função KL. Nessa posição, o erro relativo da norma L2 foi menor, como mostrado na figura 5.59.

Os resultados consolidados estão resumidos na tabela 5.4, que mostra a soma dos erros relativos para diferentes graus de suavização. Esses resultados indicam que a abordagem baseada na regularização L2 proporcionou uma melhor resolução nessas regiões mais profundas, resultando em uma reconstrução mais precisa do subsolo.

5.5 Fontes simultâneas

Foram conduzidos experimentos utilizando o modelo da bola grande com a abordagem de fontes simultâneas para as funções de Kl e L2. No contexto da inversão de forma de onda total simultânea, múltiplas fontes são ativadas ao mesmo tempo, proporcionando um processo de aquisição de dados mais eficiente (ZHANG *et al.*, 2018). Com o intuito de verificar a viabilidade da inversão multiescala, iniciamos o processo de inversão com a frequência inicial de 2,5 Hz. Ao todo, foram utilizadas quatro faixas de frequência nos testes de FWI, com frequências iniciais de 2,50, 4, 7,5 e 15 Hz, respectivamente. Cada faixa de frequência foi submetida a 100 iterações de inversão utilizando o método LBFGS.

Com o objetivo de assegurar o desempenho da inversão multiescala, foram codificados simultaneamente diferentes números de fontes para cada faixa de frequência. Nos testes subsequentes, foram utilizadas 50, 100, 50 e 150, 241 fontes simultâneas para as cinco faixas de frequência citado no parágrafo anterior. Os resultados podem ser vistos nas figuras a seguir.

5.5.1 Resultados

Nas figuras 5.64 e 5.65, pode-se observar uma melhoria significativa na cobertura em comparação com os resultados apresentados nas figuras 5.19 e 5.20, respectivamente. No entanto, as figuras 5.19 e 5.20 apresentam contornos mais nítidos e precisos. Para analisar mais detalhadamente, considera os perfis de velocidade na posição $x = 100$ metros, ilustrado nas figuras 5.66 e 5.67. Assim como no modelo da bola

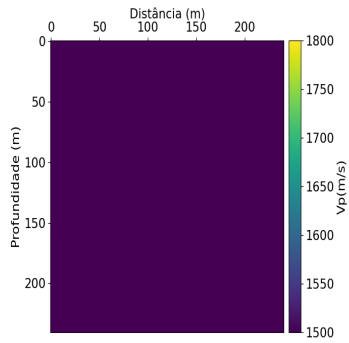


Figura 5.62: Modelo inicial

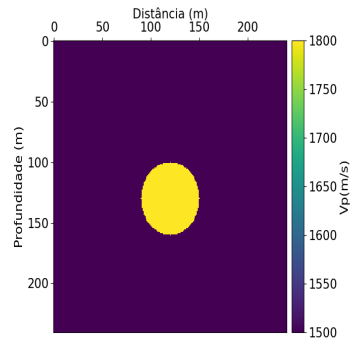


Figura 5.63: Modelo Final

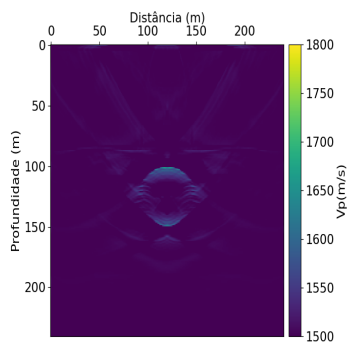


Figura 5.64: Resultado com a norma L2

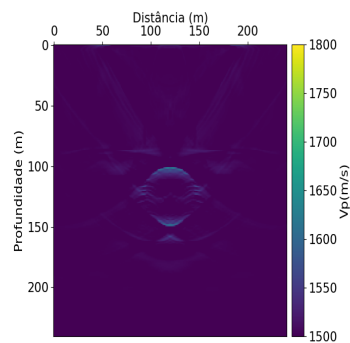


Figura 5.65: Resultado com KL

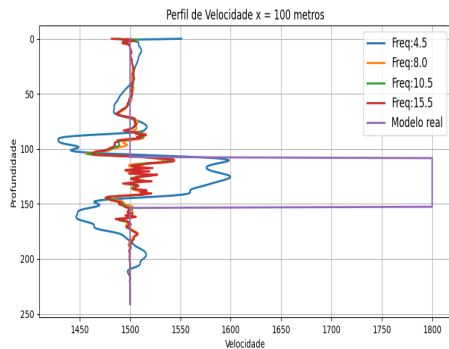


Figura 5.66: Perfil de velocidade na posição $x = 100$ metros (L2)

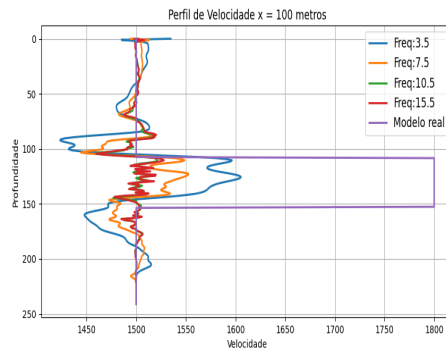


Figura 5.67: Perfil de velocidade na posição $x = 100$ metros (KL)

grande, o modelo inicial é constante, e os resultados não conseguiram se aproximar do valor máximo da velocidade do modelo real. Em ambos os casos a frequência 3.5 Hz foi a que mais ficou perto do perfil de velocidade do modelo real. As figuras 5.75 5.76 mostram que o desempenho das soluções tiveram muitas oscilações ao longo das frequências principalmente na função KL.

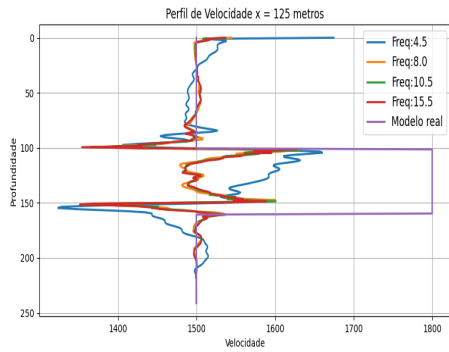


Figura 5.68: Perfil de velocidade na posição $x = 125$ metros (L2)

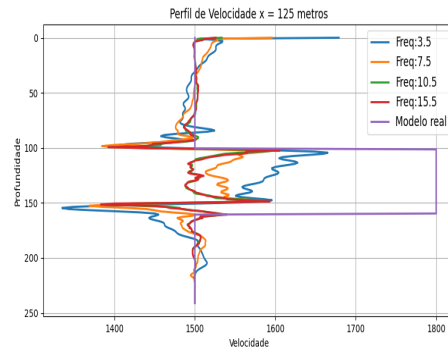


Figura 5.69: Perfil de velocidade na posição $x = 125$ metros (kl)

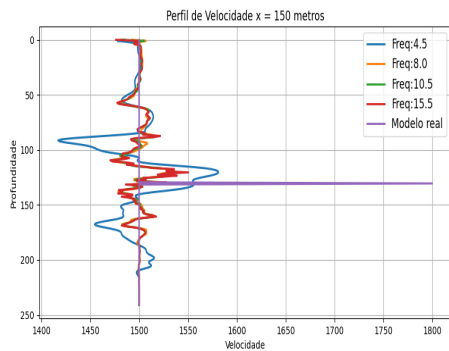


Figura 5.70: Perfil de velocidade na posição $x=150$ metros (L2)

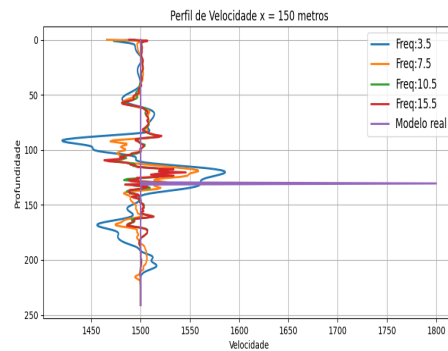


Figura 5.71: Perfil de velocidade na posição $x=150$ metros (kl)

Os percentuais de erro relativo dos resultados mostrados nas Figuras 5.72 comportaram-se de forma semelhante, no entanto, o resultado com a função KL obteve um percentual de erro menor na região onde a bola se encontra, ou seja, entre 100 e 160 metros.

Na posição $x = 125$ metros teve um desempenho semelhante à posição $x = 100$ metros, como pode ser visto nas figuras 5.68, 5.69 e 5.73. No entanto, o perfil de velocidade na posição 150 apresentou um comportamento bastante diferente. A velocidade do modelo real se concentra em profundidades próximas a 25 metros. O percentual de erro relativo máximo é menor do que nas outras posições, e a função KL teve um percentual de erro máximo maior do que a L2 na região onde o percentual de erro máximo ocorre. Em outras regiões, a norma L2 teve um percentual de erro ligeiramente maior. A soma dos erros relativos dos dois resultados é exibida na tabela 5.5, e ambos os resultados tiveram o mesmo erro total. No entanto, como demonstrado, nas regiões onde a bola se encontra, a função KL obteve um percentual de erro menor.

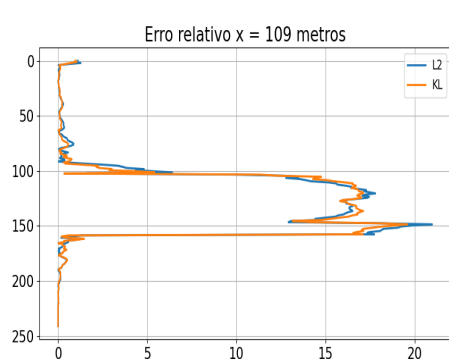


Figura 5.72: Erro relativo x = 100 metros

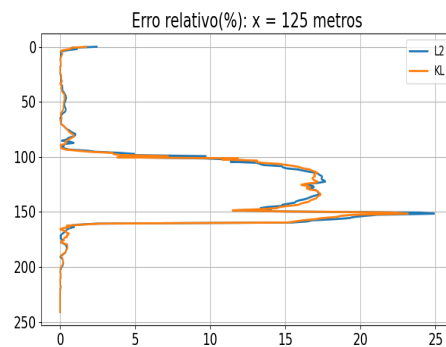


Figura 5.73: Erro relativo na posição x = 125 metros

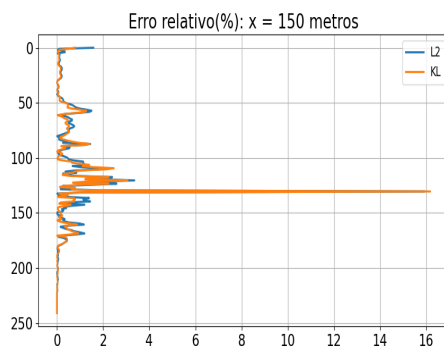


Figura 5.74: Erro relativo na posição x = 150 metros

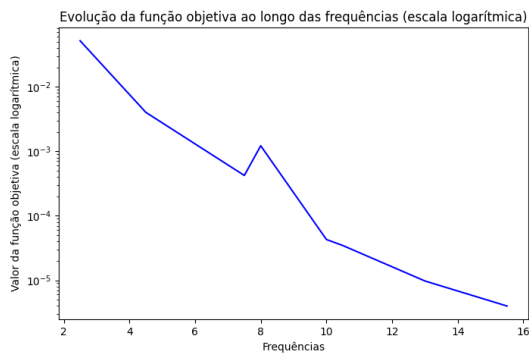


Figura 5.75: Evolução da função objetiva L2

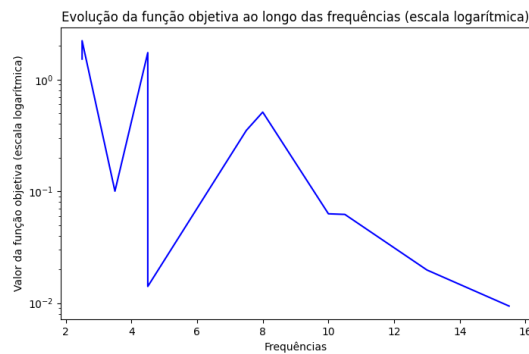


Figura 5.76: Evolução da função objetiva KI

Modelo	Erro
Norma l2	0.034
Kullback-Leibler	0.034

Tabela 5.5: Tabela de erro relativo dos resultados

5.6 Discussão

Os modelos com formato da bola demonstraram resultados mais favoráveis para ambas as funções, mesmo quando o modelo inicial estava consideravelmente distante do modelo real. Em particular, os resultados obtidos por meio da divergência de Kullback-Leibler destacaram-se em praticamente todas as posições analisadas neste estudo. Esses modelos serviram como referência para ajustes de parâmetros e para testes de novos incrementos.

O modelo que se mostrou mais desafiador de ser resolvido foi o modelo Petro, uma vez que consumia uma quantidade significativa de recursos computacionais devido ao seu tamanho. Apesar dessas complexidades, a divergência de Kullback-Leibler conseguiu obter um erro relativamente menor em comparação à função de norma L2.

Por outro lado, a norma L2 apresentou uma performance mais eficiente no modelo marmousi estudados. No entanto, é importante notar que em algumas regiões do modelo, a divergência KL conseguiu uma recuperação mais precisa.

Portanto, é fundamental ressaltar que o desempenho da função de medição de erro pode variar não apenas de acordo com o modelo, mas também depende das áreas específicas analisadas. Nesse sentido, a divergência de Kullback-Leibler surge como uma ferramenta valiosa a ser levada em consideração na resolução de problemas de FWI, dependendo da natureza específica do problema em questão.

Capítulo 6

Conclusão

A busca por uma função de discrepância robusta e prática tem sido um dos principais desafios da FWI. Neste trabalho, foi proposto o uso da divergência Kullback-Leibler (KL) para medir a diferença entre os dados calculados e observados. A KL, originada no contexto da teoria da informação, é comumente aplicada a distribuições de probabilidade. Dado que a FWI é um problema discreto, foi apresentada uma versão discreta da KL e a sugestão de utilizar a função parte positiva para lidar com valores negativos. No entanto, a função parte positiva não é continuamente diferenciável, o que dificulta o uso de alguns algoritmos, como o LBFGS. Para contornar esse problema, técnicas de aproximação suave foram empregadas.

A modelagem acústica foi adotada como base para os testes numéricos, com o objetivo de comparar os resultados obtidos com as funções de norma L2 e divergência KL. Para isso, outros parâmetros foram mantidos fixos. Os resultados da inversão demonstraram uma taxa de sucesso razoável na construção de imagens e com pouca diferença em relação à função utilizada. A norma L2 obteve melhores resultados no teste com o modelo Marmousi, enquanto a divergência KL se destacou no teste com o modelo da bola. Em todos os modelos, a divergência KL apresentou uma leve vantagem na região de maior profundidade, sugerindo que pode ser uma opção viável para resolver problemas em zonas mais profundas.

Quanto à FWI com fontes simultâneas, os resultados mostraram uma cobertura mais completa do modelo, porém com uma menor definição dos contornos. Por outro lado, a FWI tradicional resultou em uma recuperação da estrutura geral do modelo, apresentando contornos mais bem definidos. É importante ressaltar que os erros relativos obtidos para ambos os modelos foram comparáveis, indicando uma boa qualidade de ajuste em relação aos dados observados. Vale destacar a diferença significativa no custo computacional, uma vez que o método tradicional

pode demandar várias semanas de processamento, enquanto o método com fontes simultâneas requer apenas algumas horas.

Essas observações reforçam a importância de considerar diferentes funções objetivo na inversão, uma vez que cada uma delas pode oferecer vantagens e desvantagens específicas, dependendo do contexto do problema. O equilíbrio entre a cobertura do modelo e a precisão dos contornos é uma consideração crucial ao selecionar a função de objetivo mais adequada para obter os melhores resultados de inversão.

Como sugestões para trabalhos futuros, seria interessante explorar a possibilidade de desenvolver um modelo híbrido que combine elementos da norma L2 e da divergência KL. Além disso, poderiam ser testadas diferentes funções de suavização, visando melhorar ainda mais os resultados da inversão.

Referências Bibliográficas

- ASNAASHARI, A., BROSSIER, R., GARAMBOIS, S., et al., 2013, “Regularized seismic full waveform inversion with prior model information”, *Geophysics*, v. 78, n. 2, pp. R25–R36.
- BASSEVILLE, M., 1989, “Distance measures for signal processing and pattern recognition”, *Signal processing*, v. 18, n. 4, pp. 349–369.
- BREGMAN, L. M., 1967, “The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming”, *USSR computational mathematics and mathematical physics*, v. 7, n. 3, pp. 200–217.
- BUBE, K. P., LANGAN, R. T., 1997, “Hybrid l1l2 minimization with applications to tomography”, *Geophysics*, v. 62, n. 4, pp. 1183–1195.
- BUNKS, C., SALECK, F. M., ZALESKI, S., et al., 1995, “Multiscale seismic waveform inversion”, *Geophysics*, v. 60, n. 5, pp. 1457–1473.
- CAMARGO, A. W., 2019, *Análise da inversão da forma de onda completa pelo método do lagrangiano aumentado*. Tese de Doutorado, [sn].
- CARCIONE, J. M., HERMAN, G. C., TEN KROODE, A., 2002, “Seismic modeling”, *Geophysics*, v. 67, n. 4, pp. 1304–1325.
- CARNEIRO, M. D. S. R., PEREIRA-DIAS, B., SOARES FILHO, D. M., et al., 2018, “On the Scaling of the Update Direction for Multi-parameter Full Waveform Inversion: Applications to 2D Acoustic and Elastic Cases”, *Pure and Applied Geophysics*, v. 175, n. 1, pp. 217–241.
- CARNEIRO, M. D. S., 2017, *Estratégias de Inversão de multiparametros utilizando a equação de ondas completa da elastodinâmica*. Tese de Doutorado, Ph. D Thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.
- CHEN, C., 1995, *Smoothing methods in mathematical programming*. Relatório técnico.

- CHEN, F., PETER, D., 2018, “A misfit function based on entropy regularized optimal transport for full-waveform inversion”. In: *SEG Technical Program Expanded Abstracts 2018*, Society of Exploration Geophysicists, pp. 1314–1318.
- CHEN, X., 2000, “Smoothing methods for complementarity problems and their applications: a survey”, *Journal of the Operations Research Society of Japan*, v. 43, n. 1, pp. 32–47.
- CHEN, X., 2012, “Smoothing methods for nonsmooth, nonconvex minimization”, *Mathematical programming*, v. 134, n. 1, pp. 71–99.
- CHIZAT, L., PEYRÉ, G., SCHMITZER, B., et al., 2018, “Scaling algorithms for unbalanced optimal transport problems”, *Mathematics of Computation*, v. 87, n. 314, pp. 2563–2609.
- COVER, T. M., 1999, *Elements of information theory*. John Wiley & Sons.
- CRASE, E., PICA, A., NOBLE, M., et al., 1990, “Robust elastic nonlinear waveform inversion: Application to real data”, *Geophysics*, v. 55, n. 5, pp. 527–538.
- CRUZ, D. S., DE ARAÚJO, J. M., DA SILVA, C. C. N., 2019, “Dynamic Relative Entropy Regularization for Full Waveform Inversion”, .
- CRUZ, D. S., DE ARAÚJO, J. M., DA COSTA, C. A., et al., 2021, “Adding Prior Information in FWI through Relative Entropy”, *Entropy*, v. 23, n. 5, pp. 599.
- CSISZÁR, I., 1967, “Information-type measures of difference of probability distributions and indirect observation”, *studia scientiarum Mathematicarum Hungarica*, v. 2, pp. 229–318.
- DA COSTA, F. T., KARSOU, A. A., DA SILVA, R. C. M., et al., 2019, “Inversao do campo de onda completo (FWI) ao modelo conceitual do Campo de Buzios”, *Nome da Revista*, v. X, n. Y, pp. XX–YY. doi: XXXXXX.
- DA SILVA, S. L. E., CARVALHO, P. T., DA COSTA, C. A., et al., 2019, “Misfit function for full waveform inversion based on Shannon entropy for deeper velocity model updates”. In: *SEG International Exposition and Annual Meeting*, p. D033S055R001. SEG.
- DEMANET, L., 2016, “Waves and Imaging Class notes-18.325”, .

- DONG, W., BAO, A., SHAH, H., 1984, “Use of maximum entropy principle in earthquake recurrence relationships”, *Bulletin of the Seismological Society of America*, v. 74, n. 2, pp. 725–737.
- FICHTNER, A., 2010, *Full seismic waveform modelling and inversion*. Springer Science & Business Media.
- GUITTON, A., SYMES, W. W., 2003, “Robust inversion of seismic data using the Huber norm”, *Geophysics*, v. 68, n. 4, pp. 1310–1319.
- GULL, S., SKILLING, J., ROBERTS, J., 1984, “Indirect Imaging”. In: *Proc. IAU/URSI Symp.*, ed. JA Roberts, Cambridge Univ. Press, p. 267.
- HA, T., CHUNG, W., SHIN, C., 2009, “Waveform inversion using a back-propagation algorithm and a Huber function norm”, *Geophysics*, v. 74, n. 3, pp. R15–R24.
- HAYKIN, S., 2006, *Nonlinear methods of spectral analysis*, v. 34. Springer Science & Business Media.
- IEA. “world Energy Outlook 2022”. .
- JAYNES, E. T., 1957, “Information theory and statistical mechanics”, *Physical review*, v. 106, n. 4, pp. 620.
- JENSEN, J. L. W. V., 1906, “Sur les fonctions convexes et les inégalités entre les valeurs moyennes”, *Acta mathematica*, v. 30, n. 1, pp. 175–193.
- JI, S., ZHANG, H., WANG, Y., et al., 2019, “Three-dimensional inversion of full magnetic gradient tensor data based on hybrid regularization method”, *Geophysical Prospecting*, v. 67, n. 1, pp. 226–261.
- KEAREY, P., BROOKS, M., HILL, I., 2002, *An introduction to geophysical exploration*, v. 4. John Wiley & Sons.
- KULLBACK, S., 1959, “Information Theory and Statistics”, *Wiley, New York*.
- KULLBACK, S., LEIBLER, R. A., 1951, “On information and sufficiency”, *The annals of mathematical statistics*, v. 22, n. 1, pp. 79–86.
- LANTÉRI, H., 2020, “Divergences. Scale invariant Divergences. Applications to linear inverse problems. NMF Blind deconvolution”, *arXiv preprint arXiv:2003.01411*.

- LE, J., 1994, “Local estimate of surface Laplacian derivation on a realistically shaped scalp surface and its performance on noisy data.” *Electroencephalography and clinical neurophysiology*.
- LI, D., LAMOUREUX, M. P., LIAO, W., 2022, “Application of an unbalanced optimal transport distance and a mixed L1/Wasserstein distance to full waveform inversion”, *Geophysical Journal International*, v. 230, n. 2, pp. 1338–1357.
- MAININI, E., 2012, “A description of transport cost for signed measures”, *Journal of Mathematical Sciences*, v. 181, pp. 837–855.
- MARTINS, M., 2015, *Estrategias Para inversao do campo de ondas completo associado a fontes sismica reais e virtuais*. Tese de Doutorado, Ph. D Thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.
- MUNIZ, W., RAMOS, F., DE CAMPOS VELHO, H., 2000, “Entropy-and Tikhonov-based regularization techniques applied to the backwards heat equation”, *Computers & mathematics with Applications*, v. 40, n. 8-9, pp. 1071–1084.
- NOCEDAL, J., WRIGHT, S. J., 2006, *Numerical optimization*. Springer.
- NOUNAGNON, J. D., 2016, *Using Kullback-Leibler divergence to analyze the performance of collaborative positioning*. Tese de Doutorado, Virginia Tech.
- ÖSTERREICHER, F., 2002, “Csiszár’s f-divergences-basic properties”, *RGMA Res. Rep. Coll.*
- PLESSIX, R.-E., PERKINS, C., 2010, “Thematic set: Full waveform inversion of a deep water ocean bottom seismometer dataset”, *First Break*, v. 28, n. 4.
- RIETSCH, E., 1988, “THE MAXIMUM ENTROPY APPROACH TO THE INVERSION OF ONE-DIMENSIONAL SEISMOGRAMS¹”, *Geophysical prospecting*, v. 36, n. 4, pp. 365–382.
- ROBINSON, E., TREITEL, S., 1980, “Geophysical signal analysis: Prentice-Hall”, *Englewood Cliffs*.
- RYU, E. K., YIN, W., 2022, *Large-Scale Convex Optimization: Algorithms & Analyses via Monotone Operators*. Cambridge University Press.
- SHANNON, C. E., 1948, “A mathematical theory of communication”, *The Bell system technical journal*, v. 27, n. 3, pp. 379–423.

- SONG, X., XU, Y., DONG, F., 2015, “A hybrid regularization method combining Tikhonov with total variation for electrical resistance tomography”, *Flow Measurement and Instrumentation*, v. 46, pp. 268–275.
- TARANTOLA, A., 1984. “Inversion of seismic reflection data in the acoustic approximation: Geophysics”. .
- TARANTOLA, A., 2005, *Inverse problem theory and methods for model parameter estimation*. SIAM.
- THOMAS, J., 2001, “Fundamentos de engenharia do petróleo, Editora Interciência”, *Petrobrás, Rio de Janeiro*.
- THOMASSEN, E., 2008, *Full-waveform inversion studies*. Tese de Mestrado, Institutt for elektronikk og telekommunikasjon.
- ULRYCH, T. J., BISHOP, T. N., 1975, “Maximum entropy spectral analysis and autoregressive decomposition”, *Reviews of Geophysics*, v. 13, n. 1, pp. 183–200.
- VASCONCELOS, A. G., 2018, *Influencia do padrao de radiacao na inversao da forma completa da onda baseada nos metodos do gradiente*. Tese de Mestrado, Universidade Federal do Rio de Janeiro.
- VIRIEUX, J., OPERTO, S., 2009, “An overview of full-waveform inversion in exploration geophysics”, *Geophysics*, v. 74, n. 6, pp. WCC1–WCC26.
- VIRIEUX, J., ASNAASHARI, A., BROSSIER, R., et al., 2017, “An introduction to full waveform inversion”. In: *Encyclopedia of exploration geophysics*, Society of Exploration Geophysicists, pp. R1–1.
- YANG, Y., 2018, *Optimal transport for seismic inverse problems*. Tese de Doutorado, The University of Texas at Austin.
- YANG, Y., ENGQUIST, B., 2018, “Analysis of optimal transport and related misfit functions in full-waveform inversion”, *Geophysics*, v. 83, n. 1, pp. A7–A12.
- YANG, Y., ENGQUIST, B., SUN, J., et al., 2018, “Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion”, *Geophysics*, v. 83, n. 1, pp. R43–R62.
- YU, M., 2021, *Entropic Unbalanced Optimal Transport: Application to Full-Waveform Inversion and Numerical Illustration*. Tese de Doutorado, Université de Paris.

- ZHANG, J., 2004, “Divergence function, duality, and convex analysis”, *Neural computation*, v. 16, n. 1, pp. 159–195.
- ZHANG, Q., MAO, W., ZHOU, H., et al., 2018, “Hybrid-domain simultaneous-source full waveform inversion without crosstalk noise”, *Geophysical Journal International*, v. 215, n. 3, pp. 1659–1681.
- ZHU, H., LI, S., FOMEL, S., et al., 2016, “A Bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration”, *Geophysics*, v. 81, n. 5, pp. R307–R323.