



Um Tutorial de Implementação da FWI e Avaliação da Influência do Uso da Diagonal da Pseudo-Hessiana

Rafael L. Siqueira¹, Leandro Di Bartolo¹ e Bruno de Souza², ¹ Observatório Nacional, ² UFRJ

Copyright 2019, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation during the 16th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 19-22 August 2019.

Contents of this paper were reviewed by the Technical Committee of the 16th International Congress of the Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of the Brazilian Geophysical Society is prohibited.

Abstract

We present a tutorial of a basic FWI algorithm. It uses for sake of simplicity the steepest-descent method for the optimization. In addition, the use of pseudo-Hessian diagonal is discussed, as well as its influence on the results. From this work, it is shown that a basic FWI algorithm is able to provide high resolution results, especially when using information associated with Hessian.

Introdução

A inversão do campo de onda completo (FWI, do inglês, *full-waveform inversion*) é reconhecida como o melhor método de obtenção de estimativas de modelos de velocidades e outras propriedades. Ela é capaz, em teoria, de obter tais modelos com alta resolução, constituindo-se no estado da arte dos métodos de construção de modelos de velocidade.

Embora proposta há muitas décadas – independentemente por LAILLY (1983) e TARANTOLA (1984) –, o amadurecimento da FWI e sua utilização prática ocorreram bem mais recentemente. Um ponto importante é que a sua utilização prática pela Indústria do Petróleo foi possível somente quando do aprimoramento dos computadores. Uma revisão abrangente sobre a FWI pode ser encontrada em VIRIEUX e OPERTO (2009), onde é traçada a sua trajetória desde a proposição até a maturidade e seu posterior desenvolvimento.

A FWI pode ser caracterizada como de um problema de inversão altamente não linear e mal posto, no qual os dados observados (reais ou sintéticos) são comparados diretamente com os dados modelados. Para isso, utiliza-se de uma função objetivo (e.g., norma L2, norma L1, etc.). Diferentes métodos locais de minimização utilizando o gradiente podem ser considerados no processo, assim como métodos globais. Entretanto, na prática, estes são inviáveis devido ao alto custo computacional. Há ainda os métodos híbridos.

Neste trabalho, apresenta-se um tutorial que visa contribuir para a compreensão do método e, ao mesmo tempo, mostrar a sua efetividade. Também é avaliado a influência da utilização de informações da Hessiana.

Resumo da FWI e algoritmo

Um elemento fundamental no algoritmo de FWI é a função objetivo. Ela é utilizada para medir a diferença entre o dado observado e o dado modelado. Em uma aquisição *streamer*, apenas o campo de pressão acústica é utilizado para construir a função objetivo, sendo a norma L2 do vetor de resíduo entre o campo de pressão acústico medido e o modelado o exemplo mais simples:

$$E(\mathbf{m}) = \frac{1}{2} \int_0^T \sum_{j=1}^{N_f} \sum_{i=1}^{N_r} \left(p^{obs}(\mathbf{r}_i, \mathbf{r}_j, t) - p^{cal}(\mathbf{r}_i, \mathbf{r}_j, t; \mathbf{m}) \right)^2 dt, \quad (1)$$

onde p^{cal} e p^{obs} , respectivamente, o dado modelado (calculado) e o dado observado. O primeiro somatório é referente à posição das diferentes fontes e o segundo é referente aos receptores.

Existem dois principais métodos de otimização que lidam com esse problema: os métodos globais e os locais. Métodos globais têm eficiência computacional proibitiva em problemas que demandam computação extensiva, como é o caso da FWI. Os métodos locais são mais viáveis do ponto de vista computacional, pois buscam minimizar a função objetivo em torno de um modelo inicial conhecido \mathbf{m}_0 . Cabe ressaltar que tais métodos são conhecidos também sob a denominação de *métodos gradiente*. Um ponto importante no que se refere aos métodos locais, é que eles funcionam de forma mais efetiva quando se considera um modelo inicial perto da solução verdadeira. Os modelos iniciais podem ser obtidos por diferentes técnicas, sendo comumente utilizada a tomografia sísmica para este fim.

Na sequência, é apresentado o método de Newton, a base de todos os *métodos gradiente*. Ele pressupõe a aproximação em série de Taylor da função objetivo até segunda ordem em torno de um modelo inicial. Considerando, então a função objetivo quadrática, é possível mostrar que a atualização do modelo é dada pela seguinte expressão:

$$\mathbf{m}_{i+1} = \mathbf{m}_i - \mathbf{H}^{-1}(\mathbf{m}_i) \nabla E(\mathbf{m}_i), \quad (2)$$

onde \mathbf{H} e ∇E são, respectivamente, a hessiana e o gradiente da função objetivo, dados por

$$\nabla_v E(\mathbf{m}_0) \equiv \left(\frac{\partial E}{\partial m_1}(\mathbf{m}_0) \quad \frac{\partial E}{\partial m_2}(\mathbf{m}_0) \quad \dots \quad \frac{\partial E}{\partial m_N}(\mathbf{m}_0) \right)^T, \quad (3)$$

$$H(\mathbf{m}_0) \equiv \begin{pmatrix} \frac{\partial^2 E}{\partial m_1 \partial m_1}(\mathbf{m}_0) & \frac{\partial^2 E}{\partial m_1 \partial m_2}(\mathbf{m}_0) & \cdots & \frac{\partial^2 E}{\partial m_1 \partial m_N}(\mathbf{m}_0) \\ \frac{\partial^2 E}{\partial m_2 \partial m_1}(\mathbf{m}_0) & \frac{\partial^2 E}{\partial m_2 \partial m_2}(\mathbf{m}_0) & \cdots & \frac{\partial^2 E}{\partial m_2 \partial m_N}(\mathbf{m}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial m_N \partial m_1}(\mathbf{m}_0) & \frac{\partial^2 E}{\partial m_N \partial m_2}(\mathbf{m}_0) & \cdots & \frac{\partial^2 E}{\partial m_N \partial m_N}(\mathbf{m}_0) \end{pmatrix}. \quad (4)$$

O método de Newton alcança o mínimo da função objetivo (na verdade, sua aproximação em segunda ordem) em uma única iteração. Entretanto, ele é inviável computacionalmente devido ao custo de calcular a matriz Hessiana e sua inversa. Existem diversos métodos iterativos que aproximam a matriz Hessiana de diferentes formas. O cálculo do gradiente também acarreta custo computacional proibitivo caso utilizadas as definições baseadas nas derivadas de Fréchet. Na prática, o se utiliza chamado método adjunto (PLESSIX, 2006) para calcular o gradiente de forma eficiente, como será discutido.

Para o cálculo do modelo direto, utilizamos a equação acústica da onda 2D, i.e.,

$$\frac{\partial^2 p}{\partial t^2} = v^2(r) \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial z^2} \right) + v^2(r) \rho(r) s(\vec{r}, t), \quad (5)$$

onde, p é o campo de pressão, x e z são as coordenadas espaciais (considerando o eixo z para baixo, como de costume na Geofísica), v são as velocidades de onda P do modelo, t é o tempo e s é a fonte. Utilizamos o pulso de Ricker (chapéu Mexicano) como fonte, i.e.,

$$s = -At_d e^{-\pi(f_{cut} t_d)^2}, \quad (6)$$

onde A é a amplitude, f_{cut} é a frequência de corte (máxima frequência presente) e t_d é o tempo defasado:

$$f_c = \frac{f_{cut}}{3\sqrt{\pi}} \quad (7)$$

$$t_d = t - t_0, \quad (8)$$

$$t_0 = \frac{2\sqrt{\pi}}{f_{cut}}. \quad (10)$$

Para a resolução da Eq. (1), utilizamos o Método das Diferenças Finitas, com operadores de diferenças centrais de segunda e quarta ordem, respectivamente, no tempo e no espaço (ALFORD et al., 1974). A equação de marcha explícita é dada por:

$$P_{i,j}^{n+1} = \frac{\Delta t^2 V_{i,j}^2}{12h} \left[-(P_{i-2,j}^n + P_{i,j-2}^n + P_{i+2,j}^n + P_{i,j+2}^n) + 16(P_{i-1,j}^n + P_{i,j-1}^n + P_{i+1,j}^n + P_{i,j+1}^n) - 60P_{i,j}^n \right] + 2P_{i,j}^n - P_{i,j}^{n-1} + \Delta t^2 C_{i,j}^2 \rho_{i,j} S_{i,j}^n, \quad (11)$$

onde n é a variável relacionada ao tempo, h é o espaçamento entre os pontos da malha P^{n+1} , P^n e P^{n-1} são respectivamente os campos de pressão nos tempos futuro, presente e passado.

Para tentar driblar o problema do *cycle skipping*, foi proposto um método, hoje um dos aspectos chave da inversão FWI, conhecido como estratégia multiescala (VIRIEUX e OPERTO, 2009). Tal método consiste na utilização de frequências progressivamente maiores, fazendo com que o grau de complexidade da função objetivo aumente progressivamente e minimizando a chance de que seja atingido um mínimo local (Fig. 1).

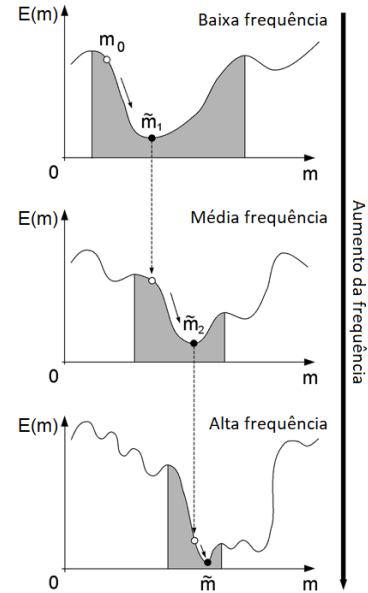


Figura 1. A estratégia Multi-escala para evitar o *cycle-skipping*. Modificado de FICHTNER A. (2011).

Um código (em Algoritmo 1) de FWI básico é apresentado abaixo, discutindo-se seus principais elementos nos próximos parágrafos.

Dados de entrada: (L5) (na linha 5) é realizada leitura do sismograma (dado de campo, pode ser também um dado sintético gerado por um modelo conhecido ou um *benchmark*) e (L1) do modelo de velocidades inicial.

Laço de frequência: (L2-24) laço responsável pela seleção dos dados, começando das frequências mais baixas para as mais altas. Este laço é responsável, portanto, pela estratégia multiescala. No caso de dados reais, é necessário filtrar os dados, enquanto, no caso sintético, para simplificar o código, as frequências foram selecionadas diretamente através da variação da frequência da fonte (variando o f_{cut}).

```

1 Ler modelo inicial
2 Laço de frequência
3 Laço de iteração (it=1,Tit)
4   Laço de tiros (t1=1,Nt)
5     Ler dado do tiro t1
6     Rotina_modelagem(mi)
7     Cálculo do resíduo
8     Calcula E(mi)
9     Calcula grad
10    Fim do laço de tiros
11    Calcula passo (p)
12    Rotina Atualiza_mod(p, grad, mi, mi+1)
13    Laço de tiros (t2=1,Nt)
14      Ler dado do tiro t2
15      Rotina_modelagem(mi+1)
16      Calcula E(mi+1)
17    Fim do laço de tiros
18    Se E(mi+1) ≥ E(mi)
19      passo = passo / 2
20      volta para L12
21    Fim se
22    mi ← mi+1
23 Fim de laço de iteração
24 Fim do laço de frequência

```

Algoritmo 1. Algoritmo básico de FWI.

Laço de iteração: (L3-23) Esse laço é responsável pelo processo iterativo, relacionado ao uso do método de otimização local. Lembrando que métodos locais visam encontrar um mínimo da função objetivo na vizinhança de um modelo inicial. O Método de Newton, apresentado anteriormente, é o único *método gradiente* capaz de obter o mínimo local da aproximação da função objetivo (aproximada até segunda ordem) em apenas um passo para cada iteração. No algoritmo apresentado, utilizamos o método do gradiente descendente, como será detalhado mais adiante.

Laços de tiro: (L4-10 e L13-17) o código apresentado, baseia-se na realização das etapas necessárias para a inversão tiro a tiro, sendo o gradiente e a função objetivo obtidas acumulando-se o resultado de cada um dos tiros. É importante se atentar para o fato de que a geometria necessita ser carregada corretamente. Por simplicidade, pode-se considerar receptores fixos, sendo um em cada ponto do *grid* de diferenças finitas, e um deslocamento da fonte sísmica, tiro a tiro.

Rotina_modelagem: (L6 e L15) resolve a modelagem direta para cada tiro sobre o modelo de entrada (no nosso caso o modelo de velocidades), ou seja, os

sismogramas (dado calculado). Repare que, no primeiro laço de tiro, é utilizado o modelo m_i enquanto que no segundo, é utilizado o modelo m_{i+1} .

Cálculo do resíduo: (L7) calcula o resíduo fazendo a diferença entre o dado observado e o dado calculado, como mostrado na Fig. 2 para um tiro próximo a borda direita.

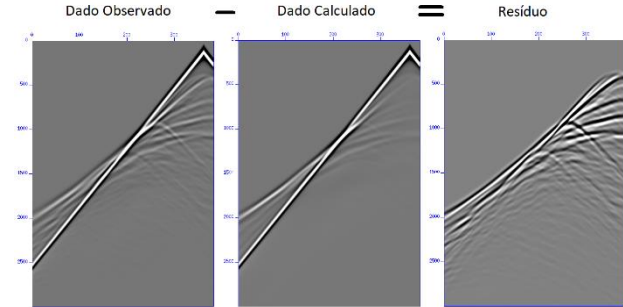


Figura 2. Exemplo de resíduo

Cálculos de E: (L8 e L16) a função objetivo (Eq. 1) é calculada utilizando-se a norma L2 entre o dado de entrada e o dado calculado. Lembrando que E quantifica o quão distante o dado calculado está do dado observado. Através dessa análise podemos avaliar se o modelo está convergindo ou não para um modelo mais perto do real, para decidir se o modelo deve ser atualizado ou não (caso em que o passo é alterado).

Calcula grad: (L9) Na prática, utiliza-se o método adjunto para o cálculo do gradiente de forma eficiente. O gradiente é calculado através da correlação entre o campo de ondas propagado a partir da fonte sísmica e a propagação reversa no tempo dos resíduos a partir dos receptores. Na verdade, o processo é análogo à Migração Reversa no Tempo.

O gradiente é então obtido com a correlação entre a derivada do campo de pressão e o campo reverso residual (caso da velocidade):

$$\frac{\partial E}{\partial v_l} = \frac{2}{v^3(\mathbf{r}_l)} \frac{\partial^2 p^{(f)}}{\partial t^2}(\mathbf{r}_l, t) \otimes p^{(f)\dagger}(\mathbf{r}_l, t) \Big|_{lag_0},$$

onde $p^{(f)}(\mathbf{r}_l, t)$, $p^{(f)\dagger}(\mathbf{r}_l, t)$ são os campos, respectivamente, da modelagem direta, a partir da fonte sísmica, e o campo adjunto. O campo adjunto nada mais é do que o resíduo entre o campo observado e o modelado propagado reversamente no tempo. Note que a correlação cruzada com atraso nulo (*lag zero*) nada mais é do que um produto escalar.

Calcula passo: (L11) O passo é um fator que escala o gradiente a fim de atualizar coerentemente os valores do modelo a ser minimizado. Existem formas mais eficientes para encontrar o valor do passo inicial, nesse tutorial o passo foi calculado da seguinte forma:

$$\alpha = \frac{\maxval}{\maxgrad},$$

sendo α o passo, \maxval o valor máximo da atualização (escolhido pelo usuário) e \maxgrad o valor máximo absoluto do gradiente calculado. A Figura 3 ilustra a influência do passo no gradiente.

Rotina Atualiza_mod: (L12) existem vários métodos que fazem a atualização do modelo através do uso do gradiente e da Hessiana (os métodos de Newton,

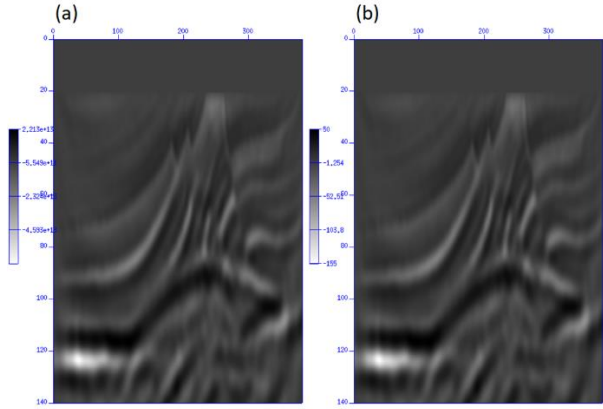


Figura 3. (a) gradiente, (b) gradiente escalado pelo passo, escolhido como 50.

gradiente descendente, gradientes conjugados, etc.). Neste tutorial é mostrado por simplicidade o método do gradiente descendente. Esse método minimiza a função objetivo utilizando um esquema iterativo, onde em cada iteração (laço de iteração, L3-23) se atualiza o modelo na direção negativa do gradiente (direção de declive máximo), utilizando a seguinte expressão:

$$\mathbf{v}_{i+1} = \mathbf{v}_i - \alpha \nabla E(\mathbf{v}_i),$$

sendo \mathbf{v}_{i+1} o modelo de velocidade atualizado e \mathbf{v}_i o modelo a ser atualizado, α o passo e ∇E o gradiente da função objetivo. A Figura 4 ilustra a atualização.

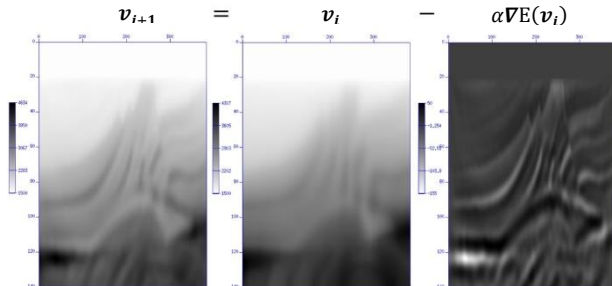


Figura 4. Processo de atualização das velocidades.

Avaliação da convergência: (L18-21) faz-se então uma avaliação para o valor da função objetivo para \mathbf{v}_{i+1} (modelo atualizado) e, se o valor for maior que para \mathbf{v}_i (modelo anterior), divide-se o passo por 2 e refaz a atualização do modelo (L12) utilizando o novo passo. Caso contrário mantém-se o modelo atualizado e passa para a próxima iteração.

Utilização da Pseudo-Hessiana

Uma solução para melhor escalar os parâmetros do modelo é levando em consideração a contribuição da Hessiana. Carneiro et al., (2018) mostra que a escala do gradiente é altamente dependente da escolha do modelo de parametrização e que o uso da pseudo-Hessiana como condicionamento torna as atualizações independentes da escolha da parametrização. Além de escalar melhor os parâmetros, a pseudo-Hessiana

proposta por SHIN et al. (2001) traz um fator de iluminação do modelo, discutido no seu artigo.

A aproximação da diagonal da pseudo-Hessiana proposta por SHIN et al. (2001) é dada por,

$$\tilde{H}_{a,v} = \sum_{s \in SRC} \int_T dt |f_{s,v}^{virt}|^2,$$

onde $f_{s,v}^{virt}$ é a fonte virtual e é dada, para o caso acústico em velocidade, por:

$$f_v^{virt} = -\frac{\partial L}{\partial v}(p) = -\frac{2}{v^3} \ddot{p}.$$

$L(p)$ é o operador direto para onda acústica. Consequentemente, a direção de atualização é dada por:

$$\tilde{H}_{a,v}^{-1} \nabla_v E = \frac{-\frac{2}{v^3} \int_T dt p^\dagger \ddot{p}}{\left(-\frac{2}{v^3}\right)^2 \int_T dt \ddot{p} \ddot{p}} = -\frac{v^3 \int_T dt p^\dagger \ddot{p}}{2 \int_T dt \ddot{p} \ddot{p}}$$

Desta maneira a atualização do modelo de velocidades é escrita como:

$$\mathbf{v}_{i+1} = \mathbf{v}_i + \alpha \frac{v_i^3 \int_T dt p^\dagger \ddot{p}}{2 \int_T dt \ddot{p} \ddot{p}},$$

onde $\int_T dt \ddot{p} \ddot{p}$ é o termo de autocorrelação do campo p .

Resultados

O algoritmo apresentado acima foi aplicado ao modelo sintético Marmousi de 383x141 pontos, no qual se aplicou uma camada de água acima (Figura 5). Um algoritmo utilizando o método de diferenças finitas foi utilizado para gerar os dados observados (sintéticos).

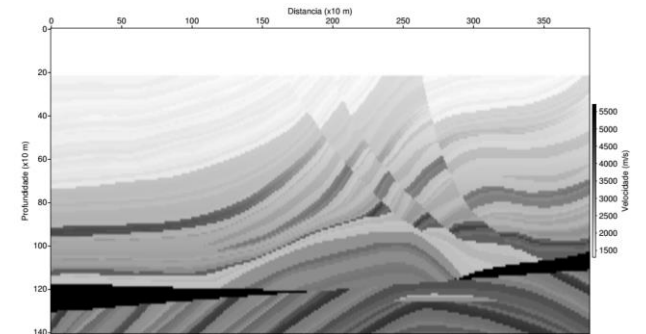


Figura 5. Modelo de velocidades Marmousi.

Neste exemplo, foram dados 19 tiros, ao longo do modelo (espaçamento entre tiros de 20 pontos). O campo de pressão foi registrado em 383 receptores distribuídos ao longo da direção horizontal do modelo.

A inversão necessita do dado observado como entrada e de um modelo inicial. O modelo inicial (Figura 6) foi obtido suavizando o modelo verdadeiro (Figura 5). Essa suavização foi obtida utilizando a função smooth2 do Seismic Unix (via mínimos quadrados amortecido), desenvolvido pela CWP da Colorado School of Mines.

A Tabela 1 contém os parâmetros de inversão utilizados. O resultado a seguir ilustra o modelo final devido a inversão (Figura 7).

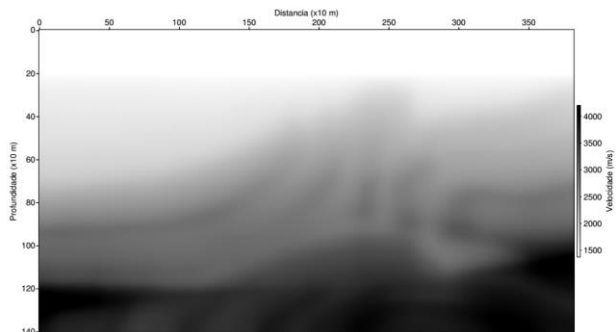


Figura 6. Modelo de velocidades inicial para inversão.

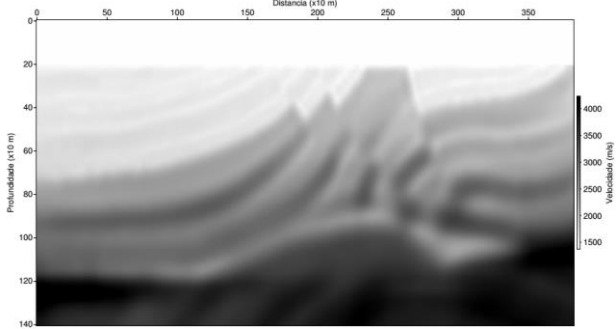


Figura 7. Resultado do modelo estimado pela inversão.

Tabela 1. Parâmetros de modelagem e inversão

Dimensões do modelo de velocidades	383×141 pontos
Espaçamento da malha	10 metros
passos de tempo	1 ms
Número total de passos de tempo	3000
Frequência inicial	3 Hz
Frequência final	30 Hz
Intervalo das frequências	3 Hz
Número total de tiros	19
Espaçamento entre tiros	20 metros
Número de receptores	383
Espaçamento entre receptores	1 ponto

Visualmente, pode-se perceber claramente a melhora da qualidade da imagem final quando utilizada a pseudo-Hessiana. Pode-se também avaliar os resultados através do comportamento de perfis de velocidades entre o modelo verdadeiro e os modelos estimados pela inversão. A Figura 9 mostra o gráfico de comparações de perfis extraídos dos modelos na posição horizontal de 2500 metros.

Conclusões

Neste trabalho, apresentou-se um tutorial básico da inversão do campo de onda completo. Um exemplo utilizando o modelo Marmousi foi apresentado, com e sem a utilização da pseudo-Hessiana. O resultado da inversão FWI, baseado no algoritmo apresentado, se mostrou bastante satisfatório, sendo o exemplo que utiliza a diagonal da pseudo-Hessiana claramente superior. Neste caso, houve uma melhora significativa na iluminação do modelo em profundidades mais altas.

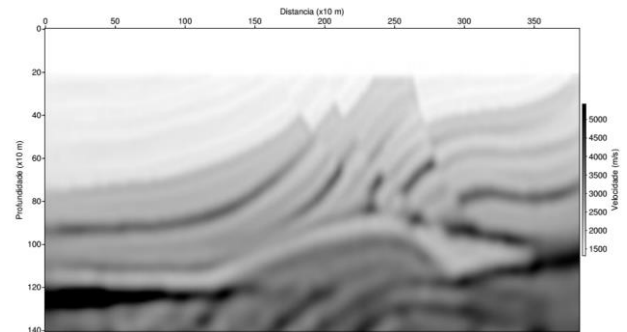


Figura 8. Resultado do modelo estimado com utilização da diagonal da pseudo-Hessiana.

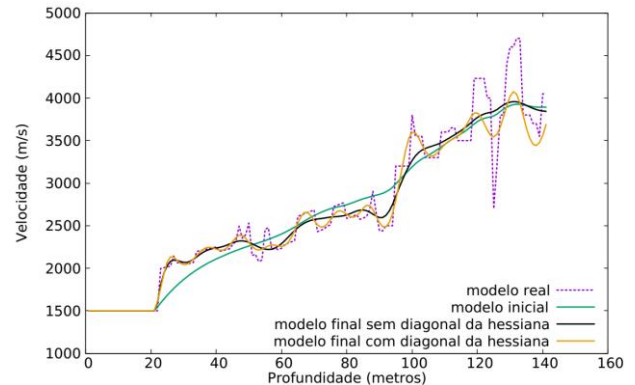


Figura 9. Perfis de velocidade.

Agradecimentos

- CAPES pela bolsa de doutorado,
- Observatório Nacional
- Universidade Federal do Rio de Janeiro

Referências

- ALFORD, R. M., KELLY, K. R., BOORE, D. M. (1974). Accuracy of finite-difference modeling of the acoustic wave equation. *Geophysics*, v.39, n.6, 834-842.
- CARNEIRO, M., DIAS, B., FILHO, D.M.S, LANDAU, L. (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*, 175, 217–241
- FICHTNER, A. (2011). *Full seismic waveform modelling and inversion*. Springer Science & Business Media.
- LAILLY, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In: J.B. Bednar, R. Redner, E. Robinson, & A. Weglein (Eds.), *Conference on inverse scattering: Theory and applications*. Philadelphia: Soc. Industr. Appl. Math.
- PLESSIX, R. E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, 167 (2), 495–503.
- SHIN, C., JANG, S., MIN, D. J., (2001), Improved amplitude preservation for prestack depth migration by inverse scattering theory", *Geophysical prospecting*, v. 49, n. 5, pp. 592-606.

TARANTOLA, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8), 1259–1266.

VIRIEUX, J., & OPERTO, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6), WCC1–WCC26.