

Inversão do campo de onda acústica: Comparação de métodos de otimização matemática

Denis Lage Ferreira da Silva¹; Marco Antonio Cetale Santos¹; Jorge Leonardo Martins²; Roger Matsumoto Moreira³; Djalma Manoel Soares Filho⁴

¹Departamento de Geologia e Geofísica, Instituto de Geociências, Universidade Federal Fluminense

²Coordenação da área de Geofísica, Observatório Nacional, Ministério da Ciência, Tecnologia & Inovação

³Escola de Engenharia, Universidade Federal Fluminense

⁴CENPES/Petrobras

Copyright 2014, SBGf - Sociedade Brasileira de Geofísica.

Este texto foi preparado para a apresentação no VI Simpósio Brasileiro de Geofísica, Porto Alegre, 14 a 16 de outubro de 2014. Seu conteúdo foi revisado pelo Comitê Técnico do VI SimBGf, mas não necessariamente representa a opinião da SBGf ou de seus associados. É proibida a reprodução total ou parcial deste material para propósitos comerciais sem prévia autorização da SBGf.

Abstract

This work compares different types of numerical optimization methods used in full waveform inversion (FWI). All methods are tested using the Marmousi model. Sequential inversions of increasing frequency mitigates the non-linearities of the process. The methods are compared in terms of normalized total run time, total number of iterations and mean square errors of the resulting synthograms. The convergence of each frequency component is further analysed.

Introdução

O desenvolvimento de técnicas de inversão do campo de onda completa (Full waveform inversion, FWI) é considerado um passo importante em direção ao melhor entendimento das propriedades da subsuperfície. A FWI é um procedimento complexo de ajuste de dados baseado na modelagem de onda completa visando extrair informações quantitativas de sismogramas.

Tarantola (1984) inferiu modelos de subsuperfície de alta resolução utilizando técnicas de minimização do resíduo dos sismogramas observados e simulados. Esta técnica consiste em encontrar uma direção de busca no subespaço dos modelos a partir de um modelo inicial. Em seguida é realizada uma segunda busca na direção calculada por um modelo que reduza o resíduo dos sismogramas. Este procedimento é repetido até que haja convergência. Existem diversos modos de encontrar a direção de busca da FWI (Fletcher, 1987), cada um com vantagens e desvantagens. Nos últimos anos, pesquisadores vem buscando formas de implementar a FWI de maneira robusta e eficiente computacionalmente.

Tarantola (1984) Inicialmente, a FWI foi implementada realizando a busca através da direção contrária ao gradiente da função objetivo (método do gradiente descendente) (Tarantola, 1984). Esta técnica, embora de fácil implementação, necessita de um número elevado de iterações para atingir a convergência. Hu et al. (2011) utiliza o método do gradiente conjugado (MGC) apresentada por Hestenes and Stiefel (1952). A facilidade de implementação para problemas com diversos parâmetros e a convergência mais rápida que o gradiente descendente torna o MGC bastante popular como otimizador na FWI. Finalmente, uma aproximação de segunda ordem da função objetivo dá origem à otimização de Newton. Desta forma, a matriz das derivadas de segunda ordem

da função objetivo (matriz Hessiana) é utilizada no cálculo da direção de busca do modelo. No caso da FWI, onde o número de parâmetros é muito grande (>10000), esta matriz não pode ser calculada de forma eficiente, inviabilizando o método. Contudo existem formas de encontrar a Hessiana de forma aproximada (Dennis and Morée, 1977). Nocedal and Wright (2006) apresentam formas implícitas de calcular as atualizações do modelo sem a necessidade de armazenar a matriz Hessiana. Viabilizando a aplicação na FWI. Alguns exemplos de aplicações de métodos baseados na técnica de Newton estão em Pratt et al. (1998), Brossier et al. (2009b), Brossier et al. (2009a), Métivier et al. (2013).

A FWI é um problema não linear malposto. Além do método de otimização numérica utilizado, outros desafios devem ser atacados visando mitigar esses problemas. Contribuem neste sentido, a inversão sequencial de frequências específicas no domínio da frequência, ou de comprimentos de ondas específicos no domínio do tempo. Técnicas de regularização e pré-condicionamento dos dados de entrada também ajudam em tornar o problema mais robusto. Virieux and Operto (2009) apresenta um resumo completo de todas as etapas do processo de FWI.

Este trabalho apresenta uma comparação entre métodos de otimização matemática utilizados na FWI. Para estas comparações, levou-se em consideração o número de iterações necessária para a convergência e o valor total da função objetivo encontrada para cada método.

Metodologia

A inversão do campo de onda completa (FWI) é um processo de otimização matemática não linear visando encontrar um modelo de velocidades sintético \hat{c} que minimiza uma função objetivo. Esta função utiliza uma métrica entre um sismograma sintético $\mathbf{u}_c(\hat{c})$ e um sismograma observado \mathbf{u}_o . Neste trabalho a métrica utilizada foi a norma l_2 da diferença entre os sismogramas.

Sendo assim, busca-se minimizar a seguinte função objetivo (Pratt, 1999)

$$\chi(\hat{c}) = \frac{1}{2} \mathbf{r}^T \mathbf{r}^* \quad (1)$$

onde $\mathbf{r} = \mathbf{u}_c - \mathbf{u}_o$ é o erro residual dos sismogramas.

O cálculo do modelo de velocidades é feito no domínio das frequências. Ao aumentar progressivamente a complexidade do modelo a ser invertido, ou seja, incluir as frequências mais altas em etapas, diminui-se a importância do modelo inicial e melhora a convergência global da FWI. Desta forma, o modelo inicial é utilizado para a inversão da componente de frequência mais baixa do

dado. A saída desta etapa é utilizada como entrada para a frequência subsequente e assim sucessivamente. A frequência inicial e a final dependem da fonte utilizada. Neste trabalho foi utilizado $\omega_i = 0.5\omega_c$ e $\omega_f = 2.5\omega_c$ onde ω_c é a frequência central da fonte.

Todos os métodos estudados seguem a seguinte sequência

```

for  $\omega = \omega_i : \omega_f$ 
  Para um dado  $\mathbf{c}_0$ 
  Calcula-se  $\chi_0$  e  $\nabla_c \chi_0$ 
  While  $(\chi_{k-1} - \chi_k) \geq \varepsilon$ 
    Calcula a direção de busca  $\mathbf{s}_k$ 
    Encontra  $\alpha_k$  satisfatório
     $\mathbf{c}_{k+1} \leftarrow \mathbf{c}_k + \alpha_k \mathbf{s}_k$ 
     $k = k + 1$ 
  end
   $\mathbf{c}_0 \leftarrow \mathbf{c}_k$ 
end
    
```

As diferenças residem no cálculo da direção de busca \mathbf{s} e do passo de iteração α .

Para a inversão do campo de onda acústica, o cálculo do gradiente da função objetivo depende da equação da onda acústica no domínio da frequência (Pratt et al., 1998) (2).

$$\mathbf{u}(\omega) = \mathbf{S}^{-1}(\omega)\mathbf{f}(\omega) \quad (2)$$

Onde $\mathbf{S} = \mathbf{K} - \omega^2\mathbf{M} + i\omega\mathbf{C}$. Neste caso, o gradiente da função objetivo é:

$$\nabla_c \chi = \text{Re} \left\{ \mathbf{F}^t [\mathbf{S}^{-1}]^t \mathbf{r}^* \right\} \quad (3)$$

Onde \mathbf{F} é uma matriz cujas colunas são os termos de fontes virtuais para cada um dos m parâmetros.

$$\mathbf{F} = [\mathbf{f}^{(1)} \quad \mathbf{f}^{(2)} \quad \dots \quad \mathbf{f}^{(m)}] \quad (4)$$

$$\mathbf{f}^{(i)} = -\frac{\partial \mathbf{S}}{\partial p_i} \quad (5)$$

O cálculo de α pode ser feito de diversas maneiras. Neste trabalho foram utilizadas duas formas. A primeira consiste em encontrar α que atenda as condições de Wolfe (Fletcher, 1987) (6).

$$\begin{aligned} \chi(\mathbf{c}_k + \alpha_k \mathbf{s}_k) &\leq \chi(\mathbf{c}_k) + \sigma_1 \alpha_k \nabla \chi_k^T \mathbf{s}_k \\ |\nabla \chi(\mathbf{c}_k + \alpha_k \mathbf{s}_k)^T \mathbf{s}_k| &\leq -\sigma_2 \nabla \chi_k^T \mathbf{s}_k \end{aligned} \quad (6)$$

Sendo $0 < \sigma_1 \leq \sigma_2 < 1/2$. O algoritmo que realiza a busca na direção \mathbf{s}_k pelo α_k que atende estas condições é chamado de algoritmo de *line search*. Este algoritmo implica em recalculer a função objetivo algumas vezes na busca do passo ótimo, o que aumenta consideravelmente o custo computacional para cada passo de iteração.

A segunda forma consiste em encontrar um α que reduz a função objetivo, ou seja, que atenda a seguinte condição:

$$\chi_{k+1}(\mathbf{c}_k + \alpha_k \mathbf{s}_k) < \chi_k(\mathbf{c}_k) \quad (7)$$

Ao utilizar como valor inicial de α_k um valor que implique em uma variação máxima definida no modelo de velocidades, caso este valor não atenda a condição 7, reduz-se α_k até que 7 seja atendida. Este algoritmo diminui consideravelmente o número de vezes que o método precisa ser calculado, aumentando a velocidade global do método. Desta forma, no decorrer do texto, os métodos que utilizem esta forma de cálculo de α serão chamados de **rápidos**.

Os algoritmos implementados neste trabalho seguem três principais linhas de cálculo da direção de busca \mathbf{s}_k . São elas:

Gradiente Descendente No método do gradiente descendente, a direção descendente é o negativo do gradiente no ponto de estudo:

$$\mathbf{s}_k = -\nabla_c \chi_k \quad (8)$$

Este método possui a vantagem de ser simples de implementar e de baixo custo computacional para cada iteração. Contudo, para funções mais complexas, o número de iterações necessárias aumenta significativamente, tornando-o pouco eficiente.

Gradiente Conjugado (Hestenes and Stiefel, 1952)

O método do gradiente conjugado é um dos métodos mais utilizados para resolver grandes problemas lineares e não lineares. As características marcantes deste método são a não necessidade de armazenamento de grandes matrizes e a taxa de convergência mais rápida que o método do gradiente descendente. Uma propriedade importante dos vetores conjugados é que podemos minimizar uma função objetivo em no máximo n passos através de minimizações sucessivas ao longo das direções conjugadas, onde n é o número de parâmetros da função. A direção de busca \mathbf{s}_k é calculada da seguinte forma:

```

if  $k = 0$ 
   $\mathbf{s}_k = -\nabla \chi_k$ 
else
   $\beta_{k+1} = \frac{\nabla \chi_{k+1}^T \nabla \chi_{k+1}}{\nabla \chi_k^T \nabla \chi_k}$ 
   $\mathbf{s}_{k+1} = -\nabla \chi_{k+1} + \beta_{k+1} \mathbf{s}_k$ 
end
    
```

Um dos maiores desafios do método gradiente conjugado é a necessidade de calcular com maior precisão o passo de iteração ótimo α_k . Caso α_k não satisfaça certas condições, as direções de busca podem não ser descendentes. Desta forma, torna-se necessário que o cálculo de α_k atenda as condições de Wolfe (6).

Newton No método de Newton, utiliza-se a informação da derivada segunda ordem da função objetivo (Fletcher, 1987). Neste caso a direção de busca passa a ser

$$\mathbf{s}_k = -\mathbf{H}_k^{-1} \nabla \chi_k \quad (9)$$

onde \mathbf{H}_k é a matriz Hessiana.

Este método, embora de convergência muito rápida, possui um alto custo computacional, pois necessita cal-

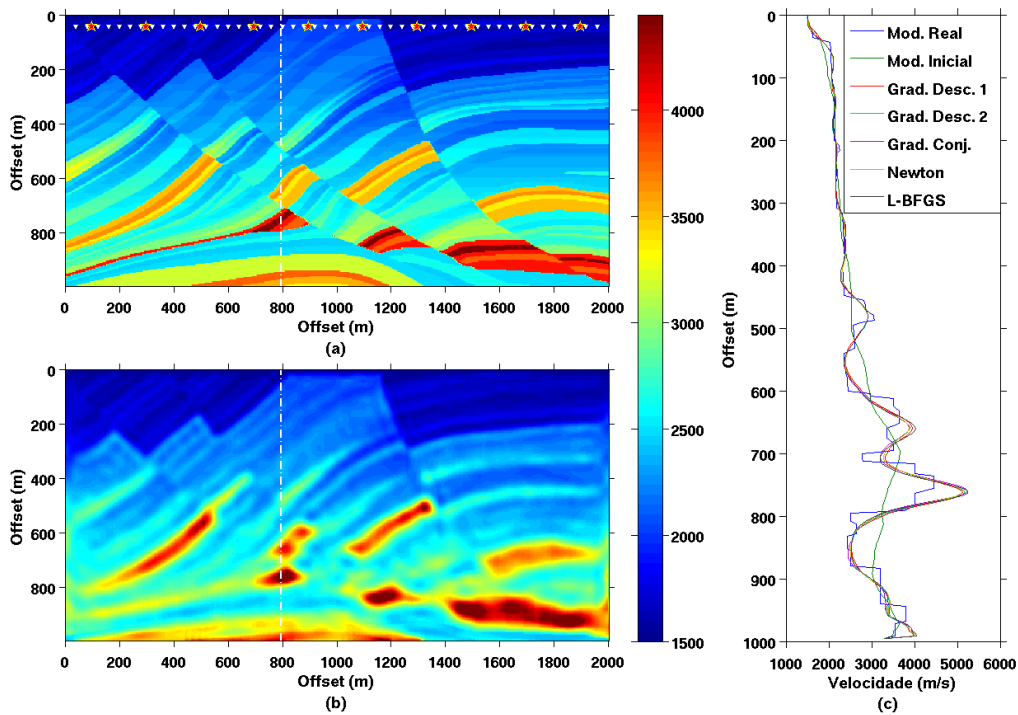


Figura 1: Comparação dos campos de velocidade. (a) - Modelo real. As estrelas vermelhas representam as posições das fontes, os triângulos brancos representam os receptores. (b) - Modelo invertido pelo método L-BFGS. (c) - Comparação de um perfil de velocidades para todos os métodos. O perfil escolhido está representado nos itens (a) e (b) por uma linha tracejada branca.

cular e inverter a matriz Hessiana para cada iteração. Para problemas com muitas variáveis, o cálculo da Hessiana apresenta um desafio em termos de armazenamento. Neste caso, implementações que calculam a inversa da Hessiana necessária para o cálculo de s_k de forma implícita resolvem este problema. Neste trabalho foram implementadas duas variações do método de Newton, ambas descritas em Nocedal and Wright (2006), Newton-CG e L-BFGS. No primeiro a direção de busca é feita aplicando o método do conjugado gradiente na equação de Newton. No segundo a inversa da matriz Hessiana é aproximada utilizando apenas informações das últimas l iterações.

Para os teste realizados, o modelo utilizado foi uma parte do Marmousi de $2004m \times 1000m$. Foram realizados dez simulações com as fontes distribuídas ao longo da profundidade de $40m$ e receptores em todos os pontos nesta mesma profundidade. A discretização espacial foi de $4m \times 4m$. A fonte utilizada foi ricker de $8Hz$. O modelo e a geometria de aquisição estão na Figura 1a. O modelo inicial foi o modelo real suavizado. Para cada método estudado, foi feita uma rodada completa do algoritmo FWI, incluindo sucessivamente as componentes de frequência desde $4Hz$ até $20Hz$. Foram comparados cinco métodos em relação ao tempo de execução relativo ao primeiro método, o número total de iterações executadas e a norma l_2 do erro do sismograma no domínio do tempo. Os métodos testados foram:

1. Gradiente Descendente com line search
2. Gradiente Descendente rápido
3. Gradiente Conjugado com line search
4. Newton-CG com line search

5. L-BFGS

Resultados

A Figura 1 apresenta o resultado da inversão. A Figura 1a e 1b é o modelo real e o resultado da inversão respectivamente. A Figura 1c mostra a comparação de um perfil de velocidades para todos os métodos. O perfil escolhido encontra-se destacado nas Figuras 1a e 1b. Visualmente, todos os métodos estudados apresentaram uma boa aproximação. Percebe-se que as bordas não apresentam um bom resultado. Isto deve-se a falta de dados na aquisição.

A Figura 2 apresenta a comparação dos sismogramas. A Figura 2a e 2b contém um sismograma gerado a partir do modelo real e do modelo invertido respectivamente. A Figura 2c mostra um traço selecionado para todos os métodos. O traço escolhido encontra-se destacado nas Figuras 2a e 2b. Nota-se que todos os métodos são capazes de aproximar com bastante precisão o traço real.

A performance dos métodos encontra-se na Tabela 1 e na Figura 3. A Tabela 1 apresenta o tempo de execução normalizado em relação ao primeiro método, número de iterações totais e o erro médio quadrático. A Figura 3 contém as evoluções dos erros para cada frequência. O valor máximo de cada barra indica o erro inicial e o valor mínimo indica o erro final para cada uma das frequências invertidas. Desta forma, uma barra muito curta, isto é, um erro final elevado, indica uma incapacidade do método de convergir para uma determinada frequência. Também é possível observar a influência que uma boa convergência em uma frequência mais baixa tem nas de-

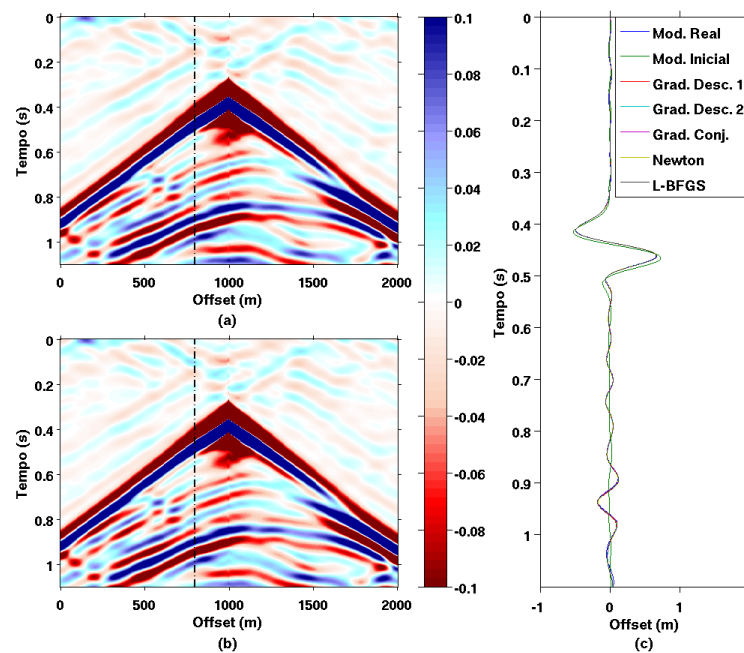


Figura 2: Comparação dos sismogramas. (a) Sismograma obtido pelo campo real. (b) Sismograma obtido pelo campo invertido pelo método L-BFGS. (c) - Comparação de um traço sísmico para todos os métodos. O traço escolhido está representado nos itens (a) e (b) por uma linha tracejada preta.

mais. Quando um método falha na convergência, o erro inicial da frequência seguinte é naturalmente maior, o que indica uma distância maior do mínimo, aumentando o tempo de convergência e o risco de encontrar um mínimo local. Vale ressaltar que é possível identificar a assinatura da banda de frequência da fonte na Figura 3.

O método do gradiente descendente com line search falhou em convergir em diversas frequências. Embora o erro total no sismograma não tenha sido muito elevado, este é um resultado preocupante, pois isto mostra que mesmo em um caso ideal, sem ruído, a convergência não é garantida. Para casos mais complexos, este problema torna-se mais grave. O resultado do gradiente descendente rápido obteve uma resposta bastante interessante, conseguindo uma boa convergência para um tempo total bastante abaixo dos demais. Como o tamanho do passo de busca neste caso não é tão restritiva como nos demais métodos, existe a possibilidade de fuga de mínimos locais através de α maiores, diminuindo a probabilidade de convergir para um mínimo local. Vale ressaltar que o número total de iterações neste caso é maior que os demais. O melhor método em termos de erro de aproximação foi o método do gradiente conjugado (MGC). Em boa parte das frequências, principalmente nas frequências mais próximas da frequência central da fonte, obteve o menor erro final. O erro total naturalmente foi o menor de todos. Em termos do número de iterações e tempo total de execução, não obteve um bom resultado, ficando atrás do gradiente descendente rápido e do L-BFGS no tempo total de execução e atrás deste último em termos de número de iterações. O método do Newton truncado obteve o segundo melhor resultado em termos de erro de convergência. Porém o tempo total de execução não tornaria este método preferível ao MGC. Por último, o L-BFGS não obteve uma boa resposta em termos de erro de convergência. É possí-

vel melhorar este resultado através de regularização e modificando a aproximação inicial da Hessiana (Métivier et al., 2013). O tempo total de execução indica uma alta eficiência computacional na atualização do modelo.

Tabela 1: Performance dos métodos: (1) - Gradiente Descendente com line search. (2) - Gradiente Descendente rápido. (3) - Gradiente Conjugado com line search. (4) - Newton Truncado com line search. (5) - L-BFGS com line search.

Método	Tempo de Execução	Nº de Iterações	Erro
(1)	1,0000	358	19,61
(2)	0,0014	720	18,91
(3)	0,0254	410	17,74
(4)	0,0299	546	18,55
(5)	0,0138	350	20,33

Conclusões

Ao analisar os métodos, percebe-se que todos obtiveram resultados satisfatórios na inversão do campo de velocidades. Todos os sismogramas gerados a partir dos modelos invertidos aproximam os sismogramas reais. O método que obteve o melhor resultado em relação ao erro da inversão foi o método de gradiente conjugado enquanto o método mais rápido foi o gradiente descendente rápido, embora tenha sido o que necessitou de maior número de iterações para convergir. Conclui-se que o método que obteve o melhor resultado foi o gradiente conjugado, mesmo este sendo mais lento. A eficiência do algoritmo de line search e técnicas de

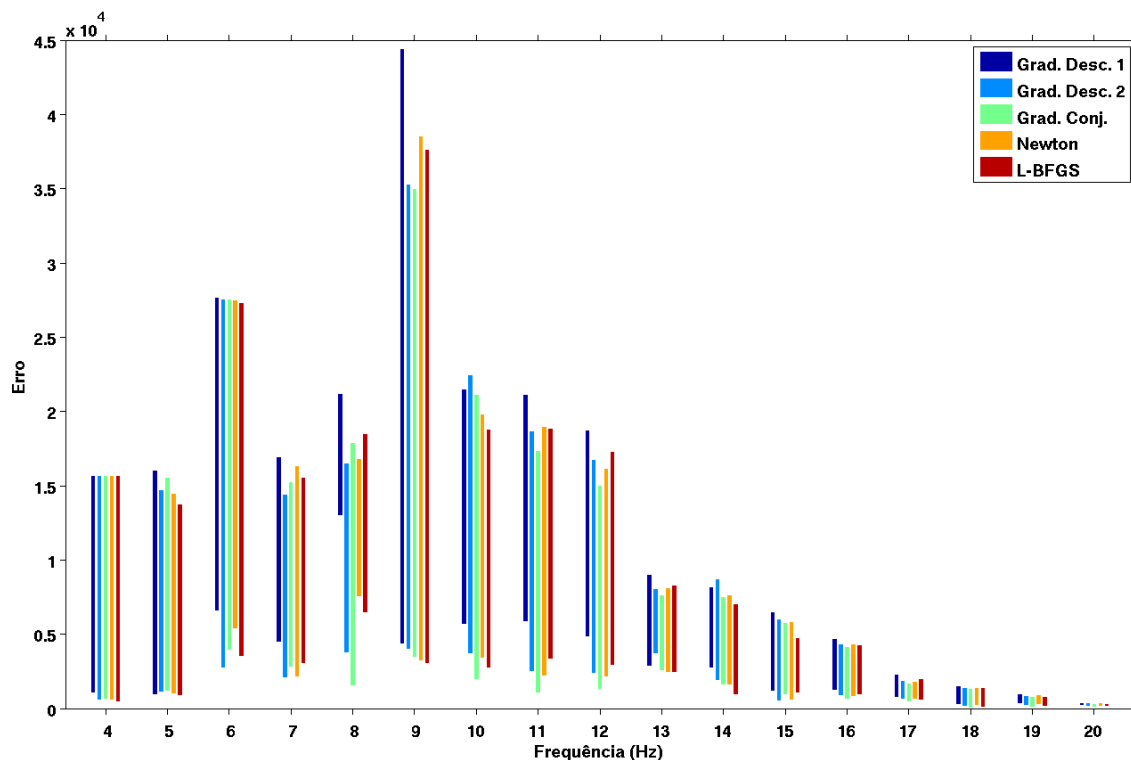


Figura 3: Evolução dos valores da função objetivo para cada frequência. Cada barra mostra os erros iniciais e finais para cada método. Um valor mínimo alto de uma barra indica uma má convergência, ao passo que um valor mínimo baixo indica uma convergência boa. O valor máximo de cada barra indica o erro inicial e demonstra a importância de uma boa aproximação para as etapas anteriores.

pré-condicionamento podem ajudar a torná-lo mais rápido. Vale ressaltar que estes resultados não são definitivos. Outras regularizações estão sendo implementadas. Aplicações em outros modelos, inclusive considerando ruído, também estão sendo testadas.

Agradecimentos

O autor deste trabalho gostaria de agradecer a CAPES pelo auxílio financeiro recebido através de bolsa de doutorado. Os autores gostariam de agradecer a PETROBRAS ao suporte dado ao Laboratório de Imagem Sismico e Inversão Sísmica (ISIS) do Departamento de Geologia e Geofísica da Universidade Federal Fluminense.

Referências

- Brossier, R., S. Operto, and J. Virieux, 2009a, Robust elastic frequency-domain full-waveform inversion using the l_1 norm: *Geophysical Research Letters*, **36**, n/a–n/a.
- , 2009b, Seismic imaging of complex onshore structures by 2d elastic frequency-domain full-waveform inversion: *Geophysics*, **74**, WCC105–WCC118.
- Dennis, and J. J. Moré, 1977, Quasi-newton methods, motivation and theory: *SIAM Review*, **19**, 46–89.
- Fletcher, R., 1987, *Practical methods of optimization*, 2nd ed.: Wiley-Interscience.

- Hestenes, M. R., and E. Stiefel, 1952, Methods of conjugate gradients for solving linear systems: *Journal of Research of the National Bureau of Standards*, **49**, 409+.
- Hu, W., A. Abubakar, T. M. Habashy, and J. Liu, 2011, Preconditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion: *Geophysical Prospecting*, **59**, 477–491.
- Métivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full waveform inversion and the truncated newton method: *SIAM Journal on Scientific Computing*, **35**, B401–B437.
- Nocedal, J., and S. J. Wright, 2006, *Numerical optimization*, 2nd ed.: Springer New York. Springer Series in Operations Research and Financial Engineering.
- Pratt, G., C. Shin, and Hicks, 1998, Gauss–newton and full newton methods in frequency–space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362.
- Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: *Geophysics*, **64**, 888–901.
- Tarantola, A., 1984, Linearized inversion of seismic reflection data: *Geophysical Prospecting*, **32**, 998–1015.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, WCC1–WCC26.