



Avaliação da convergência de algoritmos tipo gradiente aplicados à inversão de forma de onda

Sérgio Adriano Moura Oliveira, Laboratório de Engenharia e Exploração de Petróleo

Copyright 2003, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation at the 8th International Congress of The Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 14-18 September 2003.

Contents of this paper was reviewed by The Technical Committee of The 8th International Congress of The Brazilian Geophysical Society and does not necessarily represents 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

The performance of gradients type algorithms are compared and analyzed when applied to the solution of the non linear waveform inversion. Important aspects of implementation of these algorithms as pre conditioning, scaling and step length are evaluated as well. A numerical example, involving the velocity inversion of a vertically inhomogeneous medium is set up to test the convergence property of the methods.

Introdução

A inversão sísmica de forma de onda permite fazer estimativas quantitativas de parâmetros físicos das camadas geológicas, diretamente do dado registrado em superfície, e portanto aproveita informação tanto de tempo de trânsito, amplitude e fase do sinal. Este método possui um potencial de revelar detalhes da subsuperfície melhor do que as técnicas de processamento atualmente em uso na exploração de petróleo, contudo o custo computacional ainda é o maior empecilho para sua aplicação em larga escala. Isto ocorre pois esta metodologia, que é baseado em técnicas de otimização, necessita de várias etapas de modelagem direta do campo de onda para cálculo da função objetiva e de seu gradiente. Este problema se torna mais crítico ainda se a modelagem for baseada em alguma técnica numérica para solução da equação da onda, como elementos finitos ou diferenças finitas. A inversão de forma de onda pode ser vista como uma seqüência de migrações pré empilhamento (Lailly, 1984).

A função objetiva é definida em função dos parâmetros que se quer recuperar. Ela nada mais é que uma maneira de medir a distância entre os dados observados e os calculados. Sua minimização é o cerne desta metodologia. O que torna factível o uso de técnicas de otimização para este fim, é o fato do vetor gradiente desta função poder ser calculado eficientemente propagando os resíduos (diferença entre o campo calculado e observado) reversamente no tempo e correlacionando-os com o campo calculado, propagado diretamente (ver por exemplo, Tarantola, 1984). De posse da informação do vetor gradiente, pode-se então lançar mão de três algoritmos clássicos de otimização: O gradiente e conjugado gradiente e o quase Newtoniano. Os dois primeiros são de longe, os mais populares dentro da inversão sísmica e a diferença em termos de custo

computacional entre eles é praticamente nula. O quase Newtoniano é, teoricamente, o mais eficiente, mas exige o armazenamento de matrizes extra, o que pode dificultar sua aplicação, caso um elevado número de parâmetros estejam envolvidos no problema. A literatura de inversão sísmica de forma de onda ainda é carente de uma análise comparativa sobre a robustez e velocidade de convergência de tais métodos. Detalhes de implementação de tais algoritmos, mas que são de fundamental importância para o seu funcionamento eficaz, também costumam ser omitidos. Como exemplos, temos o sub problema da definição do tamanho do passo a ser dado a cada iteração e o dimensionamento das variáveis.

Neste trabalho, a comparação entre os métodos é realizada através de um exemplo numérico de inversão de velocidade, onde são analisados quesitos como o uso de pré condicionamento para acelerar a convergência e a melhor estratégia para definição do passo. Tal estudo é importante na medida em que, todo esforço é válido para tornar o inversão sísmica de forma de onda mais eficiente e próxima de ser aplicada rotineiramente na indústria.

Formulação do Problema

A estimativa de propriedades físicas a partir de dados sísmicos é posto como um problema de minimização de uma função objetiva, que mede a distância entre dados observados e calculados. Baseando-se no critério dos mínimos quadrados, esta função pode ser definida da seguinte maneira em um espaço funcional, onde as variáveis espaciais e temporal são encaradas como contínuas:

$$S(m) = \frac{1}{2} \int dx_s \int dx_r \int dt [P_c(x_r, t; x_s) - P_o(x_r, t; x_s)]^2 \quad (1)$$

Onde x_s é a posição da fonte, x_r é a posição do receptor, P_o é o campo observado, P_c é o campo calculado em função do parâmetro $m(x)$, que pode representar densidade, velocidade de onda P ou S, ETC. No presente trabalho, nos concentraremos no caso acústico, portanto P_c daqui por diante significará campo de pressão. Para obtê-lo, nos basearemos na equação acústica da onda:

$$\nabla^2 P_c - \frac{1}{c^2(x)} \frac{\partial^2 P_c}{\partial t^2} = f(x, t; x_s) \quad (2)$$

O parâmetro à ser invertido, é a velocidade de propagação: $m=c(x)$. Para solução da equação anterior, discretizamos o problema e adotamos o método das diferenças finitas. O gradiente do funcional S é calculado aplicando-se a seguinte fórmula:

$$\frac{\partial S}{\partial m} = \frac{1}{c^3(x)} \int dx_s \int_0^T dt \partial_t P_c(x, t; x_s) \partial_t \psi(x, t; x_s) \quad (3)$$

Onde ψ é o resíduo, ou seja, a diferença entre o campo calculado e observado, propagado reversamente no tempo. Para maiores detalhes ver (Kolb, 1986) ou (Tarantola, 1987).

Com a informação do gradiente, podemos recorrer então a três métodos clássicos de otimização para obter os parâmetros que minimizam S , o método gradiente, o conjugado gradiente e o Quase Newtoniano, as vezes conhecidos na literatura como métodos tipo gradiente. Ambos proporcionam a solução iterativa do problema segundo uma fórmula que pode ser posta da seguinte maneira:

$$m_{i+1} = m_i + \alpha_i A_i s_i \quad (4)$$

Onde s é o vetor que representa a direção de busca, α o tamanho do passo e A é uma matriz, positivamente definida, conhecida como matriz de pré condicionamento. A diferença entre tais métodos consiste em como é escolhida a direção de busca e a matriz A . No método gradiente, s possui direção contrária a do vetor gradiente (direção de máxima descendência). No método conjugado gradiente, s_k é obtida através da combinação linear entre as direções s_{k-1}, s_{k-2}, \dots tomadas nas iterações anteriores (Ver por exemplo: Gill et al., 1986). Neste dois últimos métodos A geralmente é mantida constante. No método Quase Newtoniano, esta matriz é modificada à cada iteração de maneira a que se aproxime da inversa da Hessiana, ou seja, da matriz de segundas derivadas. Desta forma espera-se que este atinja a convergência final do método de Newton. A aproximação inicial para A_0 pode ser qualquer matriz positivamente definida, mas na ausência de qualquer informação, geralmente esta é tomada simplesmente como a matriz identidade (Fletcher, 1987).

Outro importante aspecto prático que merece comentário diz respeito ao correto dimensionamento do problema. Isto é recomendável para se evitar que o valor da função objetiva, dos parâmetros ou das componentes do vetor gradiente sejam ou muito grandes ou muito pequenos, o que pode trazer dificuldades de representação em ponto flutuante e piorar a influência dos erros de truncamento. Supondo-se que $(b_k \geq c_k \geq a_k)$, recomendamos a seguinte transformação de variáveis:

$$C_k = \frac{2c_k}{b_k - a_k} - \frac{a_k + b_k}{b_k - a_k} \quad (5)$$

Assim C_k sempre pertencerá ao intervalo $(1 \geq C_k \geq -1)$. Este procedimento se mostrou importante para o método

conjugado gradiente e fundamental para o quase Newtoniano.

Estratégias para definição do passo

A definição do passo α é crucial para o bom funcionamento do algoritmo. O mal dimensionado deste parâmetro pode resultar em um decréscimo aquém do esperado ou então fazer com que a função objetiva assumira um valor maior do que o obtido na iteração anterior. Chamando $S(\alpha) = S(m_k + \alpha_k s_k)$, devemos escolher α onde $dS/d\alpha = 0$. A estratégia aqui adotada é interpolar um polinômio cúbico, conhecido o valor da função e de sua derivada em dois pontos distintos de um intervalo que contenha um ponto de mínimo. O valor do mínimo de tal polinômio será uma estimativa para o passo. Contudo, em um procedimento rigoroso, o valor estimado de α_{\min} deve satisfazer determinadas condições que visam, sobretudo, garantir um decréscimo significativo. Isto pode exigir várias avaliações da função objetiva e sua derivada, em um processo de tentativa e erro. No intuito de tornar o dimensionamento do passo uma tarefa mais simples e menos onerosa computacionalmente, alguns autores sugerem estratégias alternativas. Uma delas é a obtenção de fórmulas simples para cálculo de α baseando-se na linearização de $S(\alpha)$, ver por exemplo (Gauthier et All, 1986). Contudo, tais alternativas devem ser encaradas com reservas, pois o menor esforço de computação para definição do passo, pode ser compensado pela maior número de iterações, caso este venha a ser sub dimensionado. Além disto, esta estratégia pode levar a uma convergência precoce para algum mínimo local.

Exemplo numérico

Para levar a cabo a comparação entre as métodos tipo gradiente, nos basearemos em um teste numérico de inversão de velocidade em um modelo verticalmente heterogêneo. Na Figura 1 é exibido o perfil de velocidade deste modelo, bem como o modelo inicial m_0 usado em todos os métodos. Tal modelo inicial procura preservar a tendência geral de crescimento da velocidade com a profundidade. O dado observado consiste em um sismograma sintético com 48 traços, afastados de 24 metros contendo 800 amostras com intervalo de amostragem de 1.5 milissegundos. A velocidade foi discretizada em duzentas amostras c_k afastadas de 8 metros. Cada uma destas amostras é encarada como uma variável do problema. O primeiro teste demonstra a vantagem do uso do pré condicionamento do vetor gradiente, no caso a matriz A usada para tal fim, é diagonal e representa um ganho para compensar as perdas de amplitude devido ao espalhamento geométrico. Na Figura 2 são exibidas as curvas de convergência que representam a evolução da minimização da função objetiva (Equação 1) à cada passo dos algoritmos gradiente e pré condicionado gradiente. A função objetiva é normalizada pelo erro do modelo inicial, o que é sempre uma boa prática.

O segundo teste, visa comparar a evolução da minimização frente a duas estratégias para cálculo do passo. Uma destas consiste em se adotar um fórmula aproximada para cálculo do passo. A outra, é uma

abordagem mais rigorosa do problema que requer sucessivas etapas de interpolação de polinômio cúbico e avaliações do passo, até que este esteja suficientemente próximo de α_{\min} . Para garantir isto, aqui foram adotadas as seguintes condições:

$$I) \quad S(\alpha) \leq S(0) + \alpha \rho \left. \frac{dS}{d\alpha} \right|_{\alpha=0} \quad \rho \in (0, \frac{1}{2})$$

$$II) \quad \left| \frac{dS}{d\alpha} \right| \leq -\sigma \left. \frac{dS}{d\alpha} \right|_{\alpha=0} \quad \sigma \in (\rho, 1)$$

A condição *I* busca garantir que o passo estará afastado do extremo direito do intervalo, enquanto a condição *II* é que realmente controla a precisão do método, ou seja, quanto menor o valor de σ , mais próximo α estará de α_{\min} . No presente exemplo adotamos $\sigma=0.1$ e $\rho=0.001$. Pelo o que é apresentado na Figura 3, fica patente que fórmulas aproximadas para o cálculo do passo, acabam resultando em uma convergência lenta (linha tracejada) e que acaba estacionando em um valor bem aquém do que o conseguido por um algoritmo mais rigoroso de busca linear de mínimo.

A implementação do algoritmo pré condicionado conjugado gradiente, seguiu a fórmula de Polak e Ribere na obtenção das direções conjugadas, já a atualização de A_k no algoritmo quase Newtoniano, foi implementada segundo a fórmula proposta por Broyden, Fletcher, Goldfarb e Shanno (BFGS), para maiores detalhes ver (Fletcher, 1970). Para a primeira aproximação de A_0 foi adotada a mesma matriz de pré condicionamento usada anteriormente, o que mostrou vantagem com relação a adoção da matriz identidade. Na Figura 3 é exibida a curva de convergência dos três algoritmos até a iteração número 100. Ao contrário do que seria esperado teoricamente, observa-se que os métodos pré condicionados gradiente e conjugado gradiente apresentaram comportamento muito semelhantes, com uma leve vantagem para o pré condicionado gradiente. O Quase Newtoniano, apresentou uma convergência mais lenta a princípio, porém ultrapassou os outros dois a partir da iteração 20. É interessante notar que os métodos pré condicionados gradiente e conjugado gradiente praticamente pararam de convergir a partir da iteração 40 enquanto o quase Newtoniano continuou a fazer progressos. Este último foi testado até a iteração 150, onde resultou em um valor de erro quadrático igual a 10^{-6} . Na Figura 5 é observado o resultado final da inversão obtido pelos métodos pré condicionados gradiente e conjugado gradiente. Ambos, no geral, conseguiram recuperar bem o modelo, mas falharam a partir da penúltima camada, que é justamente onde ocorre uma inversão na tendência de crescimento da velocidade que não é contemplada pelo modelo inicial. O método quase Newtoniano se mostrou mais robusto com relação a isto, conseguindo um ótimo resultado final (Figura 6). Mesmo assim é bom ressaltar, o que é um fato bastante conhecido, que nenhum destes métodos é capaz de funcionar bem caso o modelo inicial se afaste muito da realidade. Este deve fornecer ao menos a informação de baixa frequência do modelo real, caso contrário, qualquer destes três algoritmos acaba ficando estacionado em um mínimo local. De qualquer forma,

ambos os métodos gradientes foram capazes de recuperar a informação de alta frequência bastante rapidamente, em no máximo 10 iterações.

Conclusões

As principais conclusões que podemos tirar dos testes numéricos apresentados neste trabalho são as seguintes; O pré condicionamento é fundamental para acelerar a convergência dos algoritmos gradiente e conjugado gradiente. No caso do algoritmo quase Newtoniano também é melhor adotar como primeira aproximação da inversa da Hessiana, a mesma matriz de pré condicionamento adotada nos métodos anteriores. O cálculo do tamanho do passo é um ponto muito importante. O uso de fórmulas aproximadas, que assume uma linearidade da função objetiva na vizinhança dos parâmetros, não é um procedimento recomendável, pois leva a uma convergência lenta e prematura. O uso de um método mais rigoroso para este fim é preferível, mesmo que esta acabe sendo a tarefa computacional mais custosa da inversão. O método quase Newtoniano mostrou um comportamento mais robusto com relação aos outros, pois foi capaz de reconstruir os parâmetros mesmo em pontos onde o modelo inicial se afastava mais da versão de baixa frequência do modelo real.

Agradecimentos

Este trabalho teve o apoio financeiro da Financiadora de Estudos e Projetos (FINEP) e Petrobras/CENPES através do projeto CTPETRO #21.01.0494.00

Referências

- Fletcher, R., 1970, A new approach to variable metric algorithms: *Comput. J.*, Vol.13, p317-322.
- Fletcher, R., 1987, *Practical Methods of Optimization: Wiley Inter. Science publication.*
- Gauthier, O., Virieux, J. e Tarantola, A., 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: *Geophysics*, Vol. 51, p387-403, July.
- Gill, P., E., Murray, W., e Wright, M. H., 1986, *Practical Optimization: Academic Press.*
- Kolb, P., Collino, F. e Lailly, P., 1986, Prestack inversion of a 1-D medium: *Proceedings Inst. Electr. Electron. Eng. special issue on seismic inversion.*
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, Vol.49, p1259-1266.
- Tarantola, A., 1987, *Inverse problem theory-methods for data fitting and model parameter estimation: Elsevier Science Publ. Co.*

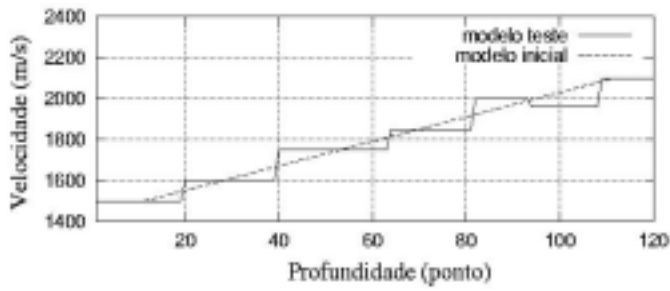


Figura 1- Modelo teste (linha sólida) e modelo inicial (linha tracejada) usado como ponto de partida para inversão.

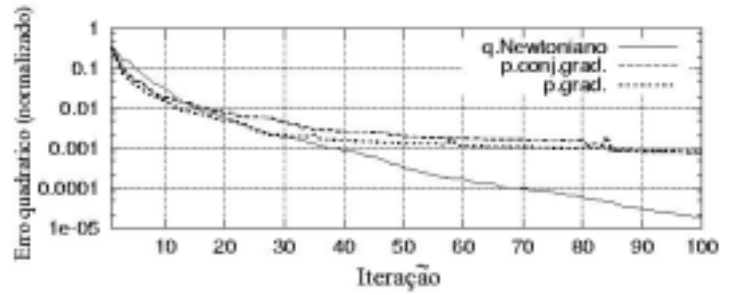


Figura 4- Convergência dos métodos pré-condicionados gradiente e conjugado gradiente e quase Newtoniano.

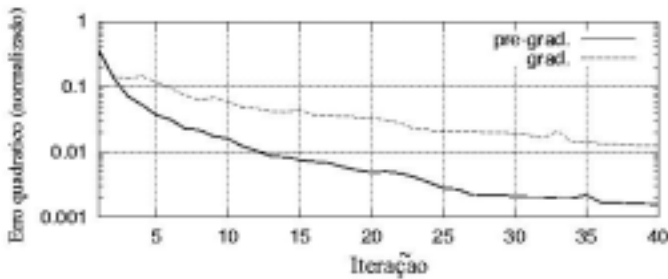


Figura 2- Convergência do método gradiente e pré condicionado Gradiente (linha sólida).

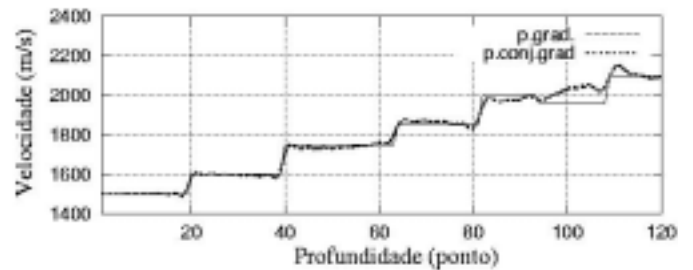


Figura 5- Resultados finais obtidos pelos métodos pré condicionados gradiente e conjugado gradiente.

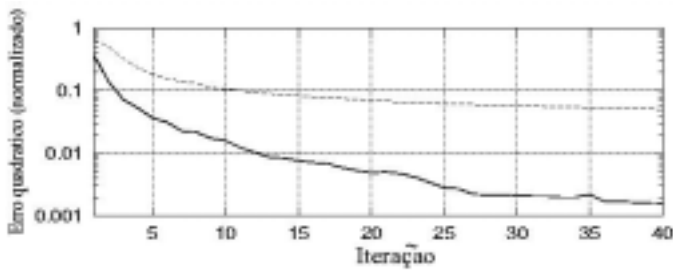


Figura 3- Convergência do método pré condicionado gradiente com cálculo rigoroso do passo (linha sólida) e cálculo aproximado do passo (linha tracejada).

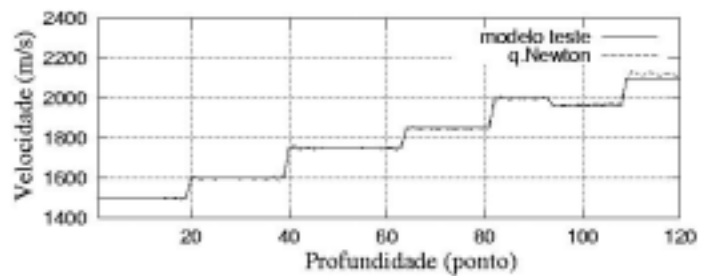


Figura 6- Resultado final obtido pelo método quase Newtoniano.