

OTIMIZAÇÃO DA SOLUÇÃO DE SISTEMAS LINEARES UTILIZANDO SOLVER PARDISO NA MODELAGEM DE DADOS CSEM

Emanuelle Machado* (emanuellesilva@on.br), Williams Lima (williamslima@on.br)* e Sergio Fontes (sergio@on.br)*
Observatório Nacional, RJ-Brasil*.

Copyright 2022, SBGf - Sociedade Brasileira de Geofísica

Este texto foi preparado para a apresentação no IX Simpósio Brasileiro de Geofísica, Curitiba, 04 a 06 de outubro de 2022. Seu conteúdo foi revisado pelo Comitê Técnico do IX 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.

Resumo

O objetivo deste estudo é avaliar a viabilidade de métodos diretos em comparação com os iterativos para problemas de grande porte em modelagem Magnetotelúrica (MT) e Eletromagnética de Fonte Controlada (CSEM). Esta investigação é especialmente importante para o caso CSEM onde o grande número de fontes pode tornar os métodos iterativos menos eficientes para esta tarefa. Construímos modelos sintéticos para as análises realizadas durante o estudo. Neste trabalho comparamos a eficácia e eficiência dos métodos BiCGStab, incluso no software ModEM, e o solver direto PARDISO, que foi incluído no código do ModEM. Reportamos aqui o desempenho de ambos os algoritmos em termos de tempo de execução, acurácia e número de iterações necessárias para atingir a acurácia desejada.

Introdução

A utilização de dados eletromagnéticos é importante para a exploração de hidrocarbonetos e tem se estabelecido firmemente como uma importante ferramenta geofísica no ambiente offshore (ver a Figura 1). O uso de dados eletromagnéticos oferece uma alternativa complementar ao método sísmico, principal método geofísico utilizado na exploração mineral e de hidrocarbonetos. Dessa forma, complementando as informações de sísmica com as informações de dados eletromagnéticos é possível reduzir os riscos nas tomadas de decisões quanto ao potencial de um determinado alvo exploratório, evitando-se a perfuração de poços "secos" ou com quantidades de petróleo que não compensem o investimento feito. Estudamos neste trabalho os métodos eletromagnéticos CSEM e Magnetotelúrico (MT) com uma aplicação particular na exploração de petróleo offshore. Para tornar a interpretação deste imenso volume de dados viável, precisamos de métodos computacionais robustos e eficientes que utilizem as tecnologias de simulações computacionais para a solução de sistemas lineares utilizados em problemas EM. Realizamos um comparativo entre as duas estratégias (métodos diretos versus métodos iterativos) tendo em vista parâmetros como, tamanho do modelo (número de células), potencial para a utilização de paralelismo em clusters, uso de memória, condicionamento do sistema linear e velocidade de convergência, possibilitando uma melhor interpretação de grandes volumes de dados EM tanto em trabalhos de modelagem quanto de inversão.

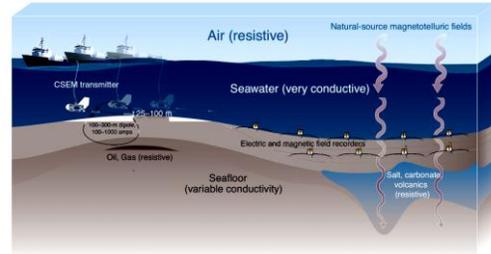


Figura 1 – Situação típica de uma aquisição de dados eletromagnéticos offshore.

Metodologia/ Problema Investigado

O processamento dos dados de uma aquisição CSEM é baseado na construção de um modelo de resistividade elétrica da subsuperfície. Formulamos o problema através das equações de Maxwell em meios condutivos no domínio da frequência (desprezando-se o termo de corrente de deslocamento) escrito para o campo elétrico:

$$\nabla \times \nabla \times \mathbf{E} - i\omega\mu_0\sigma\mathbf{E} = i\omega\mu_0\mathbf{J}_s \quad (01)$$

\mathbf{J}_s é o termo de fonte que no caso do método CSEM tratado neste trabalho corresponde a um dipolo elétrico horizontal e para o método MT é igual a zero. Após a solução da Eq.01, o campo magnético pode ser obtido em seguida através da expressão:

$$\mathbf{B} = -\frac{\nabla \times \mathbf{E}}{i\omega\mu_0} \quad (02)$$

No ModEM está implementada uma modificação da Eq.01 baseada no trabalho de EGBERT *et al.* (2018) que permite a obtenção de sistemas lineares mais bem condicionados, melhorando assim o desempenho e acurácia dos métodos de solução de sistemas lineares, particularmente dos métodos iterativos. A Eq.01 é modificada através da adição de um termo pênalti dada por $\nabla\lambda(\nabla \cdot \sigma\mathbf{E})$, tornando a Eq.03:

$$\nabla \times \nabla \times \mathbf{E} - \nabla\lambda(\nabla \cdot \sigma\mathbf{E}) + i\omega\sigma\mathbf{E} = i\omega\sigma\mathbf{J}_s \quad (03)$$

As Eq.01 e Eq.02 são discretizadas utilizando o método de diferenças finitas (KELBERT *et al.* 2014) resultando no seguinte sistema de equações algébricas lineares para as componentes do campo elétrico:

$$\mathbf{S}(\mathbf{m}, \omega)\mathbf{e} = \mathbf{b} \quad (04)$$

Em que

$$\mathbf{S}(\mathbf{m}, \omega) = [\mathbf{C}^T \mathbf{C} + \text{diag}(i\omega\mu_0\sigma(\mathbf{m}))] \mathbf{e} \quad (05)$$

em que $C: S_p \mapsto S_D$ é a versão discretizada do operador rotacional mapeando componentes dos vetores definidos nas arestas das células da malha para componentes dos vetores definidos nas faces das células da malha. A Eq.02 discretizada fica:

$$\mathbf{h} = (-i\omega\mu_0)^{-1} \mathbf{C} \mathbf{e} \quad (06)$$

O software ModEM tem sido muito utilizado na modelagem e inversão de dados CSEM/MT em todo o mundo. Estudamos a implementação da solução do sistema linear (Eq. 04) dentro do ModEM. A implementação do BiCGStab no ModEM é uma tradução para o Fortran da versão encontrada no software Matlab adaptada para o uso nas rotinas de modelagem e inversão. Um fator interessante dos métodos BiCGStab e PARDISO é a diferença entre as interfaces de processamento dos solucionadores. O PARDISO possui a interface de solver conhecido como uma biblioteca Intel oneAPI Math Kernel Solucionador PARDISO. Enquanto o BiCGStab, incluído no ModEM, utiliza a interface Open MPI Library, que possui um melhor desempenho em clusters de computação de alto desempenho baseados em processadores Intel. Para avaliarmos o tempo computacional requeridos nos dois métodos, executamos ModEM utilizando OpenMPI e o OpenMP com diversas opções de paralelização. Para a paralelização usando MPI variamos o número de processos (indicado neste trabalho pelo parâmetro n) e para a paralelização utilizando o OpenMP, utilizado pelo PARDISO, utilizamos o número de *Threads* (indicado pelo parâmetro e). A biblioteca PARDISO utilizada neste trabalho é fornecida pelo pacote do Intel MKL oneAPI. Neste trabalho propomos a implementação de um método direto como alternativa para a solução de Eq.04 utilizando a biblioteca do solver direto PARDISO (abreviação de “*Parallel Direct Solver*”) descrito em KARYPIS *et al.* (2017).

Criamos um modelo de condutividade mais complexo para simulações imitando estruturas de subsuperfície mais realistas. Este modelo (baseado em MEQBEL e RITTER, 2015) foi adotando para as análises dos solvers PARDISO e BiCGStab. O modelo foi construindo no software ModEM 3D-GRID (disponibilizado pelo ON para os estudos do projeto). A Figura 2 mostra uma seção 2D do modelo proposto, discretizado numa malha possuindo 68 células no eixo x, 80 células no eixo y e 50 células no eixo z. Este modelo consiste em cinco camadas com resistividades $10\Omega m$, $500\Omega m$, $50\Omega m$, $1000\Omega m$ e $10\Omega m$ respectivamente, as espessuras das camadas são de 400 m, 300 m, 800 m e 8200 m.

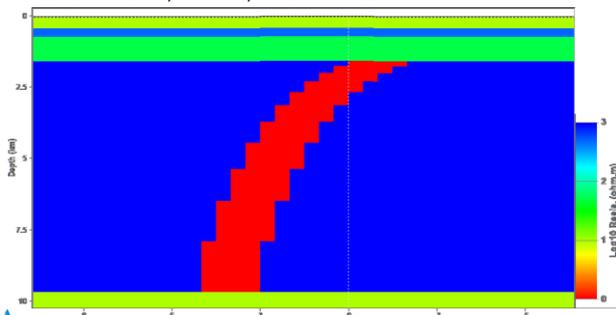


Figura 2 – O Modelo 2D com cinco camadas com diferentes resistividades e espessuras criado no 3D-GRID

Em direção ao centro no modelo, representado pela Figura 2, a segunda camada $500\Omega m$ é deformada para simular uma série de dobras com uma estrutura anticlinal localizada no centro e dois sinclinais estendendo-se para ambas as direções horizontais, indicado pela **região A** da Figura 3. A quarta camada, com valor de resistência igual a $1000\Omega m$, é cortando próxima a região do centro do modelo por um canal subvertical, representado pela **região B** indicado na Figura 3, de 2000m de largura integrado com material condutor de resistência igual a $1\Omega m$. O canal condutor se direciona para o eixo y e atinge uma profundidade de proximamente 9,7 km. A camada condutora mais profunda, com resistividade igual $10\Omega m$, permanece não deformada.

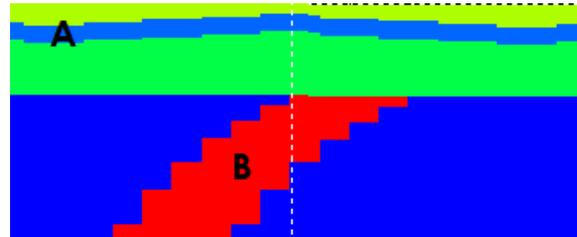


Figura 3 – O plano XY do segundo modelo em 3D. Representações das regiões sinclinal/anticlinal e da região do canal subvertical

Definimos cinco diferentes *layouts* de aquisição de dados contendo um total de 1, 17, 51, 85 e 119 fontes (transmissores), 1 período e um receptor localizado no centro da região de aquisição. A Figura 3 mostra o modelo no plano XY para caso de 119 fontes indicada pelos pontos verdes que representam os transmissores e ponto preto indicando receptor no modelo.

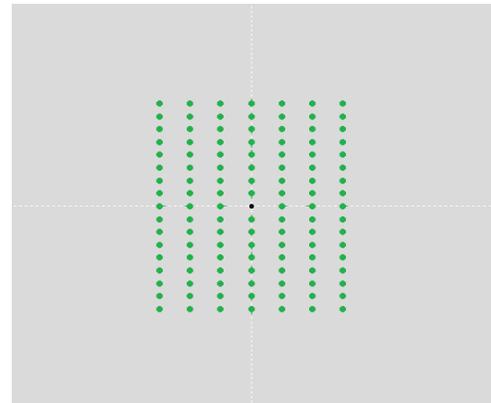


Figura 3 – Plano XY representado o arquivo de dado com 119 fontes

Resultados

Com o modelo indicando pelas Figuras 2-3, simulamos dados de CSEM com nove posições (transmissores) e 5 valores de períodos 0.1s, .177s, 0.316s, 0.562s e 1.0s. Definiu um layout de aquisição com 65 estações (possuindo 65 receptores para cada transmissor) e dados com 45 fontes. Na Figura 4 mostramos os valores do

campo elétrico em função dos dados dos transmissores/receptores nas estações. A simulação se baseia na modelagem de dados CSEM para comparação dos parâmetros envolvidos na execução do código de modelagem entre o BiCGStab, representado pela curva traçada em vermelho, e o PARDISO indicada pela a curva traçada em pontos azuis (termo GD na legenda simboliza Grad-Div).

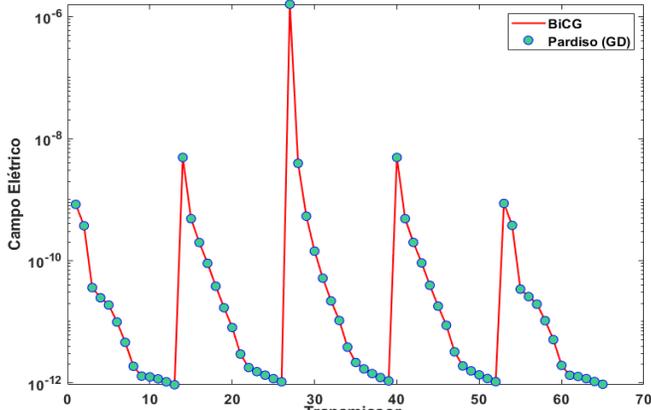


Figura 4 – Valores do campo elétrico do BiCGStab (vermelho) e PARDISO (azul).

Em uma segunda etapa de testes utilizamos o mesmo modelo para simulação de dados CSEM com valor de período igual a 1 segundo. Definiu-se um layout de aquisição com 5 arquivos de dados 1, 17, 51, 85 e 119 fontes (transmissores). A Figura 5 mostra os tempos totais de execução do programa ModEM para os cinco casos, a curva traçada em azul representa os valores de tempo do PARDISO e a curva traçada em vermelho para BiCGStab.

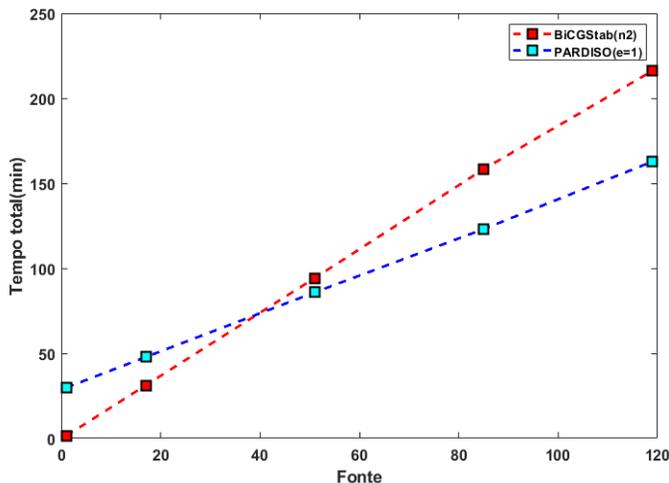


Figura 5 – Período de convergência para BiCGStab (vermelho) e PARDISO (azul)

Tabela 1 – Resultado do período de convergência do BiCGStab e PARDISO

Nº Fontes	Tempo BiCGStab (min)	Tempo do PARDISO (min)
1	1.5	29.83
17	31.22	47.82
51	94.27	85.83
85	158.28	123.04
119	216.13	162.81

A Figura 6 mostra o tempo total para o caso com 119 fontes. Nesta figura temos três curvas representando o tempo total: o BiCGStab, representado pela curva traçada em roxo, o PARDISO com curva com traçada com pontos verdes e o PARDISO (sistema) que representa somente o tempo de solução do sistema de dados representado pela curva traçada de pontos azuis.

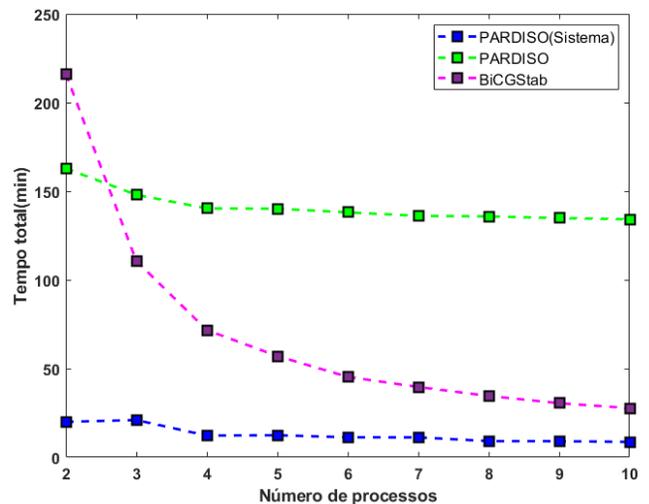


Figura 6 – Representação do gráfico do processamento total do BiCGStab (roxo) e PARDISO (verde) versus o processamento PARDISO(azul) somente solucionado o sistema sem o pré-processamento.

Discussão e Conclusões

Observando-se o gráfico da Figura 4 verificamos que ambos os métodos concordam em acurácia com os valores para o campo elétrico calculado.

De acordo com a Tabela 1 o PARDISO começa a ter uma vantagem em tempo total de processamento sobre o BiCGStab a partir de um total de 51 fontes, lembrando que nesta situação não utilizamos paralelismo.

Na Figura 06 mostramos o desempenho computacional utilizando paralelismo para o caso correspondente a 119 fontes. Para o BiCGStab variamos o número de processos MPI através do parâmetro n (1 a 10). Para o PARDISO variamos o número de *Threads* OpenMP (1 a 10). Considerando apenas o tempo de solução do sistema linear verificamos que o PARDISO se torna mais eficiente do que o BiCGStab, no entanto devido a sobrecarga de tarefas extras no software ModEM relacionadas a inicialização, o BiCGStab se torna mais eficiente quando é paralelizado.

Em virtude do problema apontado acima relativo a carga extra do software não diretamente relacionada a solução do sistema linear, estamos estudando alternativas híbridas de paralelização utilizando tanto OpenMPI e OpenMP simultaneamente. Com esta estratégia híbrida esperamos reduzir o tempo computacional relativamente ao uso de apenas OpenMP ou OpenMPI separadamente ao utilizarmos o PARDISO como *solver*.

Agradecimentos

Os autores agradecem ao Observatório Nacional pelo suporte da pesquisa e a FAPEMA pelo financiamento da bolsa de pesquisa da mestranda.

Referências

DONG, H., EGBERT, G. D., 2018, "Divergence-free solutions to electromagnetic forward and adjoint problems: a regularization approach", *Geophys. J.Int.* n. 216 (November), pp. 906–918.

KELBERT, 2014, "ModEM: A modular system for inversion of electromagnetic geophysical data", *Computers & Geosciences*, v. 66.

KELBERT, A., EGBERT, G., 2012, "Computational recipes for electromagnetic inverse problems", *Geophys.J.Int.*, v. 128, n. 251-261.

KARYPIS, G., "Intel Direct Sparse Solver for Clusters, a Research Project for Solving Large Sparse Systems of Linear Algebraic Equations on Clusters", CERFACS.

MEQBEL, N., RITTER, O., 2015, "Joint 3D inversion of multiple electromagnetic datasets", *Geophys. Prospecting*, v. 63, pp. 1450–1467.

SCHENK, O., GARTNER, K., 2002, "Two-level scheduling in PARDISO: Improved Scalability on Shared Memory Multiprocessing Systems." *Parallel Computing*, v. 28, pp. 187–197.

SCHENK, O., GARTNER, K., 2004, "Solving Unsymmetric Sparse Systems of Linear Equations with PARDISO", *Journal of Future Generation Computer Syst.*, v. 20, pp. 475–487.