| |
|---|
| AGG5740 – TEORIA DE INVERSÃO EM GEOFÍSICA |
| PROVA 1 – 03/12/2019 Entrega até 06/12/2019, às 18 horas |

**A SOLUÇÃO DA PROVA É INDIVIDUAL!**

1) Entregar um relatório sucinto, mas substanciado, contendo figuras e código fonte utilizado para resolver o problema proposto.

2) A entrega da prova pode ser feita via e-mail (calbertomc@usp.br) ou pessoalmente;

A avaliação será feita considerando a metodologia, os resultados obtidos e a clareza da formulação, ajudada por comentários e explicações adequadamente inseridos nos scripts.

**EXERCÍCIO 1** (10.0) – Considere um problema de tomografia em um determinado objeto. Deseja-se obter informações sobre a estrutura interna desse objeto, a partir de um arranjo bem simples, como mostrado na Figura 1. O tempo de percurso de uma onda elástica, $t$, é o tempo gasto para a mesma se propagar da fonte $F$ até um receptor $R$. A partir das medidas de tempo de percurso nos receptores, pretende-se estimar a variação de velocidade dentro do material, tal que se possa fazer inferências sobre as propriedades físicas do objeto.

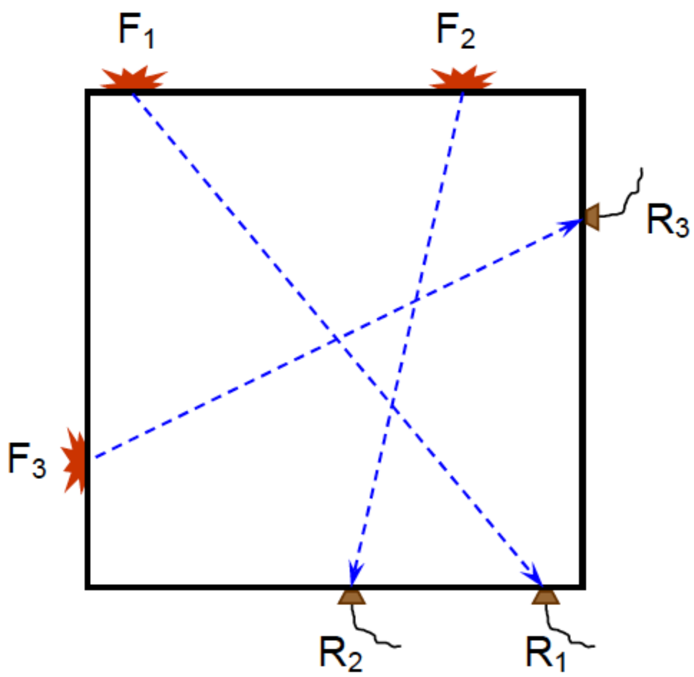


**Figura 1.** Arranjo de um experimento de tomografia sísmica. $F$ e $R$ representam fonte e receptor, respectivamente.

**Simplificações**

1) Presume-se que as ondas sejam de alta-frequência tal que a teoria do raio seja válida;
2) Presume-se que as nove células mostradas na Figura 2 sejam suficientes para descrever a variação da vagarosidade (1/velocidade) no interior do objeto;
3) Presume-se que a vagarosidade é constante dentro de uma célula quadrada e que de uma célula para outra a vagarosidade possa variar;
4) Desprezam-se os efeitos de refração do raio ao entrar ou sair de uma determinada célula.

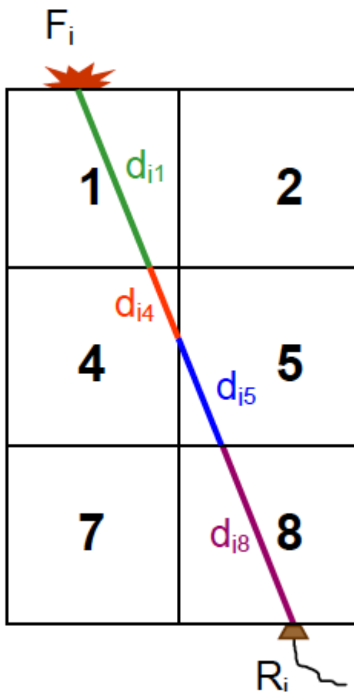


**Figura 2.** Representação da parametrização do objeto de estudo. $F$ e $R$ representam fonte e receptor, respectivamente. O parâmetro $d$ representa a distância entre a fonte e o receptor.

## Modelo físico

O tempo total de percurso $t_i$ de uma onda gerada em $F_i$ e registrada em $R_i$ será soma dos tempos gastos pela onda para atravessar cada célula envolvida em sua trajetória. No caso ilustrado na Figura 2, teremos:

$$t_i = d_{i1}p_1 + d_{i4}p_4 + d_{i5}p_5 + d_{i8}p_8 \tag{1}$$

sendo $d_{ij}$ a distância que a onda, associada à i-ésima observação, percorre dentro da j-ésima célula e $p_j$ a vagarosidade da j-ésima célula. Veja que escolhemos utilizar $p_j = \frac{1}{v_j}$, sendo $v_j$ a velocidade do meio em cada célula de forma a linearizar a equação.

## Equações de observação

Se considerarmos $N$ observações, ou seja, $N$ posições diferentes para o par Fonte-Receptor, teremos um sistema de $N$ equações lineares:

$$\begin{bmatrix} \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes \\ d_{i1} & 0 & 0 & d_{i4} & d_{i5} & 0 & 0 & d_{i8} & 0 \\ \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes & \otimes \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_{i-1} \\ t_i \\ t_{i+1} \\ \vdots \\ t_{N-1} \\ t_N \end{bmatrix}, \tag{2}$$

ou em notação matricial:

$$\boldsymbol{Gp} = \boldsymbol{t} \tag{3}$$

sendo $\boldsymbol{G}$ a chamada matriz de sensibilidade, $\boldsymbol{p}$ o vetor de parâmetros e $\boldsymbol{t}$ o vetor de dados.

Dadas as informações acima, pede-se:

**<u>Parte 1</u>**

a) Obtenha a matriz $\boldsymbol{G}$ de acordo com a metodologia exemplificada na equação (1), para o caso de nove observações, registradas em cada receptor, de cada uma das fontes associadas ao layout de fontes-receptores representados na Figura 3.

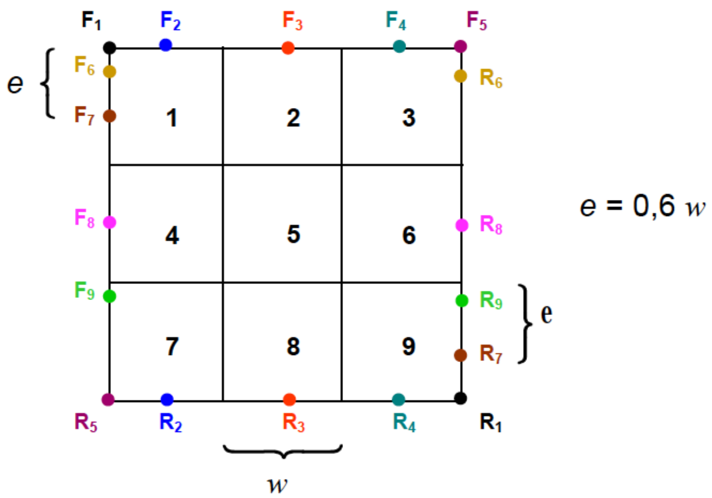


**Figura 3.** Conjunto de Fontes e Receptores utilizado no experimento tomográfico. O parâmetro $w$ representa um valor arbitrário para a largura da célula.

**Parte 2**

Obtenha observações sintéticas do seguinte modo:

a) Assinale valores numéricos aos elementos do vetor de parâmetros $\boldsymbol{p}$. Cada conjunto de nove valores de parâmetros representa um tipo de distribuição de vagarosidade no interior do objeto e constituirá o corpo "real";

b) Para cada vetor $\boldsymbol{p}$, com elementos especificados de acordo com o item (a), obtenha o vetor de observações sintéticas exatas, $\boldsymbol{t_{sin}}$, efetuando o produto $\boldsymbol{Gp}$;

c) Some a cada elemento do vetor $\boldsymbol{t_{sin}}$, obtido no item (b), a realização de uma variável aleatória com distribuição uniforme definida no intervalo $[-b, b]$, sendo $b = |0.02\ t_{MAX}|$, ou seja 2 % do maior elemento de $\boldsymbol{t_{sin}}$ em módulo. Desse modo, obtém-se o vetor de observações sintéticas com ruído, $\boldsymbol{t_{noise}}$. A realização da variável aleatória acima referida deve ser obtida através de um gerador de números pseudoaleatórios. Os números pseudoaleatórios formam sequências que podem ser reproduzidas sempre que se desejar, ao passo que uma sequência de números aleatórios nunca se repete. Uma sequência específica de números pseudoaleatórios depende apenas de uma "semente", que é um número (em geral inteiro) especificado pelo interprete. Assim, diferentes sementes geram sequências diferentes de números pseudoaleatórios, que podem ser reproduzidas exatamente, bastando especificar a mesma semente.

d) Utilizando o estimador de quadrados mínimos regularizado, dado por:

$$\widehat{\boldsymbol{p}} = (\boldsymbol{G^T G} + \mu \boldsymbol{I})^{-1} \boldsymbol{G^T t_{noise}}, \tag{5}$$

sendo $\mu$ o parâmetro de regularização, estime a solução do problema, para 5 diferentes parâmetros de regularização, variando $\mu$ em uma ordem de grandeza em relação ao valor anterior;

e) Estime a matriz de covariância de $\widehat{\boldsymbol{p}}$ e forneça o desvio padrão dos parâmetros;

e) Gere 10 vetores $\boldsymbol{t_{noise}}$, $i = 1, 2, \ldots L$, somando aos elementos do vetor exato $\boldsymbol{t_{sin}}$, $L$ sequências diferentes de números pseudoaleatórios, simulando, por exemplo, a coleta das mesmas observações em tempos distintos. A cada coleta, a sequência de ruído é diferente, mantendo-se todavia as características estatísticas do ruído. Esses vetores $\boldsymbol{t_{noise}}$ serão usados para avaliar se uma determinada solução é ou não estável, como será descrito na parte 3.

**Parte 3**

Procedimento para determinar se uma solução é estável:

a) Atribua um valor não nulo ao parâmetro de regularização $\mu$;

b) Contamine as observações com uma determinada sequência de números pseudoaleatórios de acordo com as instruções da Parte 2c;

c) Aplique o estimador dado na equação (5) e visualize a solução;

d) Modifique a semente do gerador de números pseudoaleatórios;

e) Aplique novamente o estimador dado pela equação (5) e visualize a nova solução;

f) Compare as soluções obtidas em (c) e (e). Se elas estiverem "próximas" entre si, o problema formulado é estável. Caso contrário ele é instável.

g) Estime a matriz de covariância de $\hat{\boldsymbol{p}}$ e forneça o desvio padrão dos parâmetros para cada uma das situações;

O grau de proximidade entre as soluções obtidas do modo acima descrito é estabelecido pelo interprete e depende da incerteza que ele admite para a solução em decorrência de ruído nos dados. Ao gerar um certo número de soluções (10 ou mais), cada uma com uma sequência de ruído diferente, o interprete pode inspecionar graficamente essas soluções e avaliar se elas estão dentro da margem de erro aceitável. Alternativamente, ele pode armazenar todas as estimativas obtidas para todos os parâmetros e fazer uma análise estatística, computando o desvio padrão amostral de cada parâmetro e verificando se eles estão dentro do limite de incerteza aceitável estabelecido pelo interprete. Por exemplo, o interprete estabelece a priori que o máximo erro permitido para a determinação das espessuras é 1 m. Suponha que em três inversões usando diferentes sequências de números pseudoaleatórios que contaminam as observações, os valores das estimativas do i-ésimo parâmetro são 50.4 m, 49.7 m e 51.3 m. O desvio padrão amostral dessas três estimativas é dado por:

$$s = \sqrt{\frac{\sum_{i=1}^{L}(\hat{h}_i - \bar{h})^2}{L-1}}, \tag{6}$$

sendo $L$ o número de estimativas (3 nesse exemplo), $\hat{h}_i$ a i-ésima estimativa e $\bar{h}$ a média amostral das estimativas. Para o exemplo acima, o valor de $s$ é 0.8. Assim, a solução é considerada estável em relação à determinação do i-ésimo parâmetro. Esse mesmo procedimento pode ser aplicado a todos os outros parâmetros ou a grupos selecionados de parâmetros, dependendo sempre do interesse do interprete. A decisão sobre a estabilidade de uma solução é, portanto, subjetiva e depende de como o interprete conceitua estabilidade. Estabilidade não é uma grandeza objetiva como unicidade em que só há duas possibilidades: ou a solução é única ou não. Estabilidade é uma grandeza semi-quantitativa. Só podemos afirmar objetivamente que uma solução é mais (ou menos) estável que outra. Qualquer tentativa de caracterizar estabilidade em termos absolutos (afirmar que uma solução é ou não estável) depende da conceituação prévia do interprete sobre o limiar de erro nas estimativas dos parâmetros, a partir do qual o interprete considera uma solução instável.