AGG-5740

Teoria da Inversão em Geofísica

Aula 16- Modelagem probabilística do resíduo

Introdução

Um dado geofísico resulta da sobreposição de efeitos determinísticos e aleatórios. Na inversão de dados geofísicos, considera-se a parte determinística como aquela associada às propriedades físicas do substrato. Esta parte pode ser prevista por funcionais da forma $g(x_i,\overline{p})$. Nesta formulação, os parâmetros $\overline{p}$ representam a distribuição de propriedade física envolvida, x_i um parâmetro de aquisição de dados por exemplo: posição de medida, distância entre eletrodos (sondagens EL), período (sondagens MT), O valor numérico assumido pelo funcional expressa o valor teórico associado ao modelo físico-matemático em questão. Ao proceder o ajuste de dados teóricos com dados reais estamos "modelando" os dados, ou seja associando a eles um modelo físico. Nestes casos, o ajuste ocorre de forma aproximada o que gera um resíduo entre os valores previstos e os medidos. Este resíduo pode ser atribuído tanto como decorrente de erros mensurais como da incapacidade do modelo em ajustar perfeitamente os dados. Isto é explicitamente notado quando se emprega modelos interpretativos muito simples em terrenos comparativamente mais complexos. Nesta ótica torna-se razoável atribuir um modelo também ao resíduo, seja um modelo decorrente de uma formulação determinística ou probabilística. Nos dois casos, o resíduo não pode ser confundido com as incertezas intrínsecas ao processo de medida.

A modelagem dos resíduos numa concepção probabilística leva-nos aos estimadores de veros-similhança máxima (EVM) que, como veremos nesta aula, correspondem a estimadores de mínimos quadrados ponderados, cujos pesos são definidos de acordo com a representação prevista para o resíduo.

Função de verossimilhança

Admitindo que o ruído (resíduo) η_i perturba adtivamente a parte determinística do dado, o valor medido, d_i, de um problema linear pode ser escrito

$$d_i = \overline{g}_i^t \overline{p} + \eta_i \tag{1}$$

ou para um conjunto de n dados

$$\overline{d} = \overline{\overline{G}}\,\overline{p} + \overline{\eta} \tag{2}$$

A parte aleatória (residual) é então

$$\overline{\eta} = \overline{d} - \overline{\overline{G}}\,\overline{p} \tag{3}$$

que a partir de agora será caracterizada por meio de funções de densidade de probabilidade (fdp) e seus parâmetros. Como indica a equação (1), cada dado é contaminado pela realização de uma VA específica, η_i. Considerando que tais variáveis sejam independentes temos que a probabilidade conjunta, ou de realização do conjunto de dados geofísicos, será o produto das probabilidades de cada evento particular, ou seja:

$$L(\overline{\eta}) = \prod_{i=1}^{n} f(\eta_i) \quad . \tag{4}$$

O termo $L(\overline{\eta})$ é denominado função de verossimilhança e assumirá um valor máximo quando todos os termos do produtório forem máximos (nenhum próximo a zero), ou justamente quando a realização de cada uma das variáveis aleatórias estiver numa faixa de valores com maior probabilidade. A equação (3) nos mostra que podemos obter diferentes resíduos $\overline{\eta}$ simplesmente variando os parâmetros $\overline{p}$. Podemos conceber $L(\overline{\eta}) \equiv L'(\overline{p})$ e considerar que ao maximizar $L(\overline{\eta})$ estaremos encontrando um conjunto de parâmetros que além de satisfazer um modelo determinístico para os parâmetros satisfaz um modelo probabilísitco para os resíduos. A este estimador dá-se o nome de estimador de máxima verossimilhança.

Observe que maximizar $L(\overline{p})$ eqüivale a minimizar um funcional *Q* tal que:

$$Q = -\log[L(\overline{\eta})] = \sum_{i=1}^{n} -\log[f(\eta_i)] \tag{5}$$

pois a função logarítmica é monotônica crescente. O mínimo do Q que é obtido determinando o ponto com gradiente nulo (poderia ser ponto de máximo ou inflexão?):

$$\nabla_{\overline{p}} \left\{ \sum_{i=1}^{n} -\log[f(d_i - \overline{g}_i^t \overline{p})] \right\} = 0 \tag{6}$$

Em tese, a equação (6) pode ser aplicada para qualquer fdp embora, na prática, a complexidade matemática a pode dificultar sua implementação.

Ruído/resíduo Gaussiano

Considerando que o ruído pode ser modelado pela fdp de Gauss, ou normal, tem-se que

$$f(\eta_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{\eta_i - E\{\eta_i\}}{\sigma_i} \right)^2 \right] \tag{7}$$

ou, assumindo que $E\{\eta_i\} = 0$, que

$$f(\eta_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{\eta_i}{\sigma_i} \right)^2 \right] \tag{8}$$

sendo σ_i o desvio padrão do i-ésimo resíduo. O funcional a ser minimizado, correspondente à equação (5), é então

$$Q = \sum_{i=1}^{n} -\log\left(\frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[-\frac{1}{2} (\frac{\eta_i}{\sigma_i}) \right] \right) = \sum_{i=1}^{n} -\log\left(\frac{1}{\sigma_i \sqrt{2\pi}} \right) + \log\left(\exp\left[-\frac{1}{2} (\frac{\eta_i}{\sigma_i}) \right] \right) \tag{9}$$

ou

$$Q = \sum_{i=1}^{n} -\log\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right) + \sum_{i=1}^{n} \frac{1}{2}\left(\frac{d_i - \bar{g}_i^t \bar{p}}{\sigma_i}\right)^2 \tag{10}$$

O gradiente da equação (10) é

$$\nabla_{\bar{p}}(Q) = 0 + \sum_{i=1}^{n} \frac{1}{2}\nabla_{\bar{p}}\left[\left(\frac{d_i - \bar{g}_i^t \bar{p}}{\sigma_i}\right)^2\right] = \frac{1}{2}\nabla_{\bar{p}}[(\bar{d} - \bar{\bar{G}}\bar{p})^t\, \bar{\bar{C}}_d^{-1}(\bar{d} - \bar{\bar{G}}\bar{p})] \tag{11}$$

ou

$$\nabla_{\bar{p}}(Q) = \bar{\bar{G}}^t \bar{\bar{C}}_d^{-1}(\bar{d} - \bar{\bar{G}}\bar{p}) \tag{12}$$

sendo

$$\bar{\bar{C}}_d = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix} \tag{13}$$

O ponto de mínimo é obtido fazendo $\nabla_{\bar{p}}(Q) = 0$ que fornece

$$\hat{\bar{p}} = (\bar{\bar{G}}^t \bar{\bar{C}}_d^{-1} \bar{\bar{G}})^{-1} \bar{\bar{G}}^t \bar{\bar{C}}_d^{-1} \bar{d} \tag{14}$$

ou quando todos resíduos tiverem o mesmo desvio padrão que

$$\hat{\bar{p}} = (\bar{\bar{G}}^t \bar{\bar{G}})^{-1} \bar{\bar{G}}^t \bar{d} \quad . \tag{15}$$

As equações (14) e (15) são as expressões do EMQ ponderado e simples, respectivamente, vistos na Aula 2, os quais podem ser considerados como sendo um EVM caso a matriz de pesos for composta pelas variâncias de cada resíduo. Quando as premissas forem verdadeiras, o EMQ modela a parte aleatória dos dados, caso contrário ele pode distorcer a estimativa dos parâmetros ao tentar modelar o resíduo através da parte determinística do modelo. Neste caso pode-se cogitar empregar outros modelos para o resíduo, por exemplo utilizando outras funções de densidade de probabilidade.

<u>Fdp de Poisson</u>

A fdp de Poisson é

$$f(\eta_i) = k\, e^{-|\eta_i|} \tag{16}$$

O funcional *Q* a ser minimizado é então

$$Q = \sum_{i=1}^{n} -\log(k\, e^{-|\eta_i|}) = \sum_{i=1}^{n} -\log(k) + \sum_{i=1}^{n} |\eta_i| \tag{17}$$

ou

$$Q = \sum_{i=1}^{n} -\log(k) + \sum_{i=1}^{n} |d_i - \bar{g}_i^t \bar{p}| \tag{18}$$

indicando que o EVM com esta fdp eqüivale a minimizar o valor absoluto dos resíduos. Na secção 3 será apresentado um algoritmo para obter o ponto de mínimo uma vez que a função módulo não é diferenciável.

FDP de Cauchy

A fdp de Cauchy é

$$f(\eta_i) = \frac{1}{\pi}\frac{\omega}{\eta_i^2 + \omega^2} \qquad (19)$$

sendo ω a meia largura da distribuição. O funcional a ser minimizado é então

$$Q = \sum_{i=1}^{n} -\log\left(\frac{1}{\pi}\frac{\omega}{\eta_i^2 + \omega^2}\right) = \sum_{i=1}^{n} -\log(\frac{\omega}{\pi}) + \sum_{i=1}^{n} \log(\eta_i^2 + \omega^2) \qquad (20)$$

ou

$$Q = \sum_{i=1}^{n} -\log(\frac{\omega}{\pi}) + \sum_{i=1}^{n} \log\left[(d_i - \overline{g}_i^t\overline{p})^2 + \omega^2\right]$$

cujo gradiente é

$$\nabla_{\overline{p}}(Q) = 0 + \sum_{i=1}^{n} \nabla_{\overline{p}}\{\log[(d_i - \overline{g}_i^t\overline{p})^2 + \omega^2]\} = \sum_{i=1}^{n} \frac{\nabla_{\overline{p}}(d_i - \overline{g}_i^t\overline{p})^2}{(d_i - \overline{g}_i^t\overline{p})^2 + \omega^2} \qquad 22)$$

ou

$$\nabla_{\overline{p}}(Q) = \sum_{i=1}^{n} -\frac{2(d_i - \overline{g}_i^t\overline{p})\overline{g}_i}{(d_i - \overline{g}_i^t\overline{p})^2 + \omega^2} \qquad (23)$$

O estimador é obtido fazendo o gradiente ser nulo

$$\sum_{i=1}^{n} \frac{d_i\overline{g}_i}{(d_i - \overline{g}_i^t\overline{p})^2 + \omega^2} = \sum_{i=1}^{n} \frac{\overline{g}_i^t\hat{\overline{p}}\overline{g}_i}{(d_i - \overline{g}_i^t\overline{p})^2 + \omega^2} \qquad (24)$$

ou

$$\sum_{i=1}^{n} \overline{g}_i w_i\, d_i = \sum_{i=1}^{n} \overline{g}_i w_i \overline{g}_i^t\hat{\overline{p}} \qquad (25)$$

ou

$$\overline{\overline{G}}^t\overline{\overline{W}}d = \overline{\overline{G}}^t\overline{\overline{W}}\,\overline{\overline{G}}\hat{\overline{p}} \qquad (26)$$

ou finalmente que

$$\hat{\overline{p}} = (\overline{\overline{G}}^t\overline{\overline{W}}\,\overline{\overline{G}})^{-1}\overline{\overline{G}}^t\overline{\overline{W}}d \qquad (27)$$

sendo $\overline{\overline{W}}$ uma matriz cujo i-ésimo elemento diagonal é

$$w_{ii} = \frac{1}{(d_i - \overline{g}_i^t\hat{\overline{p}})^2 + \omega^2} \qquad (28)$$

Observe que o elemento diagnonal de $\overline{\overline{W}}$ é função dos próprios parâmetros a serem estimados impedindo que se obtenha uma forma explícita para o EVM neste caso. A solução é então obtida de forma iterativa empregando algoritmo dado a seguir.

Algoritmo para Minimização de Funcionais dos Resíduos

As três exemplos anteriores, cada um com uma diferente função de densidade de probabilidade, levaram à minimização de um funcional dos resíduos genericamente designado por

$$Q = \sum_{i=1}^{n} U(r_i) \tag{29}$$

sendo

$$r_i = d_i - \overline{g}_i^t \overline{p} \, . \tag{30}$$

Dependendo da fdp adotada *U(r)* na equação (29) será de um tipo específico. Para a distribuição de Gauss $U(r_i) = r_i^2$ e para a de Poisson $U(r_i) = |r_i|$, por exemplo. Para se obter o ponto de mínimo de *Q* é necessário calcular as derivadas deste funcional em relação a cada um dos *m* parâmetros:

$$\frac{\partial}{\partial p_j}(Q) = \sum_{i=1}^{n} \frac{\partial}{\partial p_j}[U(d_i - \overline{g}_i^t \overline{p})] = \sum_{i=1}^{n} -U'(d_i - \overline{g}_i^t \overline{p}) \frac{\partial(\overline{g}_i^t \overline{p})}{\partial p_j} \tag{31}$$

Definindo

$$q_i = U'(d_i - \overline{g}_i^t \overline{p}) \, . \tag{32}$$

e considerando que

$$\frac{\partial}{\partial p_j}(\overline{g}_i^t \overline{p}) = \frac{\partial(\overline{p}^t \overline{g}_i)}{\partial p_j} = g_{i,j} \tag{33}$$

tem-se

$$\overline{\overline{G}}^t \overline{q} = \overline{0}_m \tag{34}$$

permitindo escrever

$$\overline{q} = \overline{\overline{W}} r \tag{35}$$

sendo $\overline{\overline{W}}$ uma matriz diagonal. Para resíduos com distribuição gaussiana, por exemplo,

$$q_i = 2r_i \tag{36}$$

indicando que $w_i = 2$, ou seja: igual para todos os dados. Os pesos correspondentes à distribuição de Poisson são

$$w_i = \frac{1}{|r_i|} \, . \tag{37}$$

A equação (35) fornece

$$\overline{\overline{G}}^t \overline{\overline{W}} \overline{r} = \overline{0}_m \tag{38}$$

ou

$$\overline{\overline{G}}^t \overline{\overline{W}}(\overline{d} - \overline{\overline{G}}\,\overline{p}) = \overline{0}_m \quad (39)$$

que fornece

$$\hat{\overline{p}} = (\overline{\overline{G}}^t \overline{\overline{W}}\,\overline{\overline{G}})^{-1} \overline{\overline{G}}^t \overline{\overline{W}}\,\overline{d} \quad (40)$$

A implementação do algoritmo apresentado pode ser feita segundo

passo 1- Obter a estimativa de mínimos quadrados $\hat{\overline{p}} = (\overline{\overline{G}}^t \overline{\overline{G}})^{-1} \overline{\overline{G}}^t \overline{d}$

passo 2- k=1

passo 3- $\overline{p}_k = \hat{\overline{p}}$

passo 4- k=k+1

passo 5- calcular os resíduos $r_i = d_i - \overline{g}_i^t \overline{p}$

passo 6- fazer a matriz $\overline{\overline{W}}$ ($w_{i,i} = 1/|r_i|$ caso a fdp seja de Poisson)

passo 7- calcular $\overline{p}_{k+1} = \overline{p}_k + (\overline{\overline{G}}^t \overline{\overline{W}}_k \overline{\overline{G}})^{-1} \overline{\overline{G}}^t \overline{\overline{W}}_k (\overline{d} - \overline{\overline{G}}\,\overline{p}_k)$

passo 8- se $|\overline{p}_{k+1} - \overline{p}_k| \le \varepsilon$, fim

passo 9- retorne ao passo 4

Outros Funcionais

Vide Anderson(1982) e Beltrão et al (1991).

E- Referências/Leitura Adicional

Anderson, K. R., 1982, Robust earthquake location using M-estimates: Physics of the Earth and Plan. Inter., 30, 119-130.

Beltrão, J. F., Silva, J. B. C. & Costa, J. C., 1991, Robust polynomial fitting for regional gravity estimation: Geophysics, 56, 80-89.

Claerbout, J. & Muir, F., 1973, Robust modeling with erratic data: Geophysics, 38, 826-844.

Constable, C. G., 1988, Parameter estimation in non-Gaussian noise: Geophys. J., 94, 131-142.

Gersztenkorn, A. Bednar, J. B. & Lines, L. R, 1986, Robust iterative inversion for the one-dimensional acoustic wave equation: Geophysics, 51, 357-368.

Silva, J. B. C. & Cutrim, A. O., 1989, A robust maximum likelihood method for gravity and magnetic interpretation: Geoexploration, 26, 1-31.