

INVERSÃO LINEAR

O problema linear no contínuo:

Denominamos as observações geofísicas de y^o e os parâmetros de p . Presumiremos que o sistema físico (modelo) que relaciona os parâmetros do modelo aos dados geofísicos observados y^o , é conhecido. Então podemos especificar uma função f que relaciona p aos dados y^o .

$$f(p) = y^o$$

Já vimos que os dados y^o e os parâmetros p podem ser funções contínuas do tempo e espaço, ou podem ser conjuntos discretos de medidas observacionais e parâmetros, respectivamente. Quando y^o e p são funções contínuas, então f é chamado de operador. Quando y^o e p são vetores, f é chamado de funcional. Independentemente de f ser um operador ou um funcional, chamo a atenção que é através de f que estabelecemos uma relação físico-matemática entre os parâmetros e os dados observados de um sistema físico. Ressalto que $f(p)$ pode envolver a solução de uma equação diferencial ordinária ou uma equação diferencial parcial ou a avaliação de uma integral. De qualquer forma f envolve relações de causa e efeito obtidas da resolução de um sistema de equações diferenciais advindas da física-matemática (equação de poisson, equação de Laplace, equações de Maxwell, equações da onda). Iremos adotar a convenção de denominar a relação f de um “modelo direto” ou “sistema físico” que nada mais é que uma ou mais equações que relacionam os dados geofísicos observados com os parâmetros desconhecidos. Nosso foco consiste determinar p a partir de y^o . Ressalto que o problema de identificação de modelo que consiste de determinar f a partir de p e y^o não abordaremos neste curso.

No caso de um sistema linear contínuo, verifica-se que f na equação acima pode ser sempre expresso como um operador integral linear. Então a equação $f(p) = y^o$ torna-se

$$y^o(x) = \int_a^b a(x, x') p(x') dx'$$

Esta equação acima é chamada de equação integral de Fredholm de primeira espécie, onde $p(x')$ é a função desconhecida do problema e representa alguma aspecto da estrutura da Terra ou alguma propriedade física (como por exemplo, resistividade, densidade, velocidade, susceptibilidade magnética), $y^o(x)$ é a função conhecida que representa os dados geofísicos observados e a função $a(x, x')$ é chamada de Kernel que também é uma função conhecida que relaciona os dados observados aos parâmetros desconhecidos $p(x')$. O intervalo de integração $[a, b]$ pode ser finito ou infinito. Teoricamente as observações representadas pela função $y^o(x)$ poderia teoricamente ser conhecida de em um intervalo infinito, mas na prática apenas temos medidas geofísicas de $y^o(x)$ em um intervalo finito de pontos de medidas.

$$y^o_i(x_i) = \int_a^b a_i(x_i, x') p(x') dx'$$

Vamos ver abaixo dois exemplos de problemas inversos lineares contínuos na geofísica.

Exemplos de problemas inversos lineares contínuos na geofísica:

1) Exemplo linear no contínuo em gravimetria:

Considere o problema inverso contínuo de estimar a distribuição de densidades numa linha de massas $p(x')$ situada na profundidade h a partir da anomalia gravitacional vertical $y^o(x)$ (Figura 1). Neste caso o problema direto pode ser

escrito como uma equação integral de Fredholm de primeira espécie

$$y^o(x) = \gamma \int_{-\infty}^{+\infty} \frac{-h}{[(x - x')^2 + h^2]^{3/2}} p(x') dx'$$

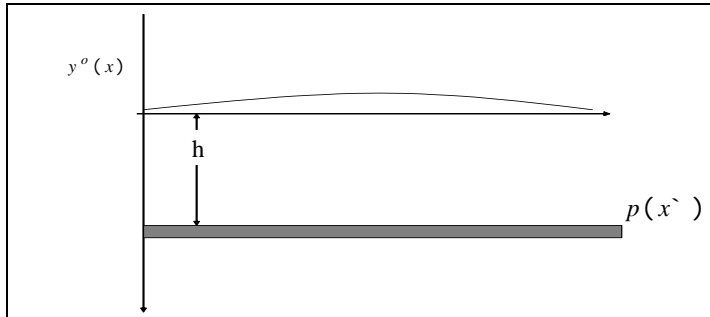


Figura 1

Como já disse anteriormente iremos trabalhar com problemas inversos discretos que iremos apresentar mais adiante em notação matricial.

2) Exemplo linear no contínuo em sismica (tomografia simplificada)

Considere o tempo de transito t de uma onda elástica gerada na fonte F e registrada no receptor R (geofone). Supõe-se que a onda atravessa o corpo (Terra) ao longo de uma trajetória L do raio, para uma velocidade contínua $v(x, z)$, então o tempo de trânsito ao longo deste caminho do raio é dado pela integral de linha:

$$t = \int_L \frac{1}{v(x, z)} dl$$

ou ainda

$$t = \int_L s(x, z) dl$$

em que $s(x, z)$ é a vagarosidade da onda (slowness). Em geral, o raio atravessa um corpo segundo uma trajetória curva devido aos efeitos de refração.

Vamos fazer algumas simplificações e estabelecer um modelo interpretativo para transformar este problema contínuo em um problema inverso discreto.

Para tanto vamos apresentar a seguir o problema linear inverso no domínio discreto.

O problema linear em notação matricial:

Existe uma relação funcional entre observações geofísicas e parâmetros do modelo que é o vetor (\bar{p}) , então a i-ésima observação geofísica coletada no ponto com coordenada x_i é

$$y_i^o = f(x_i, \bar{p}).$$

Portanto, as observações y_i^o são aproximadas pelo funcional $f(x, \bar{p})$ avaliado em $x = x_i$. Vale ressaltar que as funções na geofísica são funções diferenciáveis nas proximidades de $\bar{p} = \bar{p}^o$. Então podemos expandir a função $f(x_i, \bar{p})$ em série de Taylor (veja Apêndice - Teorema de Taylor) no entorno do ponto $\bar{p} = \bar{p}^o$:

$$y_i^o \approx f(x_i, \bar{p}) = f(x_i, \bar{p}^o) + \left. \frac{\partial}{\partial p_1} f(x, \bar{p}) \right|_{\substack{x=x_i \\ \bar{p}=\bar{p}^o}} (p_1 - p_1^o) + \left. \frac{\partial}{\partial p_2} f(x, \bar{p}) \right|_{\substack{x=x_i \\ \bar{p}=\bar{p}^o}} (p_2 - p_2^o) + \dots + \left. \frac{\partial}{\partial p_m} f(x, \bar{p}) \right|_{\substack{x=x_i \\ \bar{p}=\bar{p}^o}} (p_m - p_m^o) + \left. \frac{1}{2!} \frac{\partial^2}{\partial p_1^2} f(x, \bar{p}) \right|_{\substack{x=x_i \\ \bar{p}=\bar{p}^o}} (p_1 - p_1^o)^2 + \dots$$

Considerando a função $f(x_i, \bar{p})$ linear em relação aos parâmetros (\bar{p}) , então as derivadas superior a ordem 1 da função em relação aos parâmetros são NULAS. Isto significa que

$$y_i^o \approx f(x_i, \bar{\mathbf{p}}) = f(x_i, \bar{\mathbf{p}}) \Big|_{\substack{x=x_i \\ \bar{\mathbf{p}}=\bar{\mathbf{p}}^0}} + \frac{\partial}{\partial p_1} f(x_i, \bar{\mathbf{p}}) \Big|_{\substack{x=x_i \\ \bar{\mathbf{p}}=\bar{\mathbf{p}}^0}} (p_1 - p_1^o) +$$

$$\frac{\partial}{\partial p_2} f(x_i, \bar{\mathbf{p}}) \Big|_{\substack{x=x_i \\ \bar{\mathbf{p}}=\bar{\mathbf{p}}^0}} (p_2 - p_2^o) + \dots + \frac{\partial}{\partial p_m} f(x_i, \bar{\mathbf{p}}) \Big|_{\substack{x=x_i \\ \bar{\mathbf{p}}=\bar{\mathbf{p}}^0}} (p_m - p_m^o).$$

Na geofísica se $\bar{\mathbf{p}} = \bar{\mathbf{0}}$ a função $f(x_i, \bar{\mathbf{p}}) = 0$, ou seja, estamos tratando com funcionais para os quais:

$$\bar{\mathbf{p}} = \bar{\mathbf{0}} \Rightarrow f(x_i, \bar{\mathbf{p}}) = 0$$

Então, fazendo $\bar{\mathbf{p}}^0 \equiv \bar{\mathbf{0}}$ temos:

$$y_i^o = \frac{\partial}{\partial p_1} f(x_i, \bar{\mathbf{p}}) \Big|_{x=x_i} (p_1) + \frac{\partial}{\partial p_2} f(x_i, \bar{\mathbf{p}}) \Big|_{x=x_i} (p_2) + \dots + \frac{\partial}{\partial p_m} f(x_i, \bar{\mathbf{p}}) \Big|_{x=x_i} (p_m)$$

Note que as derivadas $\frac{\partial}{\partial p_j} f(x_i, \bar{\mathbf{p}})$ não dependem de p_j uma vez

que consideramos que f é linear em $\bar{\mathbf{p}}$. Especificamente, o termo

$\frac{\partial}{\partial p_j} f(x_i, \bar{\mathbf{p}}) \Big|_{x=x_i}$ é constante, dependendo apenas da posição da i -ésima

observação e do j -ésimo parâmetro. Assim esse termo será designado a_{ij} .

Desse modo temos:

$$y_i^o = a_{i1} p_1 + a_{i2} p_2 + a_{i3} p_3 + \dots + a_{iM} p_m.$$

Levando-se em conta que temos N observações geofísicas ($y_i^o, i = 1 \dots N$), então a equação acima é parte de um sistema de equações lineares. O sistema completo pode ser expresso em notação matricial:

$$\bar{\mathbf{y}}^o = \bar{\mathbf{A}}\bar{\mathbf{p}},$$

sendo que $\bar{\mathbf{A}}$ é a matriz de sensibilidade ($N \times M$), cujo elemento a_{ij} (ij-ésimo coeficiente de sensibilidade) é dado por:

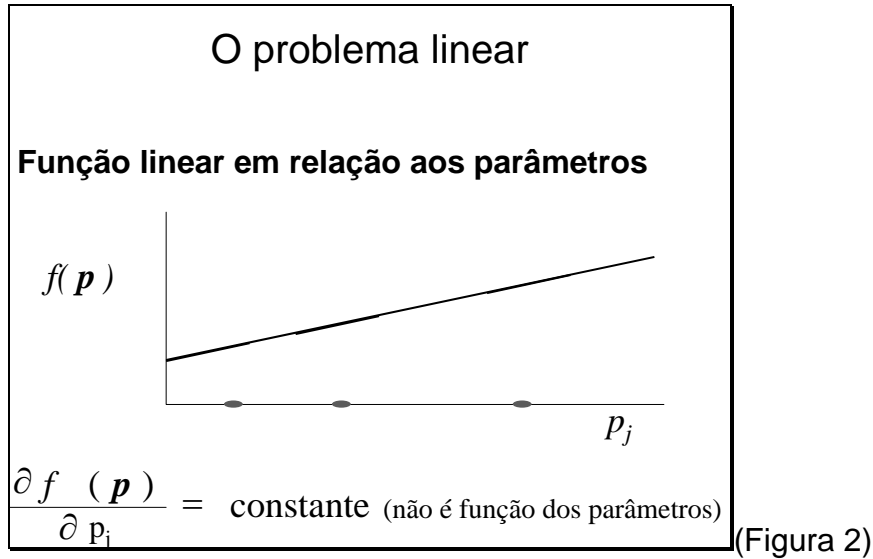
$$a_{ij} = \frac{\partial}{\partial p_j} f(x_i, \bar{\mathbf{p}}) \Big|_{x=x_i}$$

Exemplos do Problema Inverso Linear Discreto: Em Notação Matricial

Já vimos que os N dados geofísicos medidos no campo $y_i^o, i = 1, \dots, N$ representam observações de alguma grandeza física que foram coletados em pontos discretos (coordenada $x_i, i = 1, \dots, N$) na superfície da Terra. Estes dados observados, armazenados em um vetor $\bar{\mathbf{y}}^o$ ($N \times 1$), estão relacionados a alguma propriedade física da Terra que são os nossos parâmetros desconhecidos a serem estimados $\bar{\mathbf{p}}$ ($M \times 1$). Existe portanto uma equação f advinda da resolução de um sistema de equações diferenciais da física-matemática, que estabelece a relação entre $\bar{\mathbf{y}}^o$ e $\bar{\mathbf{p}}$. Esta equação f chamamos de relação funcional ou sistema físico ou modelo direto e representa uma lei DETERMINÍSTICA que estabelece a relação entre a i -ésima observação geofísica (y_i^o) medida na coordenada x_i e os parâmetros a serem estimados $\bar{\mathbf{p}}$

$$y_i^o = f(\bar{\mathbf{p}}, x_i)$$

Se a função $f(\bar{\mathbf{p}}, x_i)$ é uma função linear em relação aos parâmetros (Figura 2)



então, a $\frac{\partial f(\bar{\mathbf{p}}, x_i)}{\partial p_j}$ não é função dos parâmetros caracterizando um problema inverso linear. Em outras palavras, um problema inverso linear é aquele que a derivada da $f(\bar{\mathbf{p}}, x_i)$ em relação aos parâmetros NÃO é função dos parâmetros.

Do ponto de vista matemático a função $f(\bar{\mathbf{p}}, x_i)$ é um modelo matemático teórico onde x_i são variáveis independentes e $\bar{\mathbf{p}}$ é o vetor M-dimensional dos parâmetros deste modelo. Considerado apenas uma única função [uma única observação (N=1) em $x_1 = x$] e M parâmetros, a primeira derivada desta função $f(\bar{\mathbf{p}}, x)$ com respeito ao j-ésimo parâmetro é chamado de coeficiente de sensibilidade a_j para o parâmetro p_j

$$a_j = \frac{\partial f(\bar{\mathbf{p}}, x)}{\partial p_j}$$

Fisicamente os coeficientes de sensibilidade são importantes porque eles indicam a variação de magnitude da resposta da função $f(\bar{\mathbf{p}}, x)$ devido a perturbações nos valores dos parâmetros. Em uma situação realista, temos N observações e M parâmetros o que implica a existência de N funções $f_i(\bar{\mathbf{p}}, x_i)$,

$i = 1, \dots, N$. Neste caso a função $f_i(\bar{\mathbf{p}}, x_i)$ é chamada de função multiresposta e o coeficiente de sensibilidade a_{ij} é a primeira derivada da i-ésima função $f_i(\bar{\mathbf{p}}, x_i)$ em relação ao j-ésimo parâmetro p_j .

$$a_{ij} = \frac{\partial f_i(\bar{\mathbf{p}}, x_i)}{\partial p_j}$$

Como definimos N funções $f_i(\bar{\mathbf{p}}, x_i)$, $i = 1, \dots, N$, e M parâmetros p_j , $j = 1, \dots, M$ então teremos a definição de uma matriz de sensibilidade $\bar{\bar{\mathbf{A}}}$ de

dimensões $N \times M$, cujo i-ésimo elemento é $a_{ij} = \frac{\partial f_i(\bar{\mathbf{p}}, x_i)}{\partial p_j}$

Matriz de Sensibilidade $\mathbf{A}(N \times M)$:

j-ésima coluna

$$\bar{\bar{\mathbf{A}}} = \begin{pmatrix} \frac{\partial f_1}{\partial p_1} & \frac{\partial f_1}{\partial p_2} & \dots & \frac{\partial f_1}{\partial p_j} & \dots & \frac{\partial f_1}{\partial p_M} \\ \frac{\partial f_2}{\partial p_1} & \frac{\partial f_2}{\partial p_2} & \dots & \frac{\partial f_2}{\partial p_j} & \dots & \frac{\partial f_2}{\partial p_M} \\ \vdots & \vdots & & \vdots & & \vdots \\ \frac{\partial f_i}{\partial p_1} & \frac{\partial f_i}{\partial p_2} & \dots & \frac{\partial f_i}{\partial p_j} & \dots & \frac{\partial f_i}{\partial p_M} \\ \vdots & \vdots & & \vdots & & \vdots \\ \frac{\partial f_N}{\partial p_1} & \frac{\partial f_N}{\partial p_2} & \dots & \frac{\partial f_N}{\partial p_j} & \dots & \frac{\partial f_N}{\partial p_M} \end{pmatrix}$$

i-ésima linha

Figura 3

Veja que a i-ésima linha da matriz de sensibilidade $\bar{\bar{\mathbf{A}}}$ (Figura 3) carrega a variação de magnitude (sensibilidade) da função $f_i(\bar{\mathbf{p}}, x_i)$ devido a perturbações nos valores dos M parâmetros e a j-ésima coluna da matriz de sensibilidade $\bar{\bar{\mathbf{A}}}$ (Figura 3) é a influência da perturbação do j-ésimo parâmetro (p_j) nas N funções (N observações geofísicas)

Exemplo 1 de um problema inverso linear: Regressão linear estimar o IGRF

Este primeiro exemplo já foi anteriormente apresentado consistindo em estimar o campo geomagnético local da Terra (IGRF) a partir de N observações de uma anomalia magnética observada. Como simplificação, presumimos que o IGRF pode ser aproximado por um polinômio de ordem 1. Na Figura 4 os N pontos pretos representam as N observações da anomalia magnética $(y^o_i, i = 1, \dots, N)$.

Neste exemplo, existe uma relação funcional $f_i(\bar{\mathbf{p}}, x_i)$ entre os dados observados (vetor $\bar{\mathbf{y}}^o$) e os parâmetros a serem estimados (vetor $\bar{\mathbf{p}}$) que estão relacionados ao IGRF

$$y^o_i = f_i(\bar{\mathbf{p}}, x_i)$$

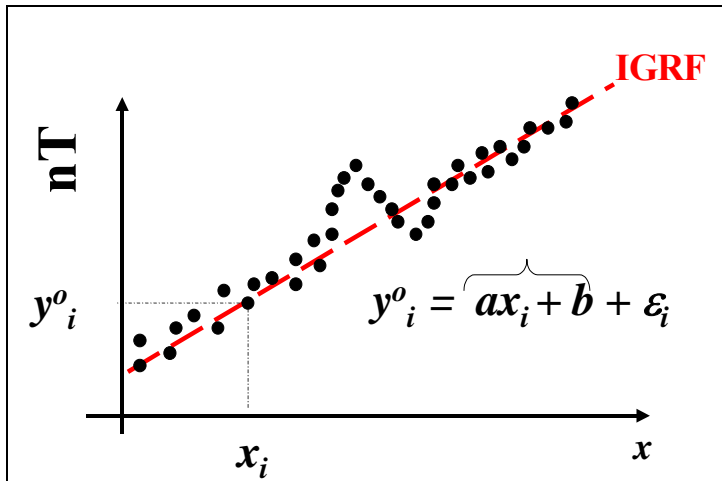


Figura 4

Como presumimos uma variação linear do campo geomagnético local da Terra (IGRF), a relação funcional entre a i -ésima observação geofísica medida na coordenada x_i e os parâmetros a serem estimados vetor $\bar{\mathbf{p}}$ é

$$f_i(\bar{\mathbf{p}}, x_i) = ax_i + b, \quad i = 1, \dots, N$$

onde N é o número de observações realizadas de uma anomalia magnética. A variável x_i é a i -ésima coordenada de localização da medida geofísica e a e b são os parâmetros desconhecidos a serem estimados. Vamos considerar que e

a e b são, respectivamente, o primeiro e segundo elementos do vetor de parâmetros $\bar{\mathbf{p}}$. Desta forma a relação funcional, neste exemplo, pode ser escrita como

$$f_i(\bar{\mathbf{p}}, x_i) = p_1 x_i + p_2, \quad i = 1, \dots, N.$$

Neste problema teremos os seguintes coeficientes de sensibilidade a_{ij} :

$$a_{i1} = \frac{\partial f_i(\bar{\mathbf{p}}, x_i)}{\partial p_1} = x_i, \quad i = 1, \dots, N$$

e

$$a_{i2} = \frac{\partial f_i(\bar{\mathbf{p}}, x_i)}{\partial p_2} = 1, \quad i = 1, \dots, N.$$

Veja que os coeficiente de sensibilidade NÃO dependem dos parâmetros, então a função $f_i(\bar{\mathbf{p}}, x_i)$ é linear em relação aos parâmetros. Neste exemplo então a matriz de sensibilidade $\bar{\bar{\mathbf{A}}}$ (N x M) é

$$\bar{\bar{\mathbf{A}}} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix}$$

Exemplo 2 de um problema inverso linear discreto em gravimetria:

Este segundo exemplo consiste em transformar o problema linear inverso contínuo apresentado na Figura 1 em um problema linear inverso discreto.

Já vimos que o problema inverso contínuo de estimar a distribuição de densidades numa linha de massas $\rho(x)$ situada na profundidade h a partir da

anomalia gravitacional vertical $y^o(x)$ (Figura 1), pode ser escrito como uma equação integral de Fredholm de primeira espécie $y^o(x) = \gamma \int_{-\infty}^{+\infty} \frac{-h}{[(x-x')^2 + h^2]^{3/2}} p(x') dx'$

No discreto faremos algumas simplificações e estabeleceremos um modelo interpretativo. Neste problema acima poderíamos discretizar a linha de massa em M prismas finitos. O j-ésimo prisma tem densidade constante p_j e dimensões conhecidas e iguais a x_1^j e x_2^j (Figura 5). Presumiremos que a densidade é constante dentro de um dos prismas e que de um prisma para outro a densidade possa variar. Presumiremos também um número finito N de observações da anomalia gravimétrica.

A relação funcional deste problema estabelece a relação entre a i-ésima observação da anomalia gravimétrica $y^o_i(x_i)$ realizada na superfície da Terra na coordenada x_i e as densidades dos M prismas que compõe o modelo interpretativo é escrita como:

$$y^o_i(x_i) = \gamma \sum_{j=1}^M p_j \int_{x_1^j}^{x_2^j} \frac{-h}{[(x_i - x')^2 + h^2]^{3/2}} dx'$$

Veja que este problema é linear em relação aos parâmetros (densidades de cada prisma).

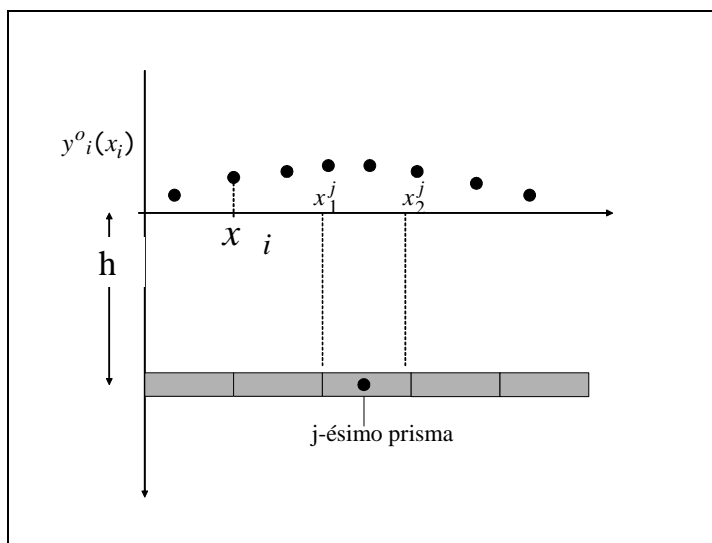


Figura 5

O problema acima poderia ser escrito como:

$$y_i^o = \sum_{j=1}^M p_j a_{ij}$$

em que a_{ij} é dado por

$$a_{ij} = \gamma \int_{x_1^j}^{x_2^j} \frac{h}{[(x_i - x_j)^2 + h^2]^{3/2}} dx$$

e representa o i -ésimo elemento da matriz $\bar{\bar{A}}$ ($N \times M$). Então em notação matricial este problema poderia ser escrito como

$$\bar{\bar{A}}\bar{\bar{p}} = \bar{\bar{y}}^o$$

Exemplo 3 de um problema inverso linear discreto em sísica (tomografia simplificada):

Este terceiro exemplo linear também já foi apresentado acima no contínuo e consiste de uma tomografia sísmica simplificada da Terra. Nosso problema inverso é estimar a variação espacial da vagarosidade das ondas sísmicas, a partir da medida de tempos de trânsito de uma onda elástica gerada por fontes F e registrada nos geofones R , ambos localizados na superfície da Terra. Presumimos as seguintes simplificações: 1) conhecemos a forma da Terra que é um quadrado; 2) conhecemos as localizações Fonte – Receptor; 3) eliminamos o efeito da refração, logo a onda elástica percorre a distância F - R em linha reta; 4) presumimos que 9 células quadradas são suficientes para descrever a variação espacial da vagarosidade das ondas sísmicas no interior de Terra e; 5) presumimos que a vagarosidade é constante dentro de uma célula quadrada e que de uma célula para outra a vagarosidade pode variar.

Neste exemplo, existe uma relação funcional $f_i(\bar{\mathbf{p}})$ entre os dados observados (tempos de transito da onda eslática) e os parâmetros a serem estimados (vetor $\bar{\mathbf{p}}$) que são as vagarosidades das 9 células. Vamos

estabelecer esta relação funcional $f_i(\bar{\mathbf{p}})$. Dado uma fonte F_i e um receptor R_i o tempo de trânsito t_i que uma onda elástica leva para percorrer a distância $F_i R_i$ será a soma dos tempos pela onda para atravessar cada célula envolvida no seu percurso. No caso da Figura 6 temos

$$t_i = f_i(\bar{\mathbf{p}})$$

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

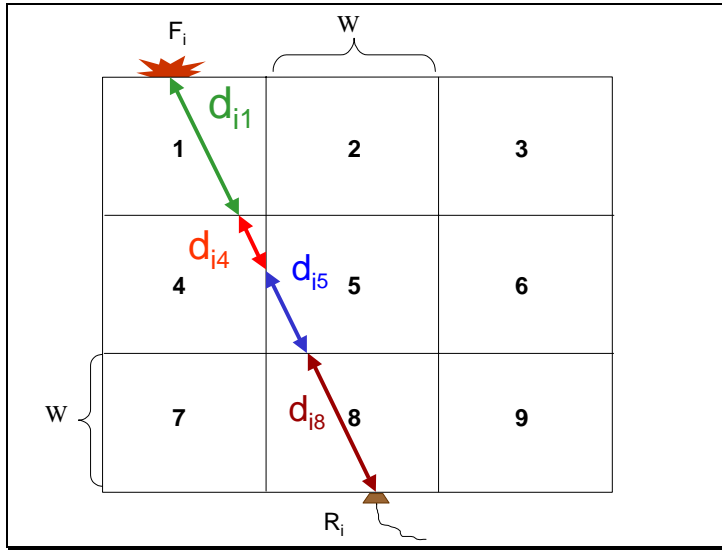


Figura 6

em que d_{ij} é a distância que a onda associada a i -ésima observação (t_i) percorre dentro da j -ésima célula e p_j é a vagarosidade da j -ésima célula. Veja que, neste segundo exemplo, podemos escrever que a relação funcional entre a i -ésima observação geofísica (tempo de trânsito t_i) e os parâmetros do modelo de 9 células (vetor $\bar{\mathbf{p}}$) é:

$$f_i(\bar{\mathbf{p}}) = \sum_{j=1}^M d_{ij} p_j, \quad i = 1, \dots, N$$

onde N é o número de observações, ou seja, N posições diferentes para o par Fonte-Receptor, M é o número de parâmetros que são as vagarosidades das 9

células (M=9). Note que os coeficientes de sensibilidade a_{ij} NÃO dependem dos parâmetros, uma vez que neste caso

$$a_{ij} = \frac{\partial f_i(\bar{\mathbf{p}})}{\partial p_j} = d_{ij}.$$

Portanto, a_{ij} depende apenas da distância que a onda associada a i-ésima observação do tempo de trânsito (t_i), que está associado ao par $F_i R_i$, leva para percorrer a j-ésima célula.

Vamos considerar um caso em que temos N=2, ou seja, temos duas medidas de tempo de trânsito associadas a duas posições diferentes para o par Fonte-Receptor e M= 9 células como mostra a Figura 7.

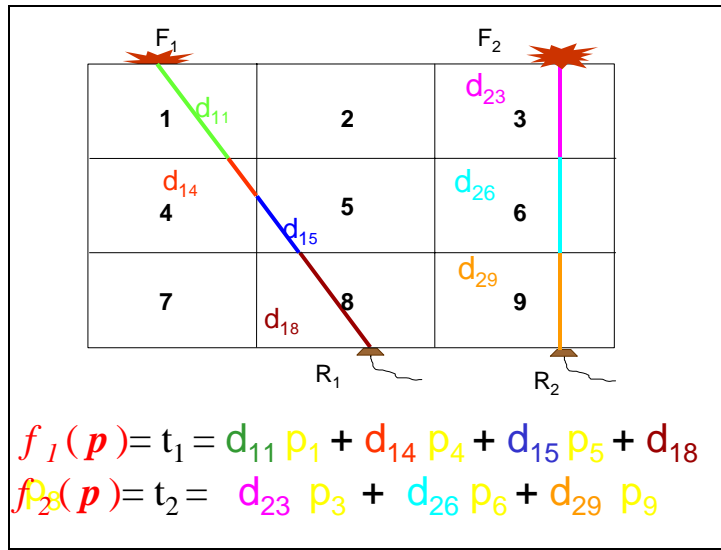


Figura 7

Matematicamente a i-ésima linha da matriz de sensibilidade é

$$\left[\frac{\partial f_i(\bar{\mathbf{p}})}{\partial p_1} \quad \frac{\partial f_i(\bar{\mathbf{p}})}{\partial p_2} \quad \frac{\partial f_i(\bar{\mathbf{p}})}{\partial p_3} \quad \dots \quad \frac{\partial f_i(\bar{\mathbf{p}})}{\partial p_9} \right]$$

Neste caso a matriz de sensibilidade terá duas linhas sendo que a primeira linha é

$$[d_{11} \quad 0 \quad 0 \quad d_{14} \quad d_{15} \quad 0 \quad 0 \quad d_{18} \quad 0]$$

e a segunda linha é

$$\begin{bmatrix} 0 & 0 & d_{23} & 0 & 0 & d_{26} & 0 & 0 & d_{29} \end{bmatrix}.$$

Logo a matriz de sensibilidade deste problema (Figura 7) é tem dimensões (2 x 9) é escrita como:

$$\overline{\overline{\mathbf{A}}} = \begin{bmatrix} d_{11} & 0 & 0 & d_{14} & d_{15} & 0 & 0 & d_{18} & 0 \\ 0 & 0 & d_{23} & 0 & 0 & d_{26} & 0 & 0 & d_{29} \end{bmatrix}.$$

O problema linear em notação matricial considerando a presença do ruído nos dados observados:

Vimos então que em um problema linear a relação funcional entre os M parâmetros e as N observações geofísicas pode ser representada por um sistema de N equações por M incógnitas:

$$\overline{\mathbf{y}}^o = \overline{\overline{\mathbf{A}}} \overline{\mathbf{p}},$$

Então a equação acima estabelece que há uma relação determinística entre os dados geofísicos observados e os parâmetros que é dada pela multiplicação da matriz de sensibilidade $\overline{\overline{\mathbf{A}}}$ pelo vetor de parâmetros $\overline{\mathbf{p}}$.

Como já discutimos anteriormente, em problemas reais não conhecemos os dados geofísicos $\overline{\mathbf{y}}^o$ com total precisão, em geral, estes dados contem componentes aleatórias que não são explicadas pela relação funcional $f(\overline{\mathbf{p}}, \overline{\mathbf{x}})$ (que são equações determinísticas). Então para cada observação geofísica y_i^o presumimos a existência de um erro (ruído experimental) ε_i que é uma variável aleatória. Conseqüentemente, em problemas reais os dados observados geofísicos $\overline{\mathbf{y}}^o$ são, matematicamente, constituídos de uma parte determinística que chamaremos dados ajustado ou calculado $\overline{\mathbf{y}}^c$ (N x 1) e uma parte aleatória aditiva representada pelo vetor

$\bar{\epsilon}$ (N x 1) . Como já vimos, em um problema linear a componente determinística (vetor dos dados ajustados ou calculados) é

$$\bar{y}^C = \bar{A} \bar{p}$$

Desta forma, matematicamente, no problema linear temos a seguinte equação:

$$\bar{y}^O = \bar{y}^C + \bar{\epsilon}$$

ou ainda

$$\bar{y}^O = \bar{A} \bar{p} + \bar{\epsilon}$$

Nosso problema de estimar os parâmetros a partir dos dados geofísicos observados \bar{y}^O é, neste caso, um problema inverso linear. No entanto, na solução que qualquer problema inverso surgem três perguntas cruciais:

- 1) A solução existe ?
- 2) Ela é única ?
- 3) Ela é estável ?

A importância destas três perguntas está relacionada a aceitação da solução estimada. De acordo com Hadamard (1902) apenas aceitamos a solução estimada de um problema inverso se, e somente se, as respostas destas três perguntas formuladas forem afirmativas. Em outras palavras, o problema matemático é chamado de bem-posto, se sua solução existe, for única e estável.

Então pergunta-se: Será que o nosso problema é bem-posto ? Para responder a esta pergunta vamos ter uma noção preliminar sobre os conceitos de existência, unicidade e estabilidade de uma solução¹ Para tanto, usaremos conceitos mais genéricos em que os dados geofísicos observados e parâmetros serão funções contínuas (vetores seriam um caso particular) e especificaremos um operador relacionando os dados observados aos parâmetros.

¹ Mais a diante, estes conceitos serão abordados de modo mais completo para fundamentar a teoria da regularização de Tikhonov

O problema da Existência e Unicidade:

Seja p uma função desconhecida pertencente ao sub-espço F do espaço de P de todos os possíveis parâmetros e seja uma função conhecida y pertencente ao sub-espço métrico AF de um espaço Y de todas as possíveis observações (Figura 8). Considere A um operador linear que mapeia p em y . Portanto, nosso problema reduz-se a resolver a equação linear dada por

$$y = Ap$$

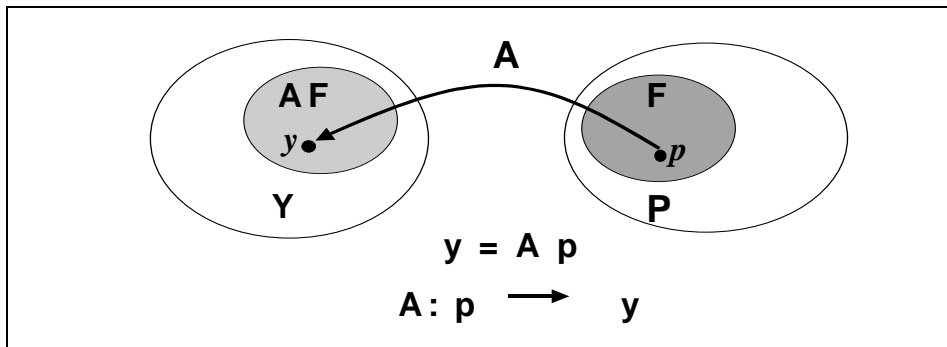


Figura 8

Note que um elemento $y \in AF$ representa um elemento produzido por uma transformação AP , cujo operador A envolve relações de causa e efeito obtidas da resolução de um sistema de equações diferenciais da física-matemática (equação de Poisson, equação de Laplace, equações de Maxwell, equação da onda). No entanto, em problemas práticos não conhecemos y com total precisão, em geral, conhecemos uma medida aproximada y^o que contém componentes aleatórias (ε) que não são explicadas por estas equações. Assim a medida y^o pode não pertencer ao sub-espço AF (Figura 9).

$$y^o = y + \varepsilon$$

Neste caso NÃO EXISTE SOLUÇÃO PARA O SISTEMA $y^o = Ap$, uma vez que o operador A não é definido.

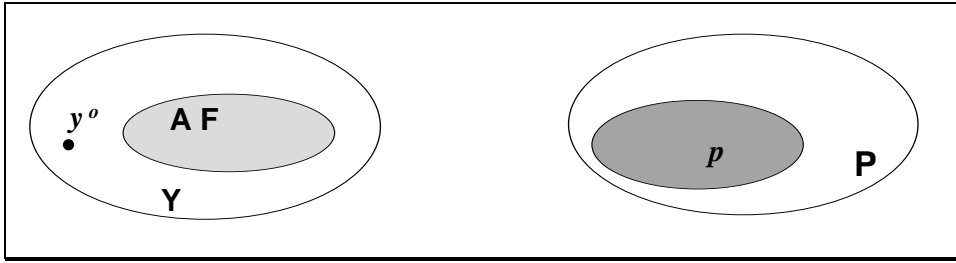


Figura 9

A inexistência de solução levou ao conceito de **Quasesolução**. Para tanto considere uma projeção de y^o no sub-espço AF (Figura 10). Esta projeção y^c , será tal que a distância entre y^c e y^o será mínima dentre todos os elementos $y \in AF$, ou seja $d(y^c, y^o) \leq \delta$ em que d é uma métrica definida no espaço Y e δ um escalar positivo que tende a zero a medida que y^o tende a y^c .

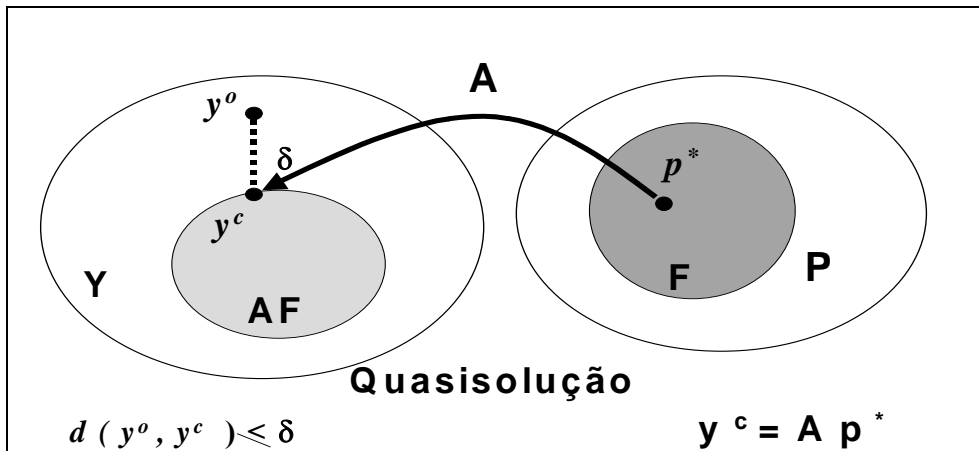


Figura 10

Agora o operador A é definido garantindo a existência da solução p^* denominadas de **Quasesolução** (Figura 10). Logo temos que o operador A mapeia p^* em y^c resultando na equação: $y^c = A p^*$.

No entanto, se o sub-espço AF for não convexo (Figura 11) existirá infinitas projeções de y^o em AF uma vez que há infinitas métricas $d(y^c, y^o) \leq \delta$. Se há infinitas projeções y^c há portanto infinitas quasisoluções p^* . Neste caso a existência da solução foi garantida mas a unicidade da solução não.

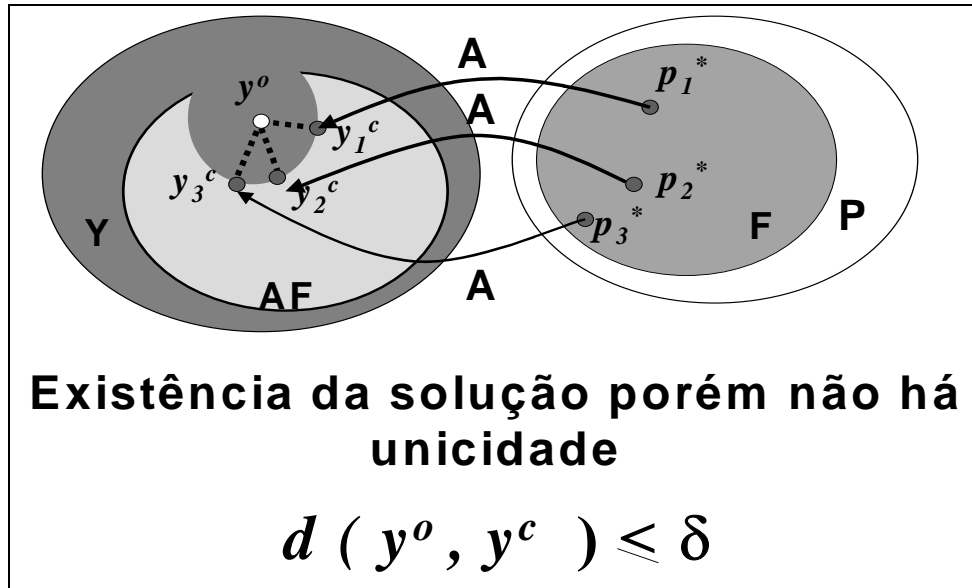


Figura 11

Desta forma temos a primeira condição de UNICIDADE DA SOLUÇÃO que garante a não ocorrência da situação mostrada na Figura 11.

PRIMEIRA CONDIÇÃO DE UNICIDADE DA SOLUÇÃO:

Para a garantia da unicidade da solução temos que garantir a unicidade da projeção y^o no sub-espço AF. Para tanto **AF tem que ser CONVEXO** (Figura 12). Como o operador A é linear se AF é convexo então F é convexo também.

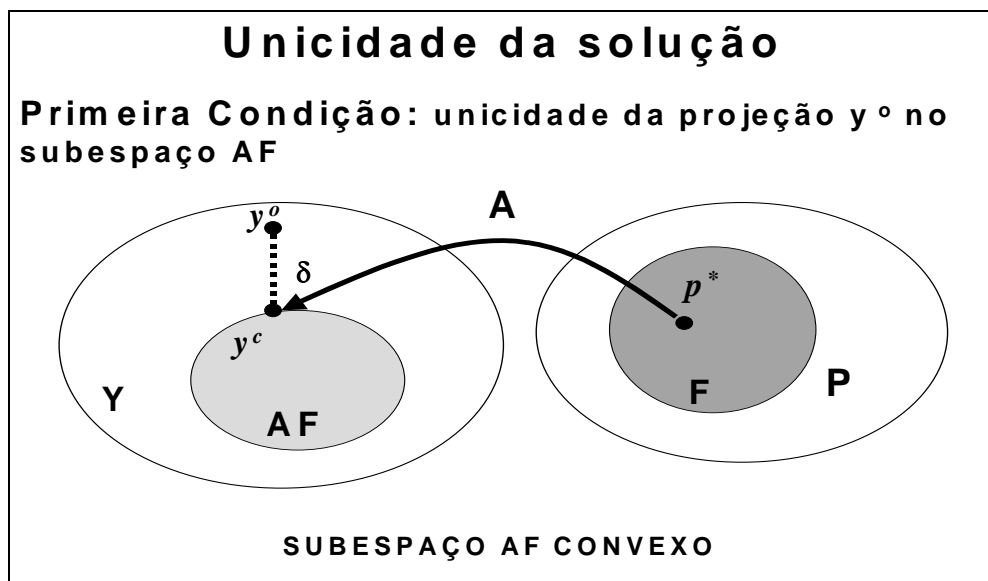


Figura 12

SEGUNDA CONDIÇÃO DE UNICIDADE DA SOLUÇÃO:

A segunda condição para a garantia da unicidade da solução é a garantia da unicidade do operador linear A que deve ser único um por um. Logo para o operador A que mapeia F em AF há um operador inverso A^{-1} que mapeia AF em F (Figura 13). Esta condição implica $p^* = 0$ é a única solução da equação $A p^* = 0$

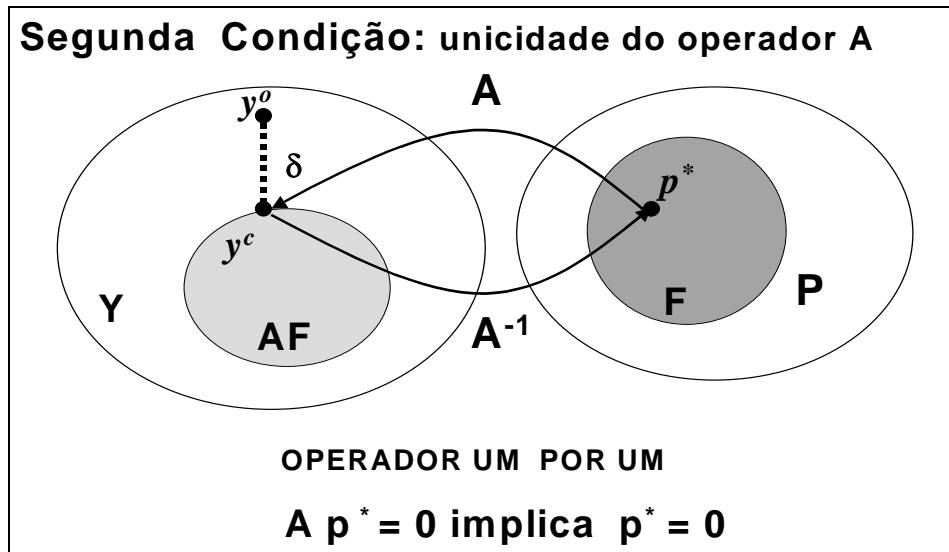


Figura 13

TERCEIRA CONDIÇÃO DE UNICIDADE DA SOLUÇÃO:

A terceira condição para a garantia da unicidade da solução é a garantia da continuidade do operador linear A . A continuidade do operador A leva ao importante conceito de estabilidade da solução. Para garantia da continuidade do operador A o sub-espaço F deve ser compacto e convexo. Esta condição nós veremos mais à frente, em detalhes, porque ela introduz um importante conceito dentro da teoria da inversão o conceito de estabilidade. Por enquanto ficaremos com o conceito heurístico de estabilidade. Uma solução é estável se uma pequena perturbação no espaço das observações (AF) leva a uma pequena perturbação no espaço dos parâmetros (espaço das soluções F). Conseqüentemente, uma solução é instável se uma pequena perturbação no espaço das observações leva a uma grande perturbação no espaço dos parâmetros.

Em outras palavras, instabilidade reflete um fato prático que dois conjuntos de observações podem diferir entre si dentro do nível de ruído experimental enquanto que os correspondentes conjuntos de parâmetros estimados podem ser completamente diferentes.

A seguir apresentaremos uma exemplificação do conceito de existência. Dentro deste exemplo, analisaremos os conceitos de unicidade e estabilidade da solução estimada.

Exemplificando a Existência e Unicidade:

Estimador de Mínimos quadrados:

Vamos considerar o caso particular de um problema inverso linear discreto, em que temos um vetor N-dimensional de observações \bar{y}^0 e queremos estimar um vetor M-dimensional de parâmetros \bar{p} . Neste caso particular, o operador A é um operador linear, a relação funcional pode ser escrita em notação matricial:

$$\bar{y}^C = \bar{A} \bar{p}$$

em que \bar{A} é a matriz M x N (matriz de sensibilidade) do operador linear A e \bar{y}^C chamaremos de vetor N-dimensional dos dados ajustado ou calculado. Particularizando agora a distância entre \bar{y}^0 e \bar{y}^C como sendo a distância Euclideana para garantir a existência da solução. Então queremos encontrar um vetor de parâmetros \hat{p} tal que $\|\bar{y}^0 - \bar{A} \hat{p}\|_2$, $\hat{p} \in F$, seja mínimo.

A Existência da solução: Então para a garantia da existência de uma solução formulamos um problema de mínimos quadrados:

$$\min_{\bar{\mathbf{p}} \in F} \left\| \bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}} \right\|_2^2$$

$$\min_{\bar{\mathbf{p}} \in F} \left(\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}} \right)^T \left(\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}} \right) = \min_{\bar{\mathbf{p}} \in F} \{Q\}$$

A condição necessária para Q ter um mínimo é

$$\bar{\nabla}_{\bar{\mathbf{p}}} \{Q\} = \bar{\mathbf{0}}$$

$$\bar{\nabla}_{\bar{\mathbf{p}}} \{Q\} = \bar{\nabla}_{\bar{\mathbf{p}}} \left\{ \left(\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}} \right)^T \left(\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}} \right) \right\}$$

$$\bar{\nabla}_{\bar{\mathbf{p}}} \{Q\} = 2 \bar{\nabla}_{\bar{\mathbf{p}}} \left\{ \left(\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}} \right)^T \right\} \left(\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}} \right)$$

$$\bar{\nabla}_{\bar{\mathbf{p}}} \{Q\} = 2 \left(\bar{\nabla}_{\bar{\mathbf{p}}} \left\{ \bar{\mathbf{y}}^0{}^T \right\} - \bar{\nabla}_{\bar{\mathbf{p}}} \left\{ \bar{\mathbf{p}}^T \bar{\mathbf{A}}^T \right\} \right) \left(\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}} \right)$$

$$\bar{\nabla}_{\bar{\mathbf{p}}} \{Q\} = 2 \left(-\bar{\mathbf{I}} \bar{\mathbf{A}}^T \right) \left(\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}} \right)$$

$$\bar{\nabla}_{\bar{\mathbf{p}}} \{Q\} = -2 \bar{\mathbf{A}}^T \left(\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}} \right)$$

Igualando $\bar{\nabla}_{\bar{\mathbf{p}}} \{Q\} = \bar{\mathbf{0}}$ Temos

$$-\bar{\mathbf{A}}^T \bar{\mathbf{y}}^0 + \bar{\mathbf{A}}^T \bar{\mathbf{A}} \hat{\bar{\mathbf{p}}} = \bar{\mathbf{0}}$$

$$\bar{\mathbf{A}}^T \bar{\mathbf{A}} \hat{\bar{\mathbf{p}}} = \bar{\mathbf{A}}^T \bar{\mathbf{y}}^0$$

$$\hat{\bar{\mathbf{p}}} = \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \bar{\mathbf{y}}^0$$

Então a escolha da distância Euclideana mínima entre $\bar{\mathbf{y}}^0$ e $\bar{\mathbf{y}}^c$ garantiu a existência de uma solução $\hat{\bar{\mathbf{p}}}$ que chamaremos de **ESTIMADOR DE MÍNIMOS QUADRADOS (MQ)**, i.e.:

$$\hat{\bar{\mathbf{p}}} = \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \bar{\mathbf{y}}^0$$

Análise da Unicidade da solução:

A solução $\hat{\mathbf{p}}$ só será única se a matriz a ser invertida ($\overline{\mathbf{A}}^T \overline{\mathbf{A}}$) for NÃO SINGULAR, $\hat{\mathbf{p}} \equiv \overline{\mathbf{0}}$ (a solução trivial) é a única solução do sistema de equações $\overline{\mathbf{A}} \hat{\mathbf{p}} = \overline{\mathbf{0}}$ (sistema homogêneo). Um meio de avaliarmos se a matriz é não singular é através do seu determinante. Se o determinante a matriz $\overline{\mathbf{A}}^T \overline{\mathbf{A}}$ é diferente de zero ($\text{Det}(\overline{\mathbf{A}}^T \overline{\mathbf{A}}) \neq 0$) esta matriz é não singular. Um segundo meio de avaliarmos se uma matriz é não singular é através do seu posto². Como a matriz de sensibilidade $\overline{\mathbf{A}}$ é N x M, o produto $\overline{\mathbf{A}}^T \overline{\mathbf{A}}$ é M x M então, a matriz $\overline{\mathbf{A}}^T \overline{\mathbf{A}}$ deve ter posto M (número de parâmetros) para ser uma matriz não singular ($r(\overline{\mathbf{A}}^T \overline{\mathbf{A}}) = M$). Como o posto de uma matriz que é obtida pelo produto de duas outras matrizes é no máximo igual ao menor posto das matrizes, então temos que o posto de $\overline{\mathbf{A}}^T \overline{\mathbf{A}} \leq \min(M, N)$, presumindo-se que $\overline{\mathbf{A}}$ tem posto completo, isto é seu posto é $\min(M, N)$. Desta forma para usarmos o estimador de mínimos quadrados é necessário, embora não suficiente que $N \geq M$, isto é que o número de observações (N) seja pelo menos igual ao número de parâmetros (M). Para garantirmos a unicidade da solução estimada via estimador MQ é necessário e suficiente que $N \geq M$ e o posto da matriz $\overline{\mathbf{A}}^T \overline{\mathbf{A}}$ seja igual a M ($r(\overline{\mathbf{A}}^T \overline{\mathbf{A}}) = M$).

Caso contrário, se ($\overline{\mathbf{A}}^T \overline{\mathbf{A}}$) for singular então, $\hat{\mathbf{p}} \equiv \overline{\mathbf{0}}$ Unão é a única solução da equação normal $\overline{\mathbf{A}}^T \overline{\mathbf{A}} \hat{\mathbf{p}} = \overline{\mathbf{A}}^T \overline{\mathbf{y}}^0$ e a segunda condição mencionada anteriormente para a unicidade da solução não é satisfeita. Neste

² posto de uma matriz é o número de vetores colunas (ou vetores linhas) que formam um conjunto LI de vetores

caso em que não há unicidade da solução a matriz $\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}}$ é singular, então o seu determinante ($\text{Det}(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}}) = 0$) é igual a zero e o seu posto é menor que o número de parâmetros a serem estimados ($r(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}}) < M$).

Vamos supor que a matriz $\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}}$ é não singular, portanto garantimos a unicidade da solução. Já falamos anteriormente, que a unicidade da solução é uma condição necessária, mas não é suficiente para afirmarmos que estamos resolvendo um problema bem-posto. Falta analisar a estabilidade da solução estimada (terceira condição mencionada anteriormente) que estudaremos aqui através de uma análise estatística.

Análise da Estabilidade da solução MQ via análise estatística:

Como já abordamos anteriormente, o estudo da estabilidade da solução pode ser realizado através da variância dos parâmetros, uma vez que o estimador $\hat{\mathbf{p}}$ é uma variável aleatória. Vamos analisar se o estimador $\hat{\mathbf{p}}$ apresenta variância mínima. Lembramos que, à luz da estatística um bom estimador $\hat{\mathbf{p}}$ deve ser não tendencioso e de variância mínima. Então vamos calcular a esperança e variância do estimador de mínimos quadrados $\hat{\mathbf{p}}$. Para tanto precisamos estabelecer alguma premissas estatísticas sobre o ruído que contamina os dados (veja Anexo 1 do tópico 3) e também sobre as demais variáveis do modelo matemático.

(1) Esperança do estimador MQ ($\hat{\mathbf{p}}$) :

Considerando as premissas estatísticas estabelecidas na Figura 14 e as propriedades da esperança, temos que

$$\mathbf{E} \left\{ \hat{\mathbf{p}} \right\} = \mathbf{E} \left\{ \left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}} \right)^{-1} \overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{y}}}^0 \right\}$$

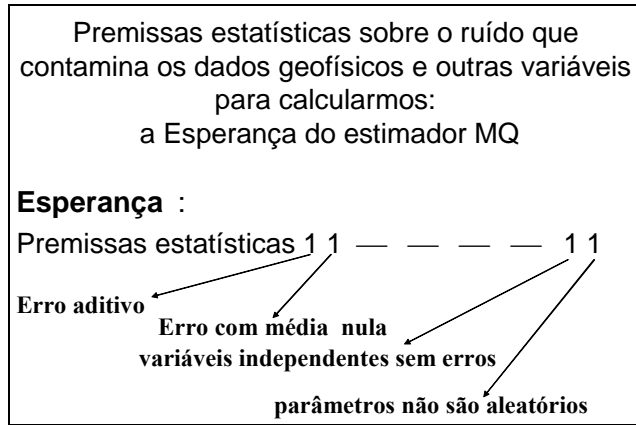


Figura 14

Usando a premissa 1 que os erros são aditivos ($\bar{y}^o = \bar{y}^c + \bar{\varepsilon}$) temos

$$E \left\{ \hat{\bar{p}} \right\} = E \left\{ \left(\bar{A}^T \bar{A} \right)^{-1} \bar{A}^T (\bar{y}^c + \bar{\varepsilon}) \right\}$$

$$E \left\{ \hat{\bar{p}} \right\} = E \left\{ \left(\bar{A}^T \bar{A} \right)^{-1} \bar{A}^T \bar{y}^c \right\} + E \left\{ \left(\bar{A}^T \bar{A} \right)^{-1} \bar{A}^T \bar{\varepsilon} \right\}$$

Pela premissa 2 os erros tem média nula o que implica $E[\bar{\varepsilon}] = \bar{0}$ então

$$E \left\{ \hat{\bar{p}} \right\} = E \left\{ \left(\bar{A}^T \bar{A} \right)^{-1} \bar{A}^T \bar{y}^c \right\}$$

Usando a informação que em um problema linear a componente determinística (vetor dos dados ajustados ou calculados) é $\bar{y}^c = \bar{A} \bar{p}$ temos que

$$E \left\{ \hat{\bar{p}} \right\} = E \left\{ \left(\bar{A}^T \bar{A} \right)^{-1} \bar{A}^T \bar{A} \bar{p} \right\}$$

$$E \left\{ \hat{\bar{p}} \right\} = \bar{p}$$

Vemos então que o estimador MQ ($\hat{\bar{p}}$) é NÃO TENDENCIOSO. De fato, o estimador MQ não introduz nenhum tipo de informação a priori sobre os parâmetros a serem estimado.

2) Matriz de covariância do estimador MQ ($\hat{\mathbf{p}}$):

Considerando as premissas estatísticas estabelecidas na Figura 15, as propriedades da variância e partindo-se da equação da matriz de covariância de um vetor (Anexo 1) temos que

$$\mathbf{COV} \left\{ \hat{\mathbf{p}} \right\} = E \left\{ \left[\hat{\mathbf{p}} - E(\hat{\mathbf{p}}) \right] \left[\hat{\mathbf{p}} - E(\hat{\mathbf{p}}) \right]^T \right\}$$

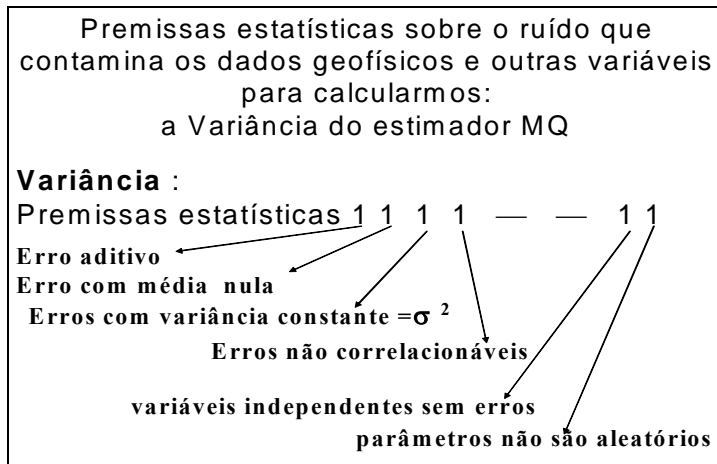


Figura 15

Usando o resultado anterior e $E \left\{ \hat{\mathbf{p}} \right\} = \bar{\mathbf{p}}$

$$\mathbf{COV} \left\{ \hat{\mathbf{p}} \right\} = E \left\{ \left[\hat{\mathbf{p}} - \bar{\mathbf{p}} \right] \left[\hat{\mathbf{p}} - \bar{\mathbf{p}} \right]^T \right\}$$

$$\mathbf{COV} \left\{ \hat{\mathbf{p}} \right\} = E \left\{ \left[\left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \bar{\mathbf{y}}^o - \bar{\mathbf{p}} \right] \left[\left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \bar{\mathbf{y}}^o - \bar{\mathbf{p}} \right]^T \right\}$$

Usando a premissa 1 que os erros são aditivos ($\bar{\mathbf{y}}^o = \bar{\mathbf{y}}^c + \bar{\boldsymbol{\varepsilon}}$) e $\bar{\mathbf{y}}^c = \bar{\mathbf{A}} \bar{\mathbf{p}}$

$$\text{COV} \left\{ \hat{\bar{\mathbf{p}}} \right\} = E \left\{ \left[\left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \left(\bar{\mathbf{A}} \bar{\mathbf{p}} + \bar{\boldsymbol{\varepsilon}} \right) - \bar{\mathbf{p}} \right] \left[\left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \left(\bar{\mathbf{A}} \bar{\mathbf{p}} + \bar{\boldsymbol{\varepsilon}} \right) - \bar{\mathbf{p}} \right]^T \right\}$$

$$\text{COV} \left\{ \hat{\bar{\mathbf{p}}} \right\} = E \left\{ \left[\bar{\mathbf{p}} + \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \bar{\boldsymbol{\varepsilon}} - \bar{\mathbf{p}} \right] \left[\bar{\mathbf{p}} + \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \bar{\boldsymbol{\varepsilon}} - \bar{\mathbf{p}} \right]^T \right\}$$

$$\text{COV} \left\{ \hat{\bar{\mathbf{p}}} \right\} = E \left\{ \left[\left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \bar{\boldsymbol{\varepsilon}} \right] \left[\left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \bar{\boldsymbol{\varepsilon}} \right]^T \right\}$$

Usando a propriedade de transposição temos :

$$\text{COV} \left\{ \hat{\bar{\mathbf{p}}} \right\} = E \left\{ \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \bar{\boldsymbol{\varepsilon}} \bar{\boldsymbol{\varepsilon}}^T \bar{\mathbf{A}} \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \right\}$$

Usando a premissa 7 que as variáveis independentes não são v.a. temos,

$$\text{COV} \left\{ \hat{\bar{\mathbf{p}}} \right\} = \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T E \left\{ \bar{\boldsymbol{\varepsilon}} \bar{\boldsymbol{\varepsilon}}^T \right\} \bar{\mathbf{A}} \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1}$$

Temos que calcular a esperança de $E \left\{ \bar{\boldsymbol{\varepsilon}} \bar{\boldsymbol{\varepsilon}}^T \right\}$ que nada mais é que a esperança de uma forma quadrática (veja Apêndice Matriz de Covariância dos Erros) que

$$E \left\{ \bar{\boldsymbol{\varepsilon}} \bar{\boldsymbol{\varepsilon}}^T \right\} = \text{cov}(\bar{\boldsymbol{\varepsilon}}) + E(\bar{\boldsymbol{\varepsilon}}) E(\bar{\boldsymbol{\varepsilon}}^T)$$

Pela premissa 2 os erros tem média nula o que implica $E[\bar{\boldsymbol{\varepsilon}}] = \bar{\mathbf{0}}$ temos que

$$E \left\{ \bar{\boldsymbol{\varepsilon}} \bar{\boldsymbol{\varepsilon}}^T \right\} = \text{cov}(\bar{\boldsymbol{\varepsilon}})$$

substituindo temos

$$\text{COV} \left\{ \hat{\bar{\mathbf{p}}} \right\} = \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \text{COV} \left\{ \bar{\boldsymbol{\varepsilon}} \right\} \bar{\mathbf{A}} \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1}$$

Usando a premissa 3 que os erros tem variância constante e a premissa 4 que os erros não são correlacionáveis temos que

$$\text{COV} \{ \bar{\epsilon} \} = \sigma^2 \bar{\mathbf{I}}$$

$$\text{COV} \{ \hat{\bar{\mathbf{p}}} \} = \sigma^2 \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1} \bar{\mathbf{A}}^T \bar{\mathbf{A}} \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1}$$

$$\text{COV} \{ \hat{\bar{\mathbf{p}}} \} = \sigma^2 \left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1}$$

Veja pela equação acima que a covariância do estimador MQ $\text{cov} \{ \hat{\bar{\mathbf{p}}} \}$ é função de dois termos:

1) σ^2 que é a variância dos erros contidos nas observações geofísicas (incerteza dos dados). Se este termo for muito grande (dados muito imprecisos) conseqüentemente a variância das estimativas dos parâmetros [elementos da diagonal da matriz de covariância de $\hat{\bar{\mathbf{p}}} (\text{COV} \{ \hat{\bar{\mathbf{p}}} \})$] será também muito grande, significando instabilidade da solução estimada $\hat{\bar{\mathbf{p}}}$.

2) $\left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1}$ que é um termo que independe dos parâmetros e sim das variáveis independentes e conseqüentemente do problema direto formulado. Este termo chamaremos de termo de propagação dos erros uma vez que ele pode funcionar como um amplificador do ruído σ^2 significando instabilidade da solução estimada $\hat{\bar{\mathbf{p}}}$.

Quando esta amplificação ocorre?

Há duas ferramentas para avaliarmos quando o termo de propagação dos erros que para o estimador MQ é $\left(\bar{\mathbf{A}}^T \bar{\mathbf{A}} \right)^{-1}$ funciona como um amplificador do ruído σ^2 . A primeira ferramenta, que usaremos agora, é a função determinante da matriz $\bar{\mathbf{A}}^T \bar{\mathbf{A}}$ e a segunda ferramenta são os autovalores da

matriz $\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}}$ (veremos mais adiante). Vamos relembra a definição da inversa de uma matriz $\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}}$

$$\left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}} \right)^{-1} = \frac{1}{\det \left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}} \right)} \text{adj} \left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}} \right)$$

Como a covariância do estimador MQ é

$$\text{COV} \left\{ \hat{\mathbf{p}} \right\} = \sigma^2 \left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}} \right)^{-1}$$

então

$$\text{COV} \left\{ \hat{\mathbf{p}} \right\} = \sigma^2 \frac{1}{\det \left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}} \right)} \text{adj} \left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}} \right)$$

Vemos pela equação acima que a amplificação do ruído σ^2 ocorrerá quando o determinante da matriz $\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}}$ for um valor muito pequeno. Estamos aqui usando o determinante da matriz $\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}}$ como uma ferramenta para testarmos a estabilidade da solução de MQ. Na prática se a matriz $\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}}$ tem um determinante próximo de zero as variâncias de \hat{p}_j , $j = 1, \dots, M$ serão valores muito grandes isto porque ocorrerá uma amplificação do ruído contido nos dados (valores das variâncias σ^2 do ruído). Se as variâncias de \hat{p}_j são grandes então a solução do problema é instável.

Aqui estamos usando o determinante para caracterização de um problema mal-posto, mais adiante usaremos a decomposição em valores singulares para fazermos a mesma avaliação.

Apêndice – Teorema de Taylor

Suponha que $f(\bar{\mathbf{x}})$ é uma função contínua com primeira e segunda derivadas contínuas. Então dado um ponto $\bar{\mathbf{x}}^o$, podemos escrever $f(\bar{\mathbf{x}})$ como:

$$f(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}}^o) + \nabla f(\bar{\mathbf{x}})^T (\bar{\mathbf{x}} - \bar{\mathbf{x}}^o) + \frac{1}{2} (\bar{\mathbf{x}} - \bar{\mathbf{x}}^o)^T \nabla^2 f(\bar{\mathbf{c}}) (\bar{\mathbf{x}} - \bar{\mathbf{x}}^o).$$

para um ponto $\bar{\mathbf{c}}$ entre $\bar{\mathbf{x}}$ e $\bar{\mathbf{x}}^o$

Se $\bar{\mathbf{x}} - \bar{\mathbf{x}}^o$ é um valor pequeno, então podemos aproximar $f(\bar{\mathbf{x}})$ como

$$f(\bar{\mathbf{x}}) \approx f(\bar{\mathbf{x}}^o) + \nabla f(\bar{\mathbf{x}})^T (\bar{\mathbf{x}} - \bar{\mathbf{x}}^o) + \frac{1}{2} (\bar{\mathbf{x}} - \bar{\mathbf{x}}^o)^T \nabla^2 f(\bar{\mathbf{x}}^o) (\bar{\mathbf{x}} - \bar{\mathbf{x}}^o).$$

O teorema de Taylor pode ser facilmente adaptada para funções de uma única variável i.e.,

$$f(x) \approx f(x^o) + f'(x)^T (x - x^o) + \frac{1}{2} (x - x^o)^T f''(x^o) (x - x^o).$$

Apêndice - Matriz de Covariância dos Erros

Temos que calcular a esperança de $E \{ \bar{\mathbf{\epsilon}} \bar{\mathbf{\epsilon}}^T \}$ que nada mais é que a esperança de uma forma quadrática. Para tanto vamos partir da matriz de covariância dos erros:

$$\text{cov}(\bar{\mathbf{\epsilon}}) = E \left\{ \left[\bar{\mathbf{\epsilon}} - E(\bar{\mathbf{\epsilon}}) \right] \left[\bar{\mathbf{\epsilon}} - E(\bar{\mathbf{\epsilon}}) \right]^T \right\}$$

$$\text{cov}(\bar{\mathbf{\epsilon}}) = E \left\{ \bar{\mathbf{\epsilon}} \bar{\mathbf{\epsilon}}^T - 2 \bar{\mathbf{\epsilon}} E(\bar{\mathbf{\epsilon}})^T + E(\bar{\mathbf{\epsilon}}) E(\bar{\mathbf{\epsilon}})^T \right\}$$

$$\text{cov}(\bar{\mathbf{\epsilon}}) = E \left\{ \bar{\mathbf{\epsilon}} \bar{\mathbf{\epsilon}}^T \right\} - 2 E \left\{ \bar{\mathbf{\epsilon}} E(\bar{\mathbf{\epsilon}})^T \right\} + E \left\{ E(\bar{\mathbf{\epsilon}}) E(\bar{\mathbf{\epsilon}})^T \right\}$$

$$\text{cov}(\bar{\mathbf{\epsilon}}) = E \left\{ \bar{\mathbf{\epsilon}} \bar{\mathbf{\epsilon}}^T \right\} - 2 E(\bar{\mathbf{\epsilon}}) E(\bar{\mathbf{\epsilon}})^T + E(\bar{\mathbf{\epsilon}}) E(\bar{\mathbf{\epsilon}})^T$$

$$\text{cov}(\bar{\mathbf{\epsilon}}) = E \left\{ \bar{\mathbf{\epsilon}} \bar{\mathbf{\epsilon}}^T \right\} - E(\bar{\mathbf{\epsilon}}) E(\bar{\mathbf{\epsilon}})^T$$

logo

$$E \left\{ \bar{\mathbf{\epsilon}} \bar{\mathbf{\epsilon}}^T \right\} = \text{cov}(\bar{\mathbf{\epsilon}}) + E(\bar{\mathbf{\epsilon}}) E(\bar{\mathbf{\epsilon}})^T$$