DEDUÇÃO DE ALGUNS ESTIMADORES VIA MÉTODO DOS MULTIPLICADORES DE LAGRANGE

Apresentamos no tópico anterior que o problema vinculado de minimizar uma função $\phi\ (\overline{\mathbf{p}})$ de M variáveis sujeito a N vínculos $\mathbf{f}\ (\overline{\mathbf{p}}) = \overline{\mathbf{0}}$, representado por um sistema de N equações nas M variáveis é um problema de otimização vinculada que pode ser resolvido através do método dos mutiplicadores de Lagrange.

Neste tópico iremos deduzir os seguintes estimadores via método dos multiplicadores de Lagrange:

1) Estimador dos Mínimos quadrados Subdeterminado
2) Estimador Ridge Regression (Regularizador de Tikhonov de ordem zero)
3) Estimador Suavidade (Regularizador de Tikhonov de primeira ordem)

1) Solução do problema de Mínimos quadrados Subdeterminado via método dos multiplicadores de Lagrange:

1.1) Dedução do estimador mínimos quadrados subdeterminado

Já dissemos que o problema genérico de Mínimos quadrados subdeterminado é formulado como:

$$\begin{cases} \min \ \overline{\mathbf{p}}^{\mathrm{T}}\ \overline{\mathbf{p}} \\ \text{sujeito a}: \ \overline{\overline{\mathbf{A}}}\ \overline{\mathbf{p}} = \overline{\mathbf{y}}^{\mathbf{o}} \end{cases}, \qquad (1.1)$$

Este problema será resolvido via Método dos Multiplicadores de Lagrange que consiste em minimizar a função não vinculada

$$\Gamma = \overline{\mathbf{p}}^{\mathrm{T}}\ \overline{\mathbf{p}} + \left(\overline{\mathbf{y}}^{\mathbf{o}} - \overline{\overline{\mathbf{A}}}\ \overline{\mathbf{p}}\right)^{\mathrm{T}}\ \overline{\boldsymbol{\lambda}},$$

Condição de mínimo da função não vinculada $\Gamma(\overline{\mathbf{p}})$ é

$$\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = \overline{0}$$

$$2\,\hat{\bar{\mathbf{p}}} - \bar{\bar{\mathbf{A}}}^{\mathsf{T}}\,\bar{\boldsymbol{\lambda}} = \bar{\mathbf{0}}$$

$$\hat{\bar{\mathbf{p}}} = \frac{1}{2}\,\bar{\bar{\mathbf{A}}}^{\mathsf{T}}\,\bar{\boldsymbol{\lambda}}$$

Substituindo $\hat{\bar{\mathbf{p}}}$ em $\bar{\bar{\mathbf{A}}}\bar{\mathbf{p}} = \bar{\mathbf{y}}^{\mathbf{o}}$, temos

$$\frac{1}{2}\,\bar{\bar{\mathbf{A}}}\;\bar{\bar{\mathbf{A}}}^{\mathsf{T}}\,\bar{\boldsymbol{\lambda}} = \bar{\mathbf{y}}^{o}$$

$$\bar{\boldsymbol{\lambda}} = 2\left(\bar{\bar{\mathbf{A}}}\;\bar{\bar{\mathbf{A}}}^{\mathsf{T}}\right)^{-1}\bar{\mathbf{y}}^{o}$$

$$\hat{\bar{\mathbf{p}}} = \frac{1}{2}\bar{\bar{\mathbf{A}}}^{\mathsf{T}}\bar{\boldsymbol{\lambda}}$$

Estimado o vetor N-dimensional ($\bar{\boldsymbol{\lambda}}$) chamado de multiplicador de Lagrange, então vamos substituir no vetor de parâmetros originais $\hat{\bar{\mathbf{p}}} = \frac{1}{2}\bar{\bar{\mathbf{A}}}^{\mathsf{T}}\,\bar{\boldsymbol{\lambda}}$ que resulta no estimador dos MQ subdeterminado

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

Em resumo a minimização da função $\Gamma(\bar{\mathbf{p}})$ leva à solução $\hat{\bar{\mathbf{p}}}$ que chamaremos de **ESTIMADOR DE MÍNIMOS QUADRADOS SUBDETERMINADO (r = N $\leq$ M)**

$$\boxed{\hat{\bar{\mathbf{p}}} = \bar{\bar{\mathbf{A}}}^{\mathbf{T}}\left(\bar{\bar{\mathbf{A}}}\;\bar{\bar{\mathbf{A}}}^{\mathsf{T}}\right)^{-1}\bar{\mathbf{y}}^{\mathbf{o}} \qquad (1.2)}$$

2) Solução do problema Regularizador de Tikhonov de ordem zero (estimador Ridge Regression) via método dos multiplicadores de Lagrange:

2.1) Dedução do estimador Ridge Regression (Regularizador de Tikhonov de ordem zero)

O estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) é formulado como:

$$\begin{cases} \min \ \|\bar{\mathbf{p}}\|_2^2 \\ \text{sujeito a}: \ \left\|\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}}\right\|_2^2 = \delta. \end{cases} \tag{2.1}$$

em que δ é o erro médio quadrático (EMQ) da seqüência de realizações da variável aleatória que contamina as observações. Dada uma seqüência $\varepsilon_1,\ldots,\varepsilon_N$, o erro médio quadrático é dado por

$$\delta \approx \sqrt{\frac{\sum_{i=1}^{N} \varepsilon_i^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N} (y_i^c - y_i^o)^2}{N}}$$

em que y_i^c é o ajuste produzido na i-ésima observação usando-se um dado estimador. Sob certas premissas estatísticas, o EMQ é uma estimativa do desvio padrão da variável aleatória que contamina as observações.

O problema vinculado (2.1) pode ainda ser expresso como

$$\begin{cases} \min \ \bar{\mathbf{p}}^{\mathrm{T}}\,\bar{\mathbf{p}} \\ \text{sujeito a}: \ \left(\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}}\right)^T \left(\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}}\right) = \delta\,. \end{cases}$$

Este problema de otimização vinculada pode ser resolvido através do método dos mutiplicadores de Lagrange.que consiste em minimizar a função não vinculada

$$\Gamma = \bar{\mathbf{p}}^{\mathrm{T}}\,\bar{\mathbf{p}} + \frac{1}{k}\left[\left(\bar{\bar{\mathbf{A}}}\bar{\mathbf{p}} - \bar{\mathbf{y}}^0\right)^{\mathrm{T}}\left(\bar{\bar{\mathbf{A}}}\bar{\mathbf{p}} - \bar{\mathbf{y}}^0\right) - \delta\right]$$

Condição de mínimo da função não vinculada $\Gamma(\bar{\mathbf{p}})$ é

$$\overline{\nabla}_{\bar{\mathbf{p}}}(\Gamma) = \bar{0}$$

O vetor gradiente em relação ao vetor de parâmetros da função não vinculada é

$$\overline{\nabla}_{\bar{\mathbf{p}}}(\Gamma) = 2\bar{\mathbf{p}} + \frac{1}{k}\left[2\bar{\bar{\mathbf{A}}}^T\left(\bar{\bar{\mathbf{A}}}\bar{\mathbf{p}} - \bar{\mathbf{y}}^0\right)\right]$$

Da condição de mínimo temos a dedução do estimador de Ridge Regression ($\bar{\mathbf{p}}^*$), i.e.,

$$2\bar{\mathbf{p}}^* + \frac{1}{k}\left[2\bar{\bar{\mathbf{A}}}^T\left(\bar{\bar{\mathbf{A}}}\bar{\mathbf{p}}^* - \bar{\mathbf{y}}^0\right)\right] = \bar{\mathbf{0}}$$

$$2\bar{\mathbf{p}}^* + \frac{1}{k}2\bar{\bar{\mathbf{A}}}^T\bar{\bar{\mathbf{A}}}\bar{\mathbf{p}}^* - \frac{1}{k}2\bar{\bar{\mathbf{A}}}^T\bar{\mathbf{y}}^0 = \bar{\mathbf{0}}$$

Multiplicando os dois lados da equação acima por $\frac{k}{2}$ temos:

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

$$\left(\bar{\bar{\mathbf{A}}}^T\bar{\bar{\mathbf{A}}} + k\,\bar{\bar{\mathbf{I}}}\right)\bar{\mathbf{p}}^* = \bar{\bar{\mathbf{A}}}^T\,\bar{\mathbf{y}}^0$$

Pré-multiplicando os dois lados da equação acima por $\left(\bar{\bar{\mathbf{A}}}^T\bar{\bar{\mathbf{A}}} + k\,\bar{\bar{\mathbf{I}}}\right)^{-1}$ temos a expressão do **ESTIMADOR RIDGE REGRESSION (Regularizador de Tikhonov de ordem zero)**

$$\bar{\mathbf{p}}^* = \left(\bar{\bar{\mathbf{A}}}^T\,\bar{\bar{\mathbf{A}}} + k\,\bar{\bar{\mathbf{I}}}\right)^{-1}\bar{\bar{\mathbf{A}}}^T\,\bar{\mathbf{y}}^0 \qquad (2.2)$$

2.2) Funcional a ser minimizado tanto via Ridge Regression como via MQ sub

Já dissemos que o problema genérico de Mínimos quadrados subdeterminado é formulado como:

$$\begin{cases} \min \ \bar{\mathbf{p}}^{T} \ \bar{\mathbf{p}} \\ \text{sujeito a}: \ \bar{\bar{\mathbf{A}}} \ \bar{\mathbf{p}} = \bar{\mathbf{y}}^{\mathbf{o}} \end{cases},$$

Por outro lado, vimos que estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) é formulado como:

$$\begin{cases} \min \ \bar{\mathbf{p}}^{T} \bar{\mathbf{p}} \\ \text{sujeito a}: \ \left\| \bar{\bar{\mathbf{A}}} \ \bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2^2 = \delta. \end{cases}$$

Note que o funcional a ser minimizado nestes dois problemas são EXATAMENTE o mesmo. Isto é, tanto no MQ Subdeterminado quanto no Ridge Regression estamos minimizando o funcional

$$\phi \ (\bar{\mathbf{p}}) = \bar{\mathbf{p}}^{T} \ \bar{\mathbf{p}} = p_1^2 + p_2^2 + \ldots + p_M^2$$

Em um problema envolvendo dois parâmetros (p_1 e p_2) a função $\phi(\bar{\mathbf{p}})$ é representada por curvas de isovalores (Figura 4) com um mínimo muito bem definido no ponto $p_1 = p_2 = 0$ (ponto vermelho na Figura 4), como mostramos no exercício 3.

Portanto, matematicamente, o vinculo de mínima norma Euclideana do vetor de parâmetros introduzido tanto no estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) quanto no estimador MQ subdeterminado força, no sentido dos mínimos quadrados, com que todos os elementos do vetor de parâmetros estejam o mais próximo possível do valor zero. Isto é, estamos forçando com que $p_1 \approx p_2 \approx \ldots \approx p_M \approx 0$. Em termos geofísicos este tipo de vínculo não faz sentido, no entanto, ele passa fazer sentido porque estamos minimizando o funcional $\phi \ (\bar{\mathbf{p}}) = \left\| \bar{\mathbf{p}} \right\|_2^2$, mas sujeito a explicar os dados geofísicos dentro de uma precisão do ruído, i.e., sujeito a $\left\| \bar{\bar{\mathbf{A}}} \ \bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2^2 = \delta$

O funcional a ser minimizado tanto via Ridge Regression como via MQ sub - Vínculo de Mínima norma Euclideana

$$\|\bar{\mathbf{p}}\|_2^2 = \bar{\mathbf{p}}^T\,\bar{\mathbf{p}} = \mathrm{p}_1^2 + \mathrm{p}_2^2$$

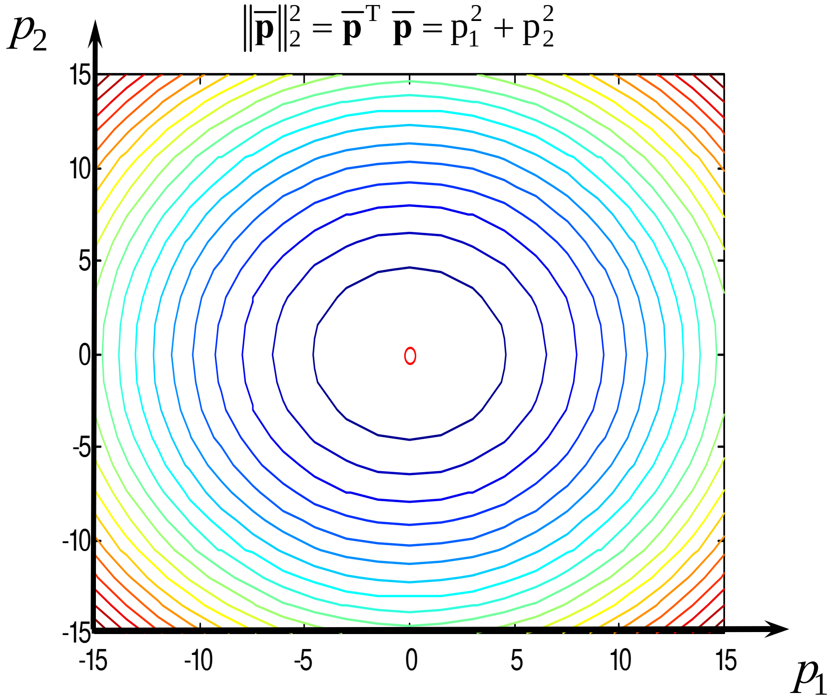


Figura 4

3) Solução do problema com Vínculos de Igualdade Aproximada via método dos multiplicadores de Lagrange:

Vimos que o estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) introduz o vínculo da norma mínima do vetor dos parâmetros. Para tanto este vinculo é introduzido ao problema através da minimização do funcional $\phi(\overline{\mathbf{p}}) = \|\overline{\mathbf{p}}\|_2^2 = \overline{\mathbf{p}}^T \ \overline{\mathbf{p}} = p_1^2 + p_2^2 + ... + p_M^2$, que matematicamente é equivalente impor uma proximidade entre o vetor de parâmetros estimados e o vetor nulo ($\mathbf{0}$).

Vamos nesta seção introduzir matematicamente dois outros vínculos: 1) Vínculos Absolutos de Igualdade e 2) Vínculos Relativos de Igualdade. Em seguida apresentaremos a formulação matemática de um problema utilizando estes dois vínculos.

Para tanto, vamos supor que o nosso problema geofísico deva ser resolvido sujeito aos seguintes vínculos:

3.1) Vínculos Absolutos de Igualdade

Vamos supor que, de algum modo, sejam conhecidos os valores exatos de H elementos do vetor de parâmetros. Por exemplo, vamos supor que as 20 células mostradas na Figura 5 sejam suficientes para descrever a variação de uma certa propriedade física do interior da Terra. Considere que um determinado furo de sondagem forneça informações sobre a propriedade física da Terra tal como mostra na Figura 5. Veja que o furo de sondagem atravessa os parâmetros p_3, p_8 e p_{13}; portanto temos informações precisas e conhecidas sobre a propriedade física destes três parâmetros (H = 3) que são respectivamente os valores numéricos $p_3^{borehole}, p_8^{borehole}$ e $p_{13}^{borehole}$. O vinculo de Absoluto de igualdade permite estabelecer uma proximidade entre estes três parâmetros (p_3, p_8 e p_{13}) e os seus correspondentes valores numéricos conhecidos a priori ($p_3^{borehole}, p_8^{borehole}$ e $p_{13}^{borehole}$). Neste caso específico apresentado na Figura 5 o vinculo de Absoluto de igualdade força, no sentido dos mínimos quadrados, com que $p_3 \approx p_3^{borehole}$, $p_8 \approx p_8^{borehole}$ e $p_{13} \approx p_{13}^{borehole}$

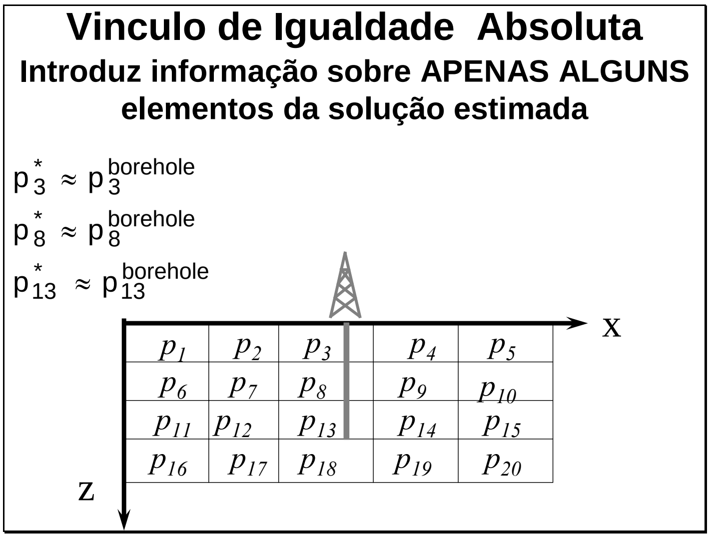


Figura 5

Formulação Matemática dos Vínculos Absolutos de Igualdade:

Considere o conhecimento a priori dos valores exatos de H elementos do vetor de parâmetros. Vamos formar uma matriz $\overline{\overline{\mathbf{D}}}$ (H x M) de vínculos absolutos cujas as H linhas desta matriz (H $\leq$M) força H parâmetros estarem próximos a H valores conhecidos a priori e que estão armazenados no vetor $\overline{\mathbf{d}}^{\,o}$ (H x 1), de modo que estabelecendo uma igualdade estrita temos:

$$\overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}} = \overline{\mathbf{d}}^{\,o} \tag{3.1}$$

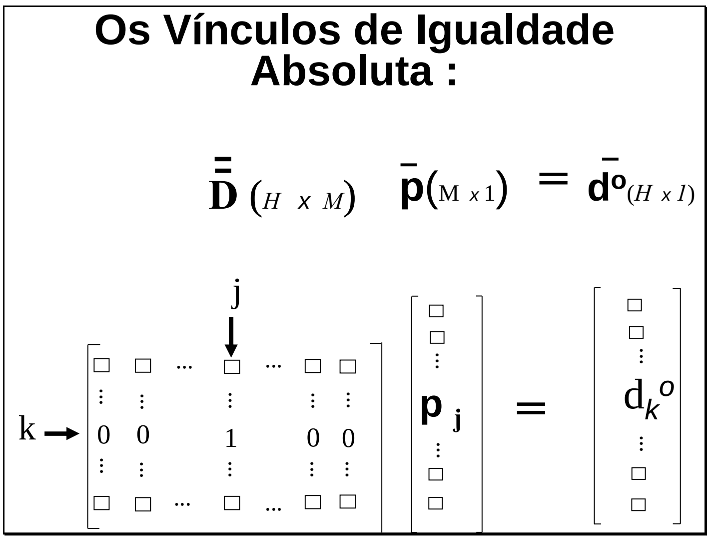


Figura 6

Note que com esta estrutura podemos introduzir qualquer informação quantitativa sobre um determinado parâmetro. Por exemplo, se houver a informação a priori que o i-ésimo parâmetro deve estar o mais próximo possível do valor 7. Então a k-ésima linha da matriz $\overline{\overline{\mathbf{D}}}$ (Figura 6) é um vetor $\overline{\mathbf{d}}_k^T$ de dimensão M com todos os elementos nulos exceto o i-ésimo elemento que é atribuído o valor 1 e o k-ésimo elemento do vetor $\overline{\mathbf{d}}^o$ será atribuído o valor numérico 7 (i.e., $d_k^o = 7$). Veja que o produto escalar

$$\overline{\mathbf{d}}_k^T \, \overline{\mathbf{p}} = d_k^o$$

ou seja,

$$\begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_{i-1} \\ p_i \\ p_{i+1} \\ \vdots \\ p_m \end{bmatrix} = 7$$

leva a relação de igualdade absoluta estrita

$$p_i = 7 \quad .$$

3.2) <u>Vínculos Relativos de Igualdade</u>

L vinculos relativos de Igualdade estabelecem L relações lineares entre pares de parâmetros. Antes de apresentarmos uma formulação matemática, vamos apresentar um possível exemplo geológico de aplicação do vínculo de igualdade relativa em um mapeamento geológico usando dados aeromagnéticos. Vamos supor que objetivamos estimar a variação de susceptibilidade magnética de uma área usando dados aeromagnéticos. Para tanto discretizamos a área a ser mapeada em 12 células 3D mostradas na Figura 7. Presumimos que esta discretização seja suficiente para descrever a variação da susceptibilidade magnética da área e que permite a produção de um mapa geológico baseado na variação desta propriedade física nesta região em estudo. Considere a existência de dois afloramentos na área em estudo. No canto esquerdo da área, coincidindo com o parâmetro p_1, há um afloramento de granito cuja a susceptibilidade magnética é 0.25 A/m e no canto esquerdo, coincidindo com o parâmetro p_{12}, há um afloramento de gabro cuja a susceptibilidade magnética é 1.0 A/m . O vinculo Relativo de igualdade permite estabelecer uma relação linear entre estes dois parâmetros (p_1 e p_{12}) que neste caso específico apresentado na Figura 7 o vinculo de Relativo de igualdade força, no sentido dos mínimos quadrados, com que $p_{12} \approx 4p_1$.

Isto significa que esperamos estimar uma susceptibilidade magnética para o parâmetro p_{12} (que corresponde ao gabro aflorante) que seja 4 vezes maior que a estimativa da susceptibilidade magnética do parâmetro p_1 (que corresponde ao granito aflorante).

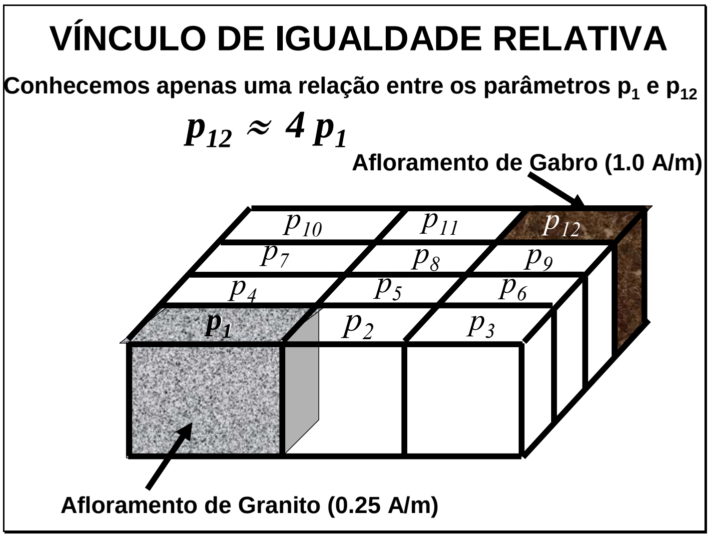


Figura 7

Formulação Matemática dos Vínculos Relativos de Igualdade:

Matematicamente este vínculo permite introduzir complexas relações lineares entre quaisquer pares de parâmetros. Para tanto vamos formular uma matriz $\overline{\overline{\mathbf{B}}}$ (L x M) cujas L linhas permitam estabelecer L relações lineares entre pares de parâmetros de forma que estabelecendo uma igualdade estrita temos:

$$\overline{\overline{\mathbf{B}}}\;\overline{\mathbf{p}} = \overline{\mathbf{b}}^{\,o} \tag{3.2}$$

Com esta igualdade podemos introduzir relações lineares entre pares de parâmetros sem, no entanto, conhecer o valor exato dos parâmetros. Por exemplo, se houver a informação a priori envolvendo uma relação linear entre o i-ésimo parâmetro é o j-ésimo parâmetro do tipo:

$$f_k \, p_i + t_k \, p_j = b_k^o$$

Neste caso a k-ésima linha da matriz $\overline{\overline{\mathbf{B}}}$ (Figura 8) é um vetor $\overline{\mathbf{b}}_k^T$ de dimensão M, com todos os elementos nulos exceto o i-ésimo elemento que é atribuído o valor f_k e o j-ésimo elemento que é atribuído o valor t_k e o k-ésimo elemento do vetor $\overline{\mathbf{b}}^o$ será atribuído o valor numérico b_o^k .

Os Vínculos de Igualdade Relativa:

$$\overline{\overline{\mathbf{B}}} \; \overline{\mathbf{p}} = \overline{\mathbf{b}}^{\,o}$$

$$\begin{array}{c c} & \begin{array}{ccccccccccc} & & & i & & & j & & & & \end{array} \\ k \rightarrow & \begin{bmatrix} & & & & & & & & & & \\ 0 & \ldots & 0 & f_k & 0 & \ldots & 0 & t_k & 0 & \ldots & 0 \\ & & & & & & & & & & \end{bmatrix}_{(L \times M)} \end{array} \begin{bmatrix} p_1 \\ \vdots \\ p_{i-1} \\ p_i \\ p_{i+1} \\ \vdots \\ p_{j-1} \\ p_j \\ p_{j+1} \\ \vdots \\ p_M \end{bmatrix}_{(M \times 1)} = \begin{bmatrix} b_1^o \\ \\ b_{k-1}^o \\ b_k^o \\ b_{k+1}^o \\ \\ b_L^o \end{bmatrix}_{(L \times 1)}$$

Figura 8

Veja que, neste caso, o produto escalar

$$\overline{\mathbf{b}}_k^T \; \overline{\mathbf{p}} = b_k^o$$

ou seja

$$\begin{bmatrix} 0 & \cdots & 0 & f_k & 0 & \cdots & 0 & t_k & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_{i-1} \\ p_i \\ p_{i+1} \\ \vdots \\ p_{j-1} \\ p_j \\ p_{j+1} \\ \vdots \\ p_M \end{bmatrix} = \begin{bmatrix} b_1^o \\ \\ b_{k-1}^o \\ b_k^o \\ b_{k+1}^o \\ \\ b_L^o \end{bmatrix}$$

leva a relação de igualdade relativa estrita

$$f_k \, p_i + t_k \, p_j = b_k^o$$

No caso específico do exemplo mostrado na Figura 7, a matriz $\bar{\bar{\mathbf{B}}}$ (1 x 2) sendo que a única linha desta matriz é um vetor $\bar{\mathbf{b}}_k^T$ de dimensão 12, com todos os elementos nulos, exceto o i-ésimo e j-ésimo elementos que serão atribuídos, respectivamente, os valores 4 e 1 e o único elemento do vetor $\bar{\mathbf{b}}^o$ será atribuído o valor numérico 0 de modo que o produto escalar $\bar{\mathbf{b}}_k^T \; \bar{\mathbf{p}} \;=\; b_k^o$

$$\begin{bmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \\ p_{10} \\ p_{11} \\ p_{12} \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

leva a relação linear

$$4p_1 - p_{12} = 0$$

ou seja

$$p_{12} = 4p_1$$

3.3) <u>Formulação do Problema utilizando os Vínculos Relativos e Absolutos de Igualdade via método dos multiplicadores de Lagrange:</u>

Vamos introduzir os vínculos de igualdade absoluta (equação 3.1) e de igualdade relativa (equação 3.2) de modo aproximado, i.e, estes vínculos não são exatamente satisfeitos mas, <u>aproximadamente</u> satisfeitos. Então esperamos que

$$\bar{\bar{\mathbf{D}}} \quad \bar{\mathbf{p}} \;\approx\; \bar{\mathbf{d}}^{\,o}$$

e

$$\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} \approx \overline{\mathbf{b}}^{o}$$

Para tanto formularemos que os vínculos de igualdade absoluta e relativa sejam aproximadamente satisfeitos no sentido dos mínimos quadrados. Desta forma, os vínculos aproximados de igualdade absoluta e relativa serão respectivamente definidos pelos funcionais $\phi_a(\overline{\mathbf{p}})$ e $\phi_r(\overline{\mathbf{p}})$ definidos no problema vinculado de:

$$\text{Minimizar} \quad \begin{cases} \phi_r(\overline{\mathbf{p}}) = \left\| \overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} - \overline{\mathbf{b}}^{o} \right\|_2^2 \\ \phi_a(\overline{\mathbf{p}}) = \left\| \overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}} - \overline{\mathbf{d}}^{o} \right\|_2^2 \end{cases}$$

$$\text{sujeito a}: \ \left\| \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^{\mathbf{o}} \right\|_2^2 = \delta. \qquad (3.3)$$

em que δ é o erro médio quadrático (EMQ) da seqüência de realizações da variável aleatória que contamina as observações. Dada uma seqüência $\varepsilon_1, \ldots, \varepsilon_N$, o erro médio quadrático é dado por

$$\delta \approx \sqrt{\frac{\sum_{i=1}^{N} \varepsilon_i^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N} (y_i^c - y_i^o)^2}{N}}$$

em que y_i^c é o ajuste produzido na i-ésima observação usando-se um dado estimador. Sob certas premissas estatísticas, o EMQ é uma estimativa do desvio padrão da variável aleatória que contamina as observações.

Este problema de otimização vinculada pode ser resolvido através do método dos mutiplicadores de Lagrange.que consiste em minimizar a função não vinculada

$$\Gamma(\overline{\mathbf{p}}) = \lambda_r \phi^{r}(\overline{\mathbf{p}}) + \lambda_a \phi^{a}(\overline{\mathbf{p}}) + \left[\left(\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^{0} \right)^{\mathrm{T}} \left(\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^{0} \right) - \delta \right]$$

Condição de mínimo da função não vinculada $\Gamma(\overline{\mathbf{p}})$ é

$$\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = \overline{0}$$

O vetor gradiente em relação ao vetor de parâmetros da função não vinculada é

$$\Gamma(\overline{\mathbf{p}}) = \lambda_r \left(\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} - \overline{\mathbf{b}}^o\right)^T \left(\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} - \overline{\mathbf{b}}^o\right) + \lambda_a \left(\overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}} - \overline{\mathbf{d}}^o\right)^T \left(\overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}} - \overline{\mathbf{d}}^o\right) + \left[\left(\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^0\right)^{\mathrm{T}} \left(\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^0\right) - \delta\right]$$

Recordando que operador gradiente $\overline{\nabla}_{\overline{\mathbf{p}}} \in R^M$ aplicado forma quadrática é

$$\underset{(M\times 1)}{\overline{\nabla}_{\overline{\mathbf{p}}}} \left\{ \underset{(1\times M)}{\overline{\mathbf{x}}^T} \underset{(\mathrm{M}\times 1)}{\overline{\mathbf{x}}} \right\} = 2\, \underset{(M\times 1)}{\overline{\nabla}_{\overline{\mathbf{p}}}} \left\{ \underset{(1\times M)}{\overline{\mathbf{x}}^T} \right\} \underset{(\mathrm{M}\times 1)}{\overline{\mathbf{x}}}$$

$$\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = \lambda_r \left[2\overline{\nabla}_{\overline{\mathbf{p}}}\left(\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} - \overline{\mathbf{b}}^o\right)^T \left(\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} - \overline{\mathbf{b}}^o\right)\right] + \lambda_a \left[2\overline{\nabla}_{\overline{\mathbf{p}}}\left(\overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}} - \overline{\mathbf{d}}^o\right)^T \left(\overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}} - \overline{\mathbf{d}}^o\right)\right] + \left[2\overline{\nabla}_{\overline{\mathbf{p}}}\left(\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^0\right)^{\mathrm{T}} \left(\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^0\right)\right]$$

$$\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = \lambda_r \left[2\overline{\nabla}_{\overline{\mathbf{p}}}\left(\overline{\mathbf{p}}^T \overline{\overline{\mathbf{B}}}^{\mathrm{T}} - \overline{\mathbf{b}}^{o^T}\right) \left(\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} - \overline{\mathbf{b}}^o\right)\right] + \lambda_a \left[2\overline{\nabla}_{\overline{\mathbf{p}}}\left(\overline{\mathbf{p}}^T \overline{\overline{\mathbf{D}}}^T - \overline{\mathbf{d}}^{o^T}\right) \left(\overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}} - \overline{\mathbf{d}}^o\right)\right] + \left[2\overline{\nabla}_{\overline{\mathbf{p}}}\left(\overline{\mathbf{p}}^T \overline{\overline{\mathbf{A}}}^T - \overline{\mathbf{y}}^{0^T}\right) \left(\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^0\right)\right]$$

$$\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = 2\lambda_r\left[\left(\overline{\nabla}_{\overline{\mathbf{p}}}\{\overline{\mathbf{p}}^T\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\} - \overline{\nabla}_{\overline{\mathbf{p}}}\{\overline{\mathbf{b}}^{o^T}\}\right)\left(\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} - \overline{\mathbf{b}}^o\right)\right] +$$
$$2\lambda_a\left[\left(\overline{\nabla}_{\overline{\mathbf{p}}}\{\overline{\mathbf{p}}^T\overline{\overline{\mathbf{D}}}^T\} - \overline{\nabla}_{\overline{\mathbf{p}}}\{\overline{\mathbf{d}}^{o^T}\}\right)\left(\overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}} - \overline{\mathbf{d}}^o\right)\right] +$$
$$2\left[\left(\overline{\nabla}_{\overline{\mathbf{p}}}\{\overline{\mathbf{p}}^T\overline{\overline{\mathbf{A}}}^T\} - \overline{\nabla}_{\overline{\mathbf{p}}}\{\overline{\mathbf{y}}^{0^T}\}\right)\left(\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^0\right)\right]$$

$$\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = 2\lambda_r\left[\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\left(\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} - \overline{\mathbf{b}}^o\right)\right] + 2\lambda_a\left[\overline{\overline{\mathbf{D}}}^T\left(\overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}} - \overline{\mathbf{d}}^o\right)\right] +$$
$$2\left[\overline{\overline{\mathbf{A}}}^T\left(\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^0\right)\right]$$

$$\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = 2\lambda_r\left[\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\,\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} - \overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\mathbf{b}}^o\right] + 2\lambda_a\left[\overline{\overline{\mathbf{D}}}^T\overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}} - \overline{\overline{\mathbf{D}}}^T\overline{\mathbf{d}}^o\right] +$$
$$2\left[\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\overline{\mathbf{A}}}^T\overline{\mathbf{y}}^0\right]$$

Condição de mínimo da função não vinculada $\Gamma(\overline{\mathbf{p}})$ é $\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = \overline{0}$ temos

$$2\lambda_r\left[\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\,\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}}^* - \overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\mathbf{b}}^o\right] + 2\lambda_a\left[\overline{\overline{\mathbf{D}}}^T\overline{\overline{\mathbf{D}}}\,\overline{\mathbf{p}}^* - \overline{\overline{\mathbf{D}}}^T\overline{\mathbf{d}}^o\right] +$$
$$2\left[\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}}^* - \overline{\overline{\mathbf{A}}}^T\overline{\mathbf{y}}^0\right] = \overline{\mathbf{0}},$$

Dividindo os dois lados da equação por 2 temos

$$\lambda_r\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\,\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}}^* - \lambda_r\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\mathbf{b}}^o + \lambda_a\overline{\overline{\mathbf{D}}}^T\overline{\overline{\mathbf{D}}}\ \overline{\mathbf{p}}^* - \lambda_a\overline{\overline{\mathbf{D}}}^T\overline{\mathbf{d}}^o + \overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{A}}}\overline{\mathbf{p}}^* - \overline{\overline{\mathbf{A}}}^T\overline{\mathbf{y}}^0 = \overline{\mathbf{0}}$$
$$\left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{A}}} + \lambda_r\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\,\overline{\overline{\mathbf{B}}} + \lambda_a\overline{\overline{\mathbf{D}}}^T\overline{\overline{\mathbf{D}}}\right)\overline{\mathbf{p}}^* = \lambda_r\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\mathbf{b}}^o + \lambda_a\overline{\overline{\mathbf{D}}}^T\overline{\mathbf{d}}^o + \overline{\overline{\mathbf{A}}}^T\overline{\mathbf{y}}^0$$

Multiplicando os dois lados da equação acima por $\left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{A}}} + \lambda_r\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\,\overline{\overline{\mathbf{B}}} + \lambda_a\overline{\overline{\mathbf{D}}}^T\overline{\overline{\mathbf{D}}}\right)^{-1}$ temos

$$\bar{\mathbf{p}}^{*} = \left[\bar{\bar{\mathbf{A}}}^{\mathrm{T}} \bar{\bar{\mathbf{A}}} + \lambda_r \bar{\bar{\mathbf{B}}}^{\mathrm{T}} \bar{\bar{\mathbf{B}}} + \lambda_a \bar{\bar{\mathbf{D}}}^{\mathrm{T}} \bar{\bar{\mathbf{D}}} \right]^{-1} \left(\bar{\bar{\mathbf{A}}}^{\mathrm{T}} \bar{\mathbf{y}}^{o} + \lambda_r \bar{\bar{\mathbf{B}}}^{\mathrm{T}} \bar{\mathbf{b}}^{o} + \lambda_a \bar{\bar{\mathbf{D}}}^{\mathrm{T}} \bar{\mathbf{d}}^{o} \right) \quad (3.4)$$

Então $\bar{\mathbf{p}}^{*}$ (equação 3.4) é o estimador do problema vinculado formulado pela equação (3.3).

3.4) Unicidade e Estabilidade da solução estimada via o estimador $\bar{\mathbf{p}}^{*}$

Note que não há garantia da solução estimada via $\bar{\mathbf{p}}^{*}$ (equação 3.4) ser única e estável, uma vez que não foi imposto que as matrizes $\bar{\bar{\mathbf{B}}}^{\mathrm{T}} \bar{\bar{\mathbf{B}}}$ e $\bar{\bar{\mathbf{D}}}^{\mathrm{T}} \bar{\bar{\mathbf{D}}}$ sejam definidas positivas.

No entanto, mesmo que a matriz $\bar{\bar{\mathbf{A}}}^{\mathrm{T}} \bar{\bar{\mathbf{A}}}$ seja semi-definida positiva (com autovalor igual a zero) ou seja definida positiva (com todos os autovalores maiores que a zero) mas apresente autovalores pequenos e as matrizes $\bar{\bar{\mathbf{B}}}^{\mathrm{T}} \bar{\bar{\mathbf{B}}}$ e $\bar{\bar{\mathbf{D}}}^{\mathrm{T}} \bar{\bar{\mathbf{D}}}$ sejam singulares, a solução estimada via $\bar{\mathbf{p}}^{*}$ será única e estável se a matriz quadrada de ordem M a ser invertida na equação (3.4), $\left[\bar{\bar{\mathbf{A}}}^{\mathrm{T}} \bar{\bar{\mathbf{A}}} + \lambda_r \bar{\bar{\mathbf{B}}}^{\mathrm{T}} \bar{\bar{\mathbf{B}}} + \lambda_a \bar{\bar{\mathbf{D}}}^{\mathrm{T}} \bar{\bar{\mathbf{D}}} \right]$, apresentar os M autovalores não nulos e grandes. Isto ocorre em duas situações:

(1) Se a matriz $\bar{\bar{\mathrm{F}}} = \left[\lambda_r \bar{\bar{\mathbf{B}}}^{\mathrm{T}} \bar{\bar{\mathbf{B}}} + \lambda_a \bar{\bar{\mathbf{D}}}^{\mathrm{T}} \bar{\bar{\mathbf{D}}} \right]$ apresentar os M autovalores não nulos.

(2) Se as regiões no espaço de parâmetros associadas ao mal-condicionamento de $\bar{\bar{\mathrm{F}}}$ e $\bar{\bar{\mathbf{A}}}^{\mathrm{T}} \bar{\bar{\mathbf{A}}}$ tenham apenas um ponto em comum.

3.5) Casos Particulares da solução estimada via o estimador $\bar{\mathbf{p}}^*$ i.e. casos particulares do uso dos Vínculos Relativos e Absolutos de Igualdade

Matematicamente formulamos um problema empregando os vínculos aproximados de igualdade absoluta e relativa através do problema vinculado (3.3), i.e.,

$$\text{Minimizar} \quad \begin{cases} \phi_r(\bar{\mathbf{p}}) = \left\| \bar{\bar{\mathbf{B}}}\,\bar{\mathbf{p}} - \bar{\mathbf{b}}^o \right\|_2^2 \\ \phi_a(\bar{\mathbf{p}}) = \left\| \bar{\bar{\mathbf{D}}}\,\bar{\mathbf{p}} - \bar{\mathbf{d}}^o \right\|_2^2 \end{cases}$$

$$\text{sujeito a}: \left\| \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2^2 = \delta.$$

Vimos que o minimante deste problema é o estimador expresso na equação (3.4), i.e.

$$\bar{\mathbf{p}}^* = \left[\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{A}}} + \lambda_r \bar{\bar{\mathbf{B}}}^{\mathrm{T}}\bar{\bar{\mathbf{B}}} + \lambda_a \bar{\bar{\mathbf{D}}}^{\mathrm{T}}\bar{\bar{\mathbf{D}}} \right]^{-1} \left(\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\mathbf{y}}^o + \lambda_r \bar{\bar{\mathbf{B}}}^{\mathrm{T}}\bar{\mathbf{b}}^o + \lambda_a \bar{\bar{\mathbf{D}}}^{\mathrm{T}}\bar{\mathbf{d}}^o \right)$$

Vamos agora estudar dois caso particulares:

Caso Particular número 1: Estimador Ridge Regression ou Regularizador de Tikhonov de ordem zero

Fazendo-se no estimador $\bar{\mathbf{p}}^*$

$\bar{\bar{\mathbf{B}}} = \bar{\bar{\mathbf{0}}}$

$\bar{\bar{\mathbf{D}}} = \bar{\bar{\mathbf{I}}}_M$

$\bar{\mathbf{d}}^o = \bar{\mathbf{0}}$

$\lambda_a = k$

Note que, sob estas condições recaímos no problema de

$$\text{Minimizar} \quad \left\| \bar{\mathbf{p}} \right\|_2^2 = \bar{\mathbf{p}}^T \bar{\mathbf{p}}$$

$$\text{sujeito a}: \left\| \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2^2 = \delta.$$

em que o minimante é

$$\overline{\mathbf{p}}^{*} = \left[\overline{\overline{\mathbf{A}}}^{\mathrm{T}} \overline{\overline{\mathbf{A}}} + k \overline{\overline{\mathbf{I}}} \right]^{-1} \left(\overline{\overline{\mathbf{A}}}^{\mathrm{T}} \overline{\mathbf{y}}^{o} \right) \qquad (3.5)$$

Note então que neste caso particular o estimador dado pela equação (3.5) é o Ridge Regression (ou Regularizador de Tikhonov de ordem zero) [equação 2.2]

Vale ressaltar que o estimador expresso na equação (3.4), i.e.

$$\overline{\mathbf{p}}^{*} = \left[\overline{\overline{\mathbf{A}}}^{\mathrm{T}} \overline{\overline{\mathbf{A}}} + \lambda_r \overline{\overline{\mathbf{B}}}^{\mathrm{T}} \overline{\overline{\mathbf{B}}} + \lambda_a \overline{\overline{\mathbf{D}}}^{\mathrm{T}} \overline{\overline{\mathbf{D}}} \right]^{-1} \left(\overline{\overline{\mathbf{A}}}^{\mathrm{T}} \overline{\mathbf{y}}^{o} + \lambda_r \overline{\overline{\mathbf{B}}}^{\mathrm{T}} \overline{\mathbf{b}}^{o} + \lambda_a \overline{\overline{\mathbf{D}}}^{\mathrm{T}} \overline{\mathbf{d}}^{o} \right)$$

é mais genérico do que o estimador Ridge Regression (ou Regularizador de Tikhonov de ordem zero)

$$\overline{\mathbf{p}}^{*} = \left[\overline{\overline{\mathbf{A}}}^{\mathrm{T}} \overline{\overline{\mathbf{A}}} + k \overline{\overline{\mathbf{I}}} \right]^{-1} \left(\overline{\overline{\mathbf{A}}}^{\mathrm{T}} \overline{\mathbf{y}}^{o} \right)$$

no sentido de permitir que apenas um subconjunto de H elementos do conjunto de M parâmetros sejam vinculados a um valor absoluto qualquer (valores que compõe o vetor $\overline{\mathbf{d}}^{o}$). No caso do Ridge Regression, TODOS OS M PARÂMETROS estão vinculados a estarem o mais próximo possível de zero, isto no sentido dos mínimos quadrados. Esta forte vinculação do Ridge Regression é muitas vezes restritiva e pouco realista do ponto de vista físico-geológico, porém há garantia da estabilidade da solução estimada basta usar um multiplicador de Lagrange suficientemente grande.

Caso Particular número 2: Estimador Suavidade ou Regularizador de Tikhonov de Primeira ordem

Vimos na seção (3.1) que os vínculos de igualdade absoluta permitem introduzir informações a priori exata sobre um conjunto de parâmetros. Vimos que um caso particular do vínculo de igualdade absoluta é o Ridge Regression (ou Regularizador de Tikhonov de ordem zero) [equação 2.2] em que impomos que todos os parâmetros devam estar o mais próximo possível do valor zero.

Vimos na seção (3.2) que os vínculos de igualdade relativa permitem introduzir informações a priori sobre relações lineares entre pares de parâmetros sem, no entanto, introduzir informações exatas (valores absolutos) dos parâmetros. De fato, matematicamente este vínculo permite introduzir complexas relações lineares entre quaisquer pares de parâmetros, como mostramos na Figura 8.

Note que o vínculo de igualdade relativa possibilita introduzir informações a priori menos quantitativas e mais qualitativas. Estas informações são mais "fracas" no sentido de não demandarem informações quantitativas e exatas sobre os parâmetros.

Por exemplo, vamos supor que a única informação que temos sobre os parâmetros a serem estimados é, simplesmente, a informação que há uma **variação suave** (não descontínua) sobre os parâmetros. Vamos supor que em um determinado modelo interpretativo o parâmetro p_i esteja adjacente ao parâmetro p_{i+1} e p_{i-1}, como mostra a Figura 9. Neste exemplo de modelo discreto de parâmetros, estabelecer uma variação suave entre os parâmetros, significa estabelecer uma igualdade aproximada entre os pares de parâmetros adjacentes. Isto significa impor que

$$\cdots \hat{p}_{i-1} \approx \hat{p}_i \approx \hat{p}_{i+1} \cdots$$

ou ainda

$$\begin{aligned} \cdots \quad \hat{p}_{i-1} - \hat{p}_i &\approx 0 \\ \hat{p}_i - \hat{p}_{i+1} &\approx 0 \quad \cdots \end{aligned} \tag{3.6}$$

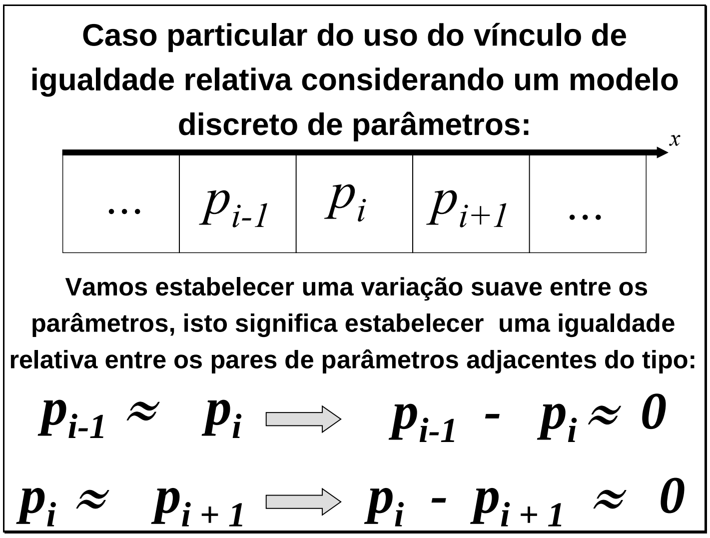


Figura 9

Note que estas informações a priori são relações lineares entre pares de parâmetros. Já vimos que, matematicamente, relações lineares entre pares de parâmetros podem ser introduzidas via vínculo de igualdade relativa [equação (3.2)]. Portanto, estas relações lineares [equações (3.6)] podem ser expressas como um caso particular da equação (3.2), i.e., um caso particular do vínculo de igualdade relativa.

$$\overline{\overline{\mathbf{B}}}\ \overline{\mathbf{p}} = \overline{\mathbf{0}} \tag{3.7}$$

em que a matriz $\overline{\overline{\mathbf{B}}}$ (L x M) cujas L linhas permitem estabelecer L relações lineares entre L pares de parâmetros que são adjacentes. Portanto L é o número de vínculos de igualdade relativa expressos pelas equações (3.6) para o modelo interpretativo da Figura 9. Neste caso particular que impomos uma variação suave considerando o modelo discreto de parâmetros da Figura 9 temos que a k-ésima linha da matriz $\overline{\overline{\mathbf{B}}}$ é um vetor $\overline{\mathbf{b}}_k^T$ de dimensão M, com todos os elementos nulos exceto o i-ésimo elemento que é atribuído o valor 1 e o (i+1)-ésimo elemento que é atribuído o valor -1 (Figura 10). Veja que, neste caso, o produto escalar da k-ésima linha da matriz $\overline{\overline{\mathbf{B}}}$ é o vetor de parâmetros leva a relação linear

$$p_i - p_{i+1} = 0$$

ou seja

$$p_i = p_{i+1}.$$

Caso particular do uso do vínculo de igualdade relativa, que chamamos de vínculo de suavidade, considerando um modelo discreto de parâmetros da Figura anterior:

$$\overline{\overline{\mathbf{B}}}\quad \overline{\mathbf{p}} \quad = \quad \overline{\mathbf{0}}$$

$$\begin{array}{c} \\ \mathrm{k\text{-}1} \rightarrow \\ \mathrm{k} \rightarrow \\ \\ \end{array}
\begin{array}{c} \qquad\quad \mathrm{i\text{-}1}\ \ \mathrm{i}\ \ \mathrm{i{+}1} \\ \qquad\quad \downarrow\ \downarrow\ \downarrow \\ \left[\begin{array}{ccccccccc} 0 & \ldots & 0 & 1 & -1 & 0 & 0\ldots & & 0 \\ 0 & \ldots & 0 & 0 & 1 & -1 & 0\ldots & & 0 \\ & & & & & & & & \end{array}\right]_{(L \times M)} \end{array}
\begin{bmatrix} p_1 \\ \vdots \\ p_{i-1} \\ p_i \\ p_{i+1} \\ \vdots \\ p_{j-1} \\ p_j \\ p_{j+1} \\ \vdots \\ p_M \end{bmatrix}_{(M \times 1)}
= \begin{bmatrix} 0 \\ \vdots \\ \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(L \times 1)}$$

Figura 10

Veja que a (k-1)-ésima linha da matriz $\overline{\overline{\mathbf{B}}}$ é um vetor $\overline{\mathbf{b}}_k^T$ de dimensão M, com todos os elementos nulos exceto o (i-1)-ésimo elemento que é atribuído o valor 1 e o (i)-ésimo elemento que é atribuído o valor -1 (Figura 10). Veja que, neste caso, o produto escalar da (k-1)-ésima linha da matriz $\overline{\overline{\mathbf{B}}}$ é o vetor de parâmetros leva a relação linear

$$p_{i-1} - p_i = 0$$

ou seja

$$p_{i-1} = p_i \, .$$

No exemplo mostrado na Figura 11 temos um modelo interpretativo composto por 4 parâmetros adjacentes em relação a direção x . Para este exemplo de modelo discreto de 4 parâmetros, estabelecer uma <u>variação suave</u> entre os parâmetros, significa estabelecer entre os pares de parâmetros adjacentes relações lineares to tipo

$$\begin{aligned} p_1 - p_2 &= 0 \\ p_2 - p_3 &= 0 \\ p_3 - p_4 &= 0 \end{aligned}$$

Neste caso da Figura 11 temos 3 relações lineares entre os 4 pares de parâmetros que são adjacentes e compõe o modelo interpretativo da Figura 11. Portanto, as relações lineares acima podem ser expressas como um caso particular do vínculo de igualdade relativa [equação (3.7)]. De acordo com o modelo interpretativo da Figura 11 a matriz $\overline{\overline{\mathbf{B}}}$ (3 x 4) é dada por

$$\overline{\overline{\mathbf{B}}} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

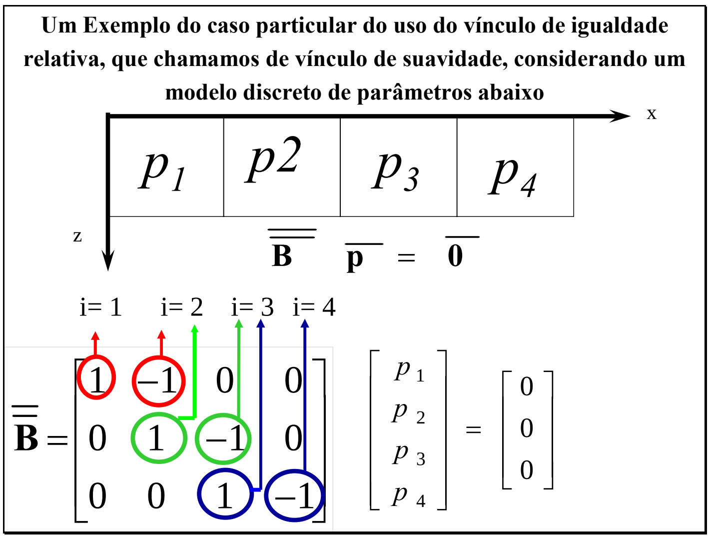


Figura 11

Note que o produto $\bar{\bar{\mathbf{B}}}\,\bar{\mathbf{p}}$ representa a diferença entre parâmetros fisicamente adjacentes. Portanto, a menos de uma constante, o produto $\bar{\bar{\mathbf{B}}}\,\bar{\mathbf{p}}$ quantifica uma aproximação da primeira derivada[1] da função contínua dos parâmetros. A matriz $\bar{\bar{\mathbf{B}}}$ representa, portanto, o operador discreto de primeiras derivadas (Twomey, 1977, Constable et al., 1987; Barbosa et al., 1997) cujas linhas contêm apenas dois elementos não nulos: 1 e -1 que estão associados com cada par de parâmetros adjacentes. No exemplo mostrado na Figura 11, o produto $\bar{\bar{\mathbf{B}}}\,\bar{\mathbf{p}}$ é uma aproximação da primeira derivada da distribuição dos parâmetros em relação a x. No entanto, dependendo do problema produto $\bar{\bar{\mathbf{B}}}\,\bar{\mathbf{p}}$ pode representar uma aproximação da primeira derivada da distribuição dos parâmetros em relação a: i) x e y; ii) x e z; iii) x, y e z. Na Figura 12 mostramos um modelo interpretativo composto por 12 parâmetros adjacentes em relação as direções x e z. Neste caso a matriz $\bar{\bar{\mathbf{B}}}$ representa o operador discreto de primeiras derivadas em relação as direções x e z.

$$\bar{\bar{\mathbf{B}}} = \begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1
\end{bmatrix}$$

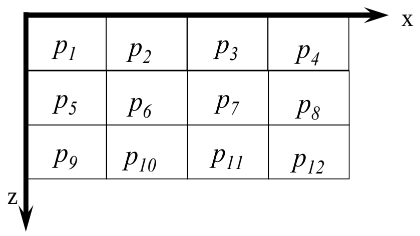
---

[1] Derivada de uma função $f(x)$ em relação a x, considerando x_1 como um número particular no domínio de $f(x)$ é dada como:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x_1 + \Delta x) - f(x_1)}{\Delta x}$$

As primeiras 9 linhas da matriz $\overline{\overline{\mathbf{B}}}$ representam o operador discreto de primeiras derivadas em relação a direção x. Já as linhas 10 a 17 da matriz $\overline{\overline{\mathbf{B}}}$ representam o operador discreto de primeiras derivadas em relação a direção z.

Um Exemplo do caso particular do uso do vínculo de igualdade relativa, que chamamos de vínculo de suavidade, considerando um modelo discreto de 12 parâmetros distribuidos ao longo das direções x e z.



Para este exemplo de modelo discreto de 12 parâmetros, estabelecer uma <u>variação suave</u> entre os parâmetros, significa estabelecer 17 relações lineares entre os pares de parâmetros adjacentes nas duas direções x e z, veja abaixo.

• Relações lineares na direção x

$p_1 - p_2 = 0 \quad p_5 - p_6 = 0 \quad p_9 - p_{10} = 0$

$p_2 - p_3 = 0 \quad p_6 - p_7 = 0 \quad p_{10} - p_{11} = 0$

$p_3 - p_4 = 0 \quad p_9 - p_8 = 0 \quad p_{11} - p_{12} = 0$

• Relações lineares na direção z

$p_1 - p_5 = 0 \quad p_5 - p_9 = 0$

$p_2 - p_6 = 0 \quad p_6 - p_{10} = 0$

$p_3 - p_7 = 0 \quad p_7 - p_{11} = 0$

$p_4 - p_8 = 0 \quad p_8 - p_{12} = 0$

Figura 12

O uso do vínculo de primeira derivada dos parâmetros é conhecido na literatura como Estimador Suavidade (smoothness) ou Estimador Flatness ou Regularizador de Tikhonov de Primeira ordem. Tal como o estimador Ridge Regression (ou Regularizador de Tikhonov de ordem zero), o Estimador Suavidade ou Regularizador de Tikhonov de Primeira ordem, é um caso particular do estimador $\overline{\mathbf{p}}^*$ (equação 3.4) i.e., é um caso particular do problema vinculado formulado pela equação (3.3). Para tanto fazemos no estimador $\overline{\mathbf{p}}^*$ [equação (3.4)]

$\overline{\overline{\mathbf{D}}} = \overline{\overline{\mathbf{0}}}$,

$\overline{\mathbf{b}}^{o} = \overline{\mathbf{0}}$,

$\lambda_r = \lambda$,

e

$\overline{\overline{\mathbf{B}}}$ (Figura 8) uma matriz cujas linhas contém apenas dois elementos não nulos iguais a 1 e –1, localizados nas colunas correspondentes aos parâmetros *i* e *j* cujas estimativas devem estar o mais próximas possível (como mostramos no exemplo da Figura 12).

$$\text{Minimizar} \qquad \left\| \overline{\overline{\mathbf{B}}}\, \overline{\mathbf{p}} \right\|_2^2 = \overline{\mathbf{p}}^T\, \overline{\overline{\mathbf{B}}}^T\, \overline{\overline{\mathbf{B}}}\, \overline{\mathbf{p}}$$

$$\text{sujeito} \quad \text{a}: \quad \left\| \overline{\overline{\mathbf{A}}}\, \overline{\mathbf{p}} - \overline{\mathbf{y}}^{\mathbf{o}} \right\|_2^2 = \delta\,.$$

em que o minimante é

$$\overline{\mathbf{p}}^* = \left[\overline{\overline{\mathbf{A}}}^{\mathrm{T}} \overline{\overline{\mathbf{A}}} + \lambda\, \overline{\overline{\mathbf{B}}}^{\mathrm{T}} \overline{\overline{\mathbf{B}}} \right]^{-1} \left(\overline{\overline{\mathbf{A}}}^{\mathrm{T}} \overline{\mathbf{y}}^{o} \right) \tag{3.8}$$

Este estimador (3.8) é conhecido como **Regularizador de Tikhonov de Primeira ordem**

Neste caso o estimador (3.8) é a solução que impõe, no sentido dos mínimos quadrados, o vínculo da primeira derivada do vetor de parâmetros sujeito a explicar os dados geofísicos observados dentro da precisão imposta pelo ruído dos dados (δ). Este estimador (3.8) recebe na literatura vários nomes dentre eles:

- Estimador de Suavidade (Smoothness)
- Estimador de Suavidade Global (Global Smoothness)
- Estimador de Igualdade ou Achatamento (Flatness)
- Estimador com vínculo do mínimo da primeira derivada do vetor de parâmetros
- Regularizador de Tikhonov de Primeira ordem.

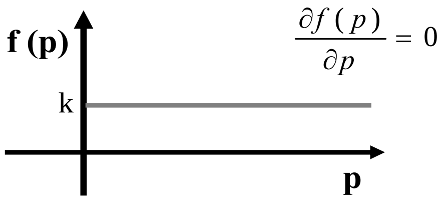
Todos estes nomes têm sentido tanto fisco como matemático. Quando minimizamos $\left\| \overline{\overline{\mathbf{B}}}\, \overline{\mathbf{p}} \right\|_2^2$ estamos exigindo que, no sentido dos mínimos quadrados, a primeira derivada do vetor de parâmetros seja próxima de zero, i.e., $\overline{\overline{\mathbf{B}}}\,\overline{\mathbf{p}} \approx \overline{\mathbf{0}}$. Veja, se considerarmos os parâmetros como uma função contínua $f(p)$, então estabelecer que a primeira derivada $f\,'(p)$ é igual a zero[2] estamos estabelecendo indiretamente que

[2] Rigorosamente, impomos que $f\,'(p)$ deva estar o mais próximo possível de zero, no sentido dos mínimos quadrados.

esta função dos parâmetros é igual a uma constante, i.e., $f(p) = k$ [veja Figura 13]. Em outras palavras, estamos impondo que a distribuição dos parâmetros (solução estimada) seja achatada ou suave. Estamos impondo que haja uma igualdade entre pares de parâmetros geometricamente adjacentes.

O estimador $\bar{\mathbf{p}}^* = \left[\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{A}}} + \lambda\, \bar{\bar{\mathbf{B}}}^{\mathrm{T}}\bar{\bar{\mathbf{B}}}\right]^{-1}\left(\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\mathbf{y}}^{o}\right)$ recebe na literatura vários nomes dentre eles:

Estimador de Suavidade (Smoothness)
Estimador de Suavidade Global (Global Smoothness)
Estimador de Igualdade ou Achatamento (Flatness)
Estimador com vínculo do mínimo da primeira derivada do vetor de parâmetros
Regularizador de Tikhonov de Primeira ordem.



O significado de se estabelecer que a primeira derivada dos parâmetros é igual a zero é dizendo que a função dos parâmetros é igual a uma constante k. Em outras palavras, significa impor que a distribuição dos parâmetros (solução) seja suave

Figura 13

Sobre a Unicidade e Estabilidade da solução estimada via o estimador Estimador Suavidade ou Regularizador de Tikhonov de Primeira ordem.

Vimos que o estimador da suavidade, i.e.,

$$\bar{\mathbf{p}}^* = \left[\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{A}}} + \lambda\, \bar{\bar{\mathbf{B}}}^{\mathrm{T}}\bar{\bar{\mathbf{B}}}\right]^{-1}\left(\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\mathbf{y}}^{o}\right)$$

envolve a inversa de uma matriz dada pela soma $\left[\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{A}}} + \lambda\, \bar{\bar{\mathbf{B}}}^{\mathrm{T}}\bar{\bar{\mathbf{B}}}\right]$. A matriz $\bar{\bar{\mathbf{B}}}^{\mathrm{T}}\bar{\bar{\mathbf{B}}}$ é uma matriz que representa um operador discreto da segunda derivada. Por exemplo, considerando o modelo interpretativo da Figura 11 a matriz $\bar{\bar{\mathbf{B}}}$ (3 x 4) é dada por

$$\bar{\bar{\mathbf{B}}} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

logo

$$\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\overline{\mathbf{B}}} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Esta matriz $\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\overline{\mathbf{B}}}$ tem posto 3, logo há um vetor pertencente ao espaço nulo desta matriz, i.e., há um vetor-solução que satisfaz o sistema homogêneo de equações $\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\overline{\mathbf{B}}}\,\overline{\mathbf{p}} = \overline{\mathbf{0}}$ é o vetor

$$\overline{\mathbf{p}}^{\text{null}} = \alpha \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Generalizando, o posto da matriz $\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\overline{\mathbf{B}}}$ é igual ao número de vínculos linearmente independentes, porque as colunas de $\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\overline{\mathbf{B}}}$ são combinações lineares de $\overline{\mathbf{b}}_k$. Além disso o posto de $\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\overline{\mathbf{B}}}$ é no máximo M-1, isto pode ser verificado pelo fato dos vínculos $p_i = p_j$ e $p_j = p_l$ implica que $p_i = p_l$. Logo o espaço nulo da matriz $\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\overline{\mathbf{B}}}$ é um vetor de dimensão M dado por

$$\overline{\mathbf{p}}^{\text{null}} \; (\in R^M) = \alpha \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}$$

Se temos dois parâmetros o funcional $\left\| \overline{\overline{\mathbf{B}}}\,\overline{\mathbf{p}} \right\|_2^2 = \overline{\mathbf{p}}^T \overline{\overline{\mathbf{B}}}^T \overline{\overline{\mathbf{B}}}\,\overline{\mathbf{p}}$ é representado pela Figura 14 em que

$$\overline{\overline{\mathbf{B}}} = [1 \quad -1]$$

$$\overline{\mathbf{p}}^T \overline{\overline{\mathbf{B}}}^T \overline{\overline{\mathbf{B}}}\,\overline{\mathbf{p}} = [p_1 \quad p_2] \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = [p_1 \quad p_2] \begin{pmatrix} p_1 - p_2 \\ p_2 - p_1 \end{pmatrix} = p_1^2 - 2\,p_1 p_2 + p_2^2$$

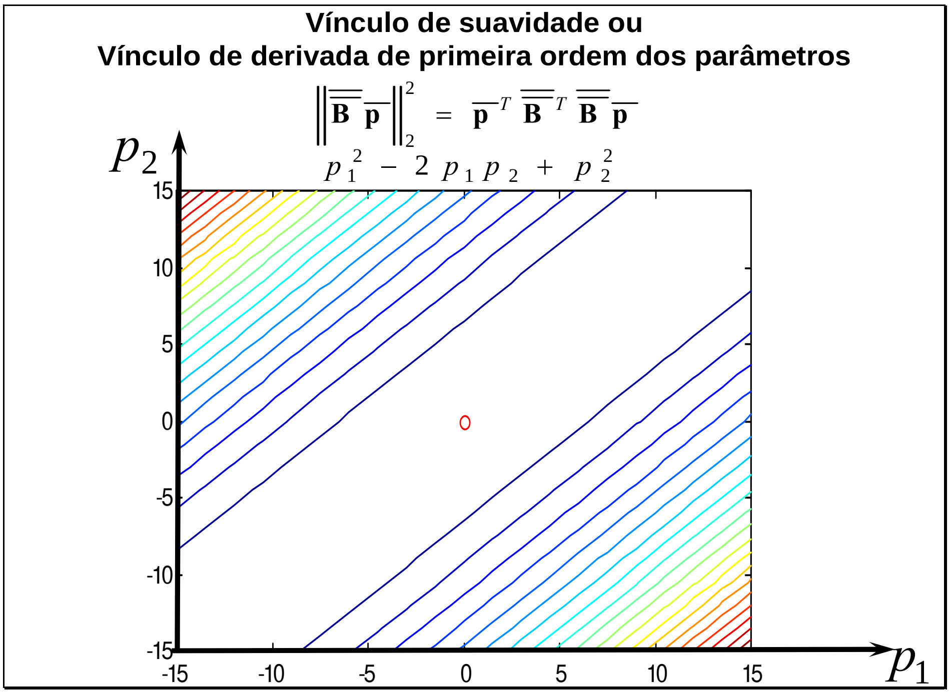


Figura 14

A Figura 14 mostra o funcional $\left\|\overline{\overline{\mathbf{B}}}\,\overline{\mathbf{p}}\right\|_2^2 = \overline{\mathbf{p}}^T\,\overline{\overline{\mathbf{B}}}^T\,\overline{\overline{\mathbf{B}}}\,\overline{\mathbf{p}}$ em que notamos claramente a região de mínimos, i.e. uma região em que este funcional é igual a zero. Portanto este funcional $\left\|\overline{\overline{\mathbf{B}}}\,\overline{\mathbf{p}}\right\|_2^2 = \overline{\mathbf{p}}^T\,\overline{\overline{\mathbf{B}}}^T\,\overline{\overline{\mathbf{B}}}\,\overline{\mathbf{p}}$ apresenta um mínimo mal definido. Apesar da matriz $\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\,\overline{\overline{\mathbf{B}}}$ ter um espaço nulo, este é um vetor proporcional a média dos parâmetros e na geofísica, em geral, o espaço nulo da matriz $\overline{\overline{\mathbf{A}}}^{\mathrm{T}}\overline{\overline{\mathbf{A}}}$ é um vetor proporcional a contraste dos parâmetros. Em outras palavra o espaço nulo da matriz $\overline{\overline{\mathbf{A}}}^{\mathrm{T}}\overline{\overline{\mathbf{A}}}$ é ortogonal ao espaço nulo da matriz $\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\,\overline{\overline{\mathbf{B}}}$. Portanto, a matriz dada pela soma $\left[\overline{\overline{\mathbf{A}}}^{\mathrm{T}}\overline{\overline{\mathbf{A}}} + \lambda\,\overline{\overline{\mathbf{B}}}^{\mathrm{T}}\overline{\overline{\mathbf{B}}}\right]$ não apresenta espaço nulo.

Seleção do parâmetro de regularização (selecionando uma boa solução)

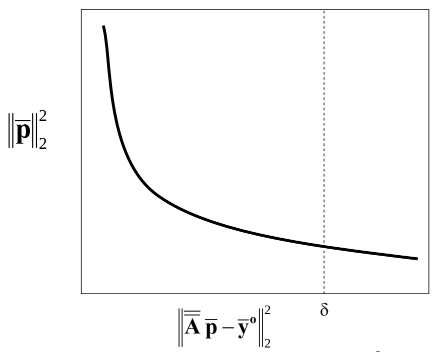
Para a maioria dos problemas lineares em geofísica resolvidos via estimador de mínimos quadrados resultam em infinitas soluções que ajustam os dados. Se considerarmos que os dados contem ruído não nenhuma explicação para ajustarmos perfeitamente tal ruido, portanto torna-se evidente que pode haver muitas soluções que podem adequadamente ajustar os dados no sentido que $\left\| \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2$ seja bastante pequeno. Na regularização de Tikhonov regularization, consideramos todas as soluções com $\left\| \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2 \leq \delta$ e selecionamos dentre essas soluções aquela que minimiza a norma de $\bar{\mathbf{p}}$, i.e.,

$$\text{Minimizar} \qquad \left\| \bar{\mathbf{p}} \right\|_2 = \bar{\mathbf{p}}^T \, \bar{\mathbf{p}}$$

$$\text{sujeito a :} \quad \left\| \bar{\bar{\mathbf{A}}}\;\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2 \leq \delta \,. \qquad (1)$$

Por que selecionamos a solução de mínima norma dos parâmetros dentre aquelas soluções que ajustam adequadamente os dados ? Uma explicação intuitiva é que toda a característica diferente de zero que aparecer na solução regularizada aumentará a norma de $\bar{\mathbf{p}}$. Tais características aparecem na solução porque são necessárias para ajustar os dados.

Note que como δ aumenta, o conjunto de possíveis soluções também aumenta e o valor mínimo de $\left\| \bar{\mathbf{p}} \right\|_2$ diminui. Nós podemos assim construir uma curva de valores mínimos de $\left\| \bar{\mathbf{p}} \right\|_2$ versus $\left\| \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2$,i.e., uma curva relacionada ao problema condicionado acima:



Caso particular em que o funcional ajustante $\left\| \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2^2 = \delta$

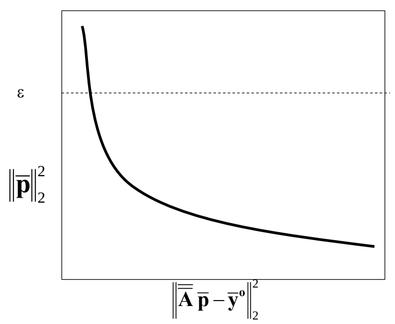
e sua associda norma mínima dos parâmetros $\left\| \bar{\mathbf{p}} \right\|_2^2$

Nós também podemos construir uma outra curva considerando o problema

$$\text{Minimizar} \quad \left\| \overline{\overline{\mathbf{A}}}\ \overline{\mathbf{p}} - \overline{\mathbf{y}}^{\mathbf{o}} \right\|_2$$

$$\text{sujeito a :} \quad \left\| \overline{\mathbf{p}} \right\|_2 \le \varepsilon \,. \tag{2}$$

Considerando que ε diminui, o conjunto de possíveis soluções também começa diminuir, por outro lado o mínimo de $\left\| \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^{\mathbf{o}} \right\|_2$ aumemta.



Caso particular em que $\left\|\overline{\mathbf{p}}\right\|_2^2 = \varepsilon$ e esta está associada a $\left\| \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^{\mathbf{o}} \right\|_2^2 = \delta$

Uma terceira opção considerara o problema dos minimos quadrados amortecido (Damped Least-Square) i.e.,

$$\min \left\{ \left\| \overline{\overline{\mathbf{A}}}\ \overline{\mathbf{p}} - \overline{\mathbf{y}}^{\mathbf{o}} \right\|_2 + \mu \left\| \overline{\mathbf{p}} \right\|_2 \right\} \tag{3}$$

Este problema (3) surge quando aplicamos o método dos multiplicadores de Lagrange ao problema (1) em que μ é o parâmetro de regularização. Pode ser mostrado que uma escolha apropriada de μ, δ e ε os três problemas acima (1), (2) e (3) resultam na mesma solução.

Métodos de escolha de μ, os quais se baseiam na tarefa de se encontrar $\mu(\varepsilon)$, tal que $\left\|\overline{\mathbf{p}}\right\|_2 = \varepsilon$, são denominados métodos “a priori” [problema (2)]. A premissa deste tipo de método, é a de que se tem uma boa informação prévia da solução a ser estimada. No entanto na maioria das vezes este dado não é disponível ou então não é suficientemente confiável ou preciso. Em geral, uma alternativa mais viável consiste em se empregar métodos, onde a escolha do parâmetro de regularização, μ, seja orientada pela busca de um $\mu(\delta)$, o qual resulte em $\left\| \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^{\mathbf{o}} \right\|_2 = \delta$. Este é o chamado Princípio da Discrepância de Morozov (1984). Parte-se do pressuposto de que o erro (ou *discrepância*) δ, para $\mu(\delta)$, é conseqüencia de ruído nos dados observados, e este é corretamente estimado. Sob certas premissas estatísticas para a contaminação por ruído dos dados experimentais, δ é o erro médio quadrático (EMQ) da seqüência de realizações da variável aleatória que contamina as observações. Logo, dada uma seqüência $\sigma_i, i = 1,\ldots,N$ o erro médio quadrático é dado por

$$\delta = \sqrt{\frac{\sum_{i=1}^{N} \sigma_i^2}{N}}$$

Sob certas premissas estatísticas, o EMQ é uma estimativa do desvio padrão da variável aleatória que contamina as observações.

Logo, tendo uma estimativa para o desvio padrão do ruído a busca da solução utilizando o critério de Morozov pode ser feita utilizando o Algoritmo

Início

passo 1: k = 0

passo 2: $\mu_k = 0$

passo 3: $\sigma =$ desvio padrão do ruído dos dados (valor presumido pelo interprete)

passo 4: compute $\tilde{\bar{\mathbf{p}}} = \min \left\{ \left\| \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2 + \mu_k \left\| \bar{\mathbf{p}} \right\|_2 \right\}$

passo 5: **Enquanto** $\left\| \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2 > \sqrt{\frac{\sum_{i=1}^{N} \sigma_i^2}{N}}$ **faça**

paso 6: k = k + 1

passo 7: $\mu_k = \mu_{k-1} \cdot 10^{1/10}$

passo 7: compute $\tilde{\bar{\mathbf{p}}} = \min \left\{ \left\| \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2 + \mu_k \left\| \bar{\mathbf{p}} \right\|_2 \right\}$

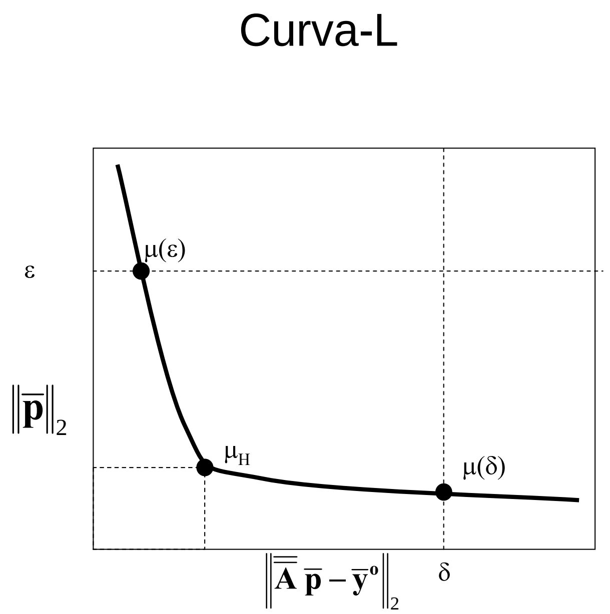
Fim Enquanto

Fim

Nós iremos nos concentrar na solução do problema (3) Mínimos quadrados amortecido.

Curva L

Vamos agora construir um gráfico na scala log-log da função $\|\bar{\mathbf{p}}\|_2$ versus $\|\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}}\|_2$. em função do parametro de regularização $\mu(\delta)$. Esta curva é chamada convenientemente de **curva-L**. Esta forma L ocorre porque $\|\bar{\mathbf{p}}\|_2$ é uma função estritamente decrescente com μ e $\|\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}}\|_2$ é uma função estritamente crescente com μ.

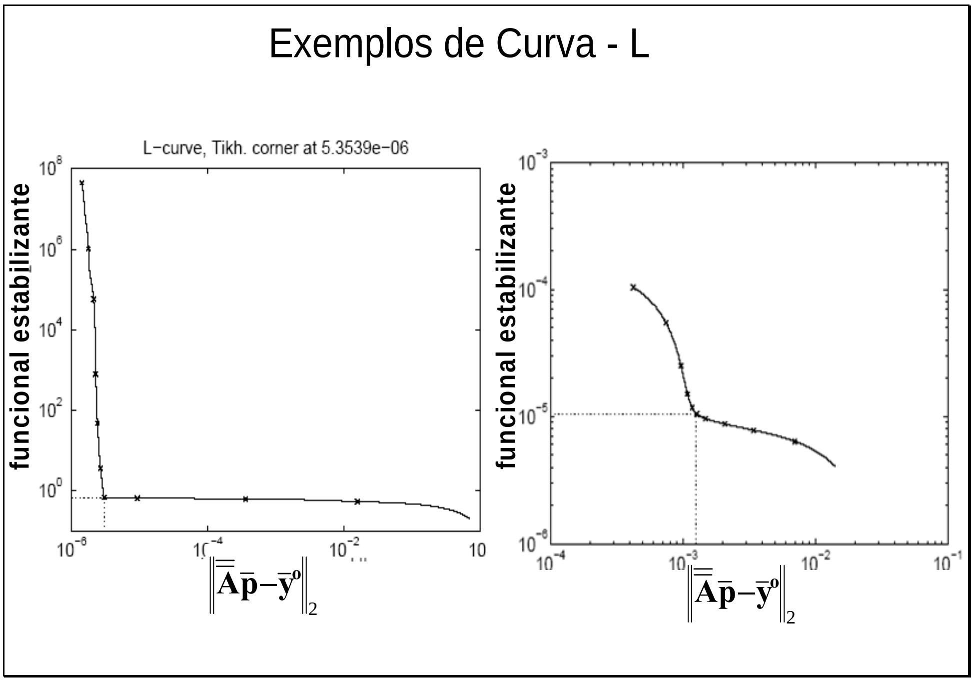


Curva L genérica em escala n log-log. O funcional estabilizante é o Regularizador de Tikhonov de ordem ZERO, $\|\bar{\mathbf{p}}\|_2$ *,versus a norma Euclideana dos resíduos,* $\|\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}}\|_2$*, ambos como funções do parâmetro de regularização* μ*.*

Note que em intervalos da curva-L mais próximos ao eixo vertical, uma variação pequena de μ, resulta numa acentuada variação de $\|\bar{\mathbf{p}}\|_2$. O mesmo se observa com relação a $\|\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}}\|_2$, para intervalos da curva-L mais próximos do eixo horizontal. Portanto, o passo logarítmico ameniza o efeito destas variações bruscas nestes intervalos críticos, tornando a busca de μ mais eficiente.

No gráfico, aparece um novo valor de parâmetro de regularização μ_H que corresponde um terceiro método de busca de μ, localizado no ponto da curva onde a sua curvatura é máxima. A escolha recai neste ponto, porque é onde se dá a relação mais equilibrada de suavidade(estabilidade) × ajuste, e onde a diferença no efeito da variação de é reduzido. Denominado de método de Hansen (1992), ou método da curva-L, não requer nem conhecimento prévio de ε nem qualquer informação sobre o erro (ou *discrepância*) δ.

A nitidez do "canto" varia de um problema para outro, mas é freqüentemente bem definido. Por esta razão, a curva é chamada uma curva L. Portanto além do princípio da discrepância (critério acima descrito) para escolher o valor de μ temos o critério da L-curva em que o valor de μ selecionado, μ_H, é aquele cuja a solução localiza-se o mais próximo ao canto da L-curva. Veja alguns exemplos reais de curva L.



V. A. Morozov. Methods for Solving Incorrectly Posed Problems. Springer– Verlag, New York, 1984.

Per Christian Hansen. Analysis of discrete ill-posed problems by means of the L–curve. SIAM Review, 34(4):561–580, December 1992.