DEDUÇÃO DO ESTIMADOR INVERSA GENERALIZADA VIA MÉTODO DOS MULTIPLICADORES DE LAGRANGE

Vimos que quando há M-r valores singulares nulos temos infinitas soluções uma vez que a solução que minimiza a soma dos quadrados dos resíduos é

$$\bar{\boldsymbol{\alpha}}_r{}' = \bar{\bar{\mathbf{S}}}_r^{\,\mathbf{-1}}\ \bar{\boldsymbol{\beta}}_r$$ [1]

(1)

Portanto a parte do vetor $\bar{\boldsymbol{\alpha}}$ correspondente a $\bar{\boldsymbol{\alpha}}_{M-r}$ não é usada para minimizar o funcional

$$Q = \left\| \bar{\boldsymbol{\beta}}_r - \bar{\bar{\mathbf{S}}}_r\ \bar{\boldsymbol{\alpha}}_r \right\|_2^2 + \left\| \bar{\boldsymbol{\beta}}_{M-r} \right\|_2^2 + \left\| \bar{\boldsymbol{\beta}}_{N-M} \right\|_2^2 \qquad (2)$$

Então, se temos M-r valores singulares nulos temos

$$\bar{\mathbf{p}}^* = \bar{\bar{\mathbf{V}}}\ \bar{\boldsymbol{\alpha}}' = \begin{bmatrix} \underset{(M\times r)}{\bar{\bar{\mathbf{V}}}_r} & \underset{(M\times M-r)}{\bar{\bar{\mathbf{V}}}_{M-r}} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\alpha}}'_r \\ \bar{\boldsymbol{\alpha}}'_{M-r} \end{bmatrix}$$

$$\bar{\mathbf{p}}^* = \bar{\bar{\mathbf{V}}}\ \bar{\boldsymbol{\alpha}}' = \begin{bmatrix} \underset{(M\times r)}{\bar{\bar{\mathbf{V}}}_r} & \underset{(M\times M-r)}{\bar{\bar{\mathbf{V}}}_{M-r}} \end{bmatrix} \begin{bmatrix} \bar{\bar{\mathbf{S}}}_r^{\,-1}\ \bar{\boldsymbol{\beta}}_r \\ \bar{\bar{\mathbf{S}}}_{M-r}^{\,-1}\ \bar{\boldsymbol{\beta}}_{M-r} \end{bmatrix}$$

$$\bar{\mathbf{p}}^* = \bar{\bar{\mathbf{V}}}_r\ \bar{\bar{\mathbf{S}}}_r^{\,-1}\ \bar{\boldsymbol{\beta}}_r + \bar{\bar{\mathbf{V}}}_{M-r}\ \bar{\bar{\mathbf{S}}}_{M-r}^{\,-1}\ \bar{\boldsymbol{\beta}}_{M-r}$$

[1] Veja a dedução do estimador $\bar{\boldsymbol{\alpha}}'$ na aula Tópico 15 SVD como ferramenta

$$\overline{\mathbf{p}}^* = \sum_{i=1}^{r} \frac{\beta_i}{S_i} \overline{\mathbf{v}}_i + \underbrace{\sum_{i=r+1}^{M} \frac{\beta_i}{S_i} \overline{\mathbf{v}}_i}_{\in \text{ Espaço Nulo}}$$

Vimos que no caso de M - r valores singulares nulos, podemos atribuir ao vetor $\overline{\boldsymbol{\alpha}}'_{M-r}$ qualquer valor, então teremos por conseqüência infinitas soluções (não unicidade da solução)

$$\overline{\mathbf{p}}^* = \overline{\overline{\mathbf{V}}}_r \; \overline{\boldsymbol{\alpha}}'_r + \underbrace{\overline{\overline{\mathbf{V}}}_{M-r} \; \overline{\boldsymbol{\alpha}}'_{M-r}}_{\overline{\mathbf{p}}^{null}} \qquad (3)$$

produzindo o mesmo mínimo da função Q dado pela equação (2). Em outras palavras temos a existência do espaço nulo e conseqüentemente M-r vetores soluções nulas, implicando uma solução geral dada por

$$\overline{\mathbf{p}}_{geral} = \overline{\mathbf{p}}_{particular} + \overline{\mathbf{p}}^{null}$$

Vimos também que no caso de M - r valores singulares próximos a zero temos:

$$\overline{\mathbf{p}}^* = \overline{\overline{\mathbf{V}}}_r \; \overline{\boldsymbol{\alpha}}'_r + \overline{\overline{\mathbf{V}}}_{M-r} \; \overline{\boldsymbol{\alpha}}'_{M-r} \qquad (4)$$

$$\overline{\mathbf{p}}^* = \sum_{i=1}^{r} \alpha'_i \, \overline{\mathbf{v}}_i + \sum_{i=r+1}^{M} \alpha'_i \, \overline{\mathbf{v}}_i \qquad (5)$$

$$\overline{\mathbf{p}}^* = \sum_{i=1}^{r} \frac{\beta_i}{S_i} \overline{\mathbf{v}}_i + \underbrace{\sum_{i=r+1}^{M} \frac{\beta_i}{S_i} \overline{\mathbf{v}}_i}_{\text{valores pequenos de } S_i} \qquad (6)$$

Como temos $S_i, i = r+1,..., M$ muito pequenos teremos uma instabilidade da solução promovida pela segunda parcela da equação (6), especificamente devida a parcela de $\overline{\boldsymbol{\alpha}}'$ correspondente a $\overline{\boldsymbol{\alpha}}'_{M-r}$ [equação (4)]. Neste caso, a solução é única porém, instável e produzirão um mínimo de Q [equação (2)].

Em resumo, em qualquer destes dois casos apresentados [caso (1) M - r valores singulares nulos e caso (2) M - r valores singulares muito pequenos] os elementos de $\overline{\boldsymbol{\alpha}}'_{M-r}$ não podem ser determinados apenas das observações geofísicas. Então buscaremos um modo de tornar o espaço desconhecido composto pelo vetor $\overline{\boldsymbol{\alpha}}'_{M-r}$ em um espaço conhecido. Para isto usaremos informações independentes das observações geofísicas para determinarmos os elementos do vetor $\overline{\boldsymbol{\alpha}}'_{M-r}$.

Formularemos então o seguinte problema matemático vinculado

Minimize $\left\| \overline{\overline{\mathbf{W}}}_{\mathbf{p}}^{\mathbf{1/2}} \ \overline{\mathbf{p}} \right\|_2^2$ (7)

Sujeito a

$$\left\| \overline{\overline{\mathbf{W}}}_{\mathbf{e}}^{\mathbf{1/2}} \ \overline{\boldsymbol{\varepsilon}} \ \right\|_2^2 = \delta$$

que pode ser re-escrito como

Minimize $\overline{\mathbf{p}}^{\mathrm{T}} \ \overline{\overline{\mathbf{W}}}_{\mathbf{p}} \ \overline{\mathbf{p}}$ (8)

Sujeito a

$$\left(\overline{\overline{\mathbf{A}}}\overline{\mathbf{p}} - \overline{\mathbf{y}}^0 \right)^{\mathrm{T}} \overline{\overline{\mathbf{W}}}_{\mathbf{e}} \left(\overline{\overline{\mathbf{A}}}\overline{\mathbf{p}} - \overline{\mathbf{y}}^0 \right) = \delta$$

Façamos

$$\bar{\bar{\mathbf{W}}}_{\mathbf{p}} = \bar{\bar{\mathbf{V}}}\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{0}}} & \bar{\bar{\boldsymbol{\eta}}} \end{bmatrix}\bar{\bar{\mathbf{V}}}^{\mathrm{T}}$$

em que $\bar{\bar{\boldsymbol{\eta}}}$ é uma matriz diagonal, não singular com dimensão $(\mathrm{M} - \mathrm{r} \times \mathrm{M} - \mathrm{r})$ e

$$\bar{\bar{\mathbf{W}}}_{\mathbf{e}} = \bar{\bar{\mathbf{U}}}\begin{bmatrix} \bar{\bar{\boldsymbol{\gamma}}} & \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{0}}} \end{bmatrix}\bar{\bar{\mathbf{U}}}^{\mathbf{T}}$$

em que $\bar{\bar{\boldsymbol{\gamma}}}$ é uma matriz diagonal, não singular com dimensão $(\mathrm{r} \times \mathrm{r})$.

Particionando a matriz $\bar{\bar{\mathbf{V}}}$ podemos re-escrever a matriz $\bar{\bar{\mathbf{W}}}_{\mathbf{p}}$ como

$$\bar{\bar{\mathbf{W}}}_{\mathbf{p}} = \begin{bmatrix} \bar{\bar{\mathbf{V}}}_{r} & \bar{\bar{\mathbf{V}}}_{M-r} \end{bmatrix}\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{0}}} & \bar{\bar{\boldsymbol{\eta}}} \end{bmatrix}\begin{bmatrix} \bar{\bar{\mathbf{V}}}_{r}^{T} \\ \bar{\bar{\mathbf{V}}}_{M-r}^{T} \end{bmatrix} = \bar{\bar{\mathbf{V}}}_{M-r}\ \bar{\bar{\boldsymbol{\eta}}}\ \bar{\bar{\mathbf{V}}}_{M-r}^{T} \tag{9}$$

Particionando a matriz $\bar{\bar{\mathbf{U}}}$ podemos re-escrever a matriz $\bar{\bar{\mathbf{W}}}_{\mathbf{e}}$ como

$$\bar{\bar{\mathbf{W}}}_{\mathbf{e}} = \begin{bmatrix} \bar{\bar{\mathbf{U}}}_{r} & \bar{\bar{\mathbf{U}}}_{N-r} \end{bmatrix}\begin{bmatrix} \bar{\bar{\boldsymbol{\gamma}}} & \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{0}}} \end{bmatrix}\begin{bmatrix} \bar{\bar{\mathbf{U}}}_{r}^{T} \\ \bar{\bar{\mathbf{U}}}_{N-r}^{T} \end{bmatrix} = \bar{\bar{\mathbf{U}}}_{r}\bar{\bar{\boldsymbol{\gamma}}}\bar{\bar{\mathbf{U}}}_{r}^{T} \tag{10}$$

Em resumo, o nosso problema consiste em resolver o seguinte problema vinculado

$$\text{Minimize} \quad \bar{\mathbf{p}}^{\mathrm{T}}\ \bar{\bar{\mathbf{W}}}_{\mathbf{p}}\ \bar{\mathbf{p}} \tag{11}$$

Sujeito a

$$\left(\bar{\bar{\mathbf{A}}}\bar{\mathbf{p}} - \bar{\mathbf{y}}^{0}\right)^{\mathrm{T}}\bar{\bar{\mathbf{W}}}_{\mathbf{e}}\left(\bar{\bar{\mathbf{A}}}\bar{\mathbf{p}} - \bar{\mathbf{y}}^{0}\right) = \delta$$

em que

$$\bar{\bar{\mathbf{W}}}_{\mathbf{p}} = \bar{\bar{\mathbf{V}}}_{M-r}\ \bar{\bar{\boldsymbol{\eta}}}\ \bar{\bar{\mathbf{V}}}_{M-r}^{T} \tag{12}$$

e

$$\overline{\overline{\mathbf{W}}}_{\mathbf{e}} = \overline{\overline{\mathbf{U}}}_r \;\overline{\overline{\boldsymbol{\gamma}}}\; \overline{\overline{\mathbf{U}}}_r^{T} \qquad (13)$$

Transformando um problema vinculado em um problema não vinculado via método dos multiplicadores de Lagrange temos que a função objeto não vinculada a ser minimizada é

$$\Gamma = \overline{\mathbf{p}}^{\mathrm{T}}\overline{\overline{\mathbf{W}}}_{\mathbf{p}}\;\overline{\mathbf{p}} + \frac{1}{\lambda}\left[\left(\overline{\overline{\mathbf{A}}}\overline{\mathbf{p}} - \overline{\mathbf{y}}^{0}\right)^{\mathrm{T}}\overline{\overline{\mathbf{W}}}_{\mathbf{e}}\left(\overline{\overline{\mathbf{A}}}\overline{\mathbf{p}} - \overline{\mathbf{y}}^{0}\right) - \delta\right] \qquad (14)$$

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

$$\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = 2\overline{\nabla}_{\overline{\mathbf{p}}}\left\{\overline{\mathbf{p}}^{\mathrm{T}}\right\}\overline{\overline{\mathbf{W}}}_{\mathbf{p}}\;\overline{\mathbf{p}} + \frac{1}{\lambda}2\left[\left(\overline{\nabla}_{\overline{\mathbf{p}}}\{\overline{\mathbf{p}}^{T}\overline{\overline{\mathbf{A}}}^{T}\} - \overline{\nabla}_{\overline{\mathbf{p}}}\{\overline{\mathbf{y}}^{0^{T}}\}\right)\overline{\overline{\mathbf{W}}}_{\mathbf{e}}\left(\overline{\overline{\mathbf{A}}}\overline{\mathbf{p}} - \overline{\mathbf{y}}^{0}\right)\right]$$

$$\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = 2\overline{\overline{\mathbf{W}}}_{\mathbf{p}}\;\overline{\mathbf{p}} + \frac{1}{\lambda}2\overline{\overline{\mathbf{A}}}^{T}\overline{\overline{\mathbf{W}}}_{\mathbf{e}}\left(\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - \overline{\mathbf{y}}^{0}\right)$$

$$\overline{\nabla}_{\overline{\mathbf{p}}}(\Gamma) = 2\lambda\overline{\overline{\mathbf{W}}}_{\mathbf{p}}\;\overline{\mathbf{p}} + 2\overline{\overline{\mathbf{A}}}^{T}\overline{\overline{\mathbf{W}}}_{\mathbf{e}}\;\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} - 2\overline{\overline{\mathbf{A}}}^{T}\overline{\overline{\mathbf{W}}}_{\mathbf{e}}\;\overline{\mathbf{y}}^{0}$$

igualando o vetor gradiente acima ao vetor nulo temos

$$2\lambda\overline{\overline{\mathbf{W}}}_{\mathbf{p}}\;\overline{\mathbf{p}}^{+} + 2\overline{\overline{\mathbf{A}}}^{T}\overline{\overline{\mathbf{W}}}_{\mathbf{e}}\;\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}}^{+} - 2\overline{\overline{\mathbf{A}}}^{T}\overline{\overline{\mathbf{W}}}_{\mathbf{e}}\;\overline{\mathbf{y}}^{0} = \overline{\mathbf{0}}$$

Dividindo ambos os lados da equação acima por 2 temos

$$\left(\overline{\overline{\mathbf{A}}}^{T}\overline{\overline{\mathbf{W}}}_{\mathbf{e}}\;\overline{\overline{\mathbf{A}}} + \lambda\overline{\overline{\mathbf{W}}}_{\mathbf{p}}\right)\overline{\mathbf{p}}^{+} = \overline{\overline{\mathbf{A}}}^{T}\overline{\overline{\mathbf{W}}}_{\mathbf{e}}\;\overline{\mathbf{y}}^{0} \qquad (15)$$

Vamos agora usar a decomposição em valores singulares da matriz de sensibilidade (i.e. $\overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{U}}}_r\,\overline{\overline{\mathbf{S}}}_r\,\overline{\overline{\mathbf{V}}}_r^{T}$) e substituir as matrizes (12) e (13) na equação acima

$$\left(\bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{S}}}_r \bar{\bar{\mathbf{U}}}_r^T \left(\bar{\bar{\mathbf{U}}}_r \;\bar{\bar{\boldsymbol{\gamma}}}\; \bar{\bar{\mathbf{U}}}_r^T\right) \bar{\bar{\mathbf{U}}}_r \bar{\bar{\mathbf{S}}}_r \bar{\bar{\mathbf{V}}}_r^T + \lambda \bar{\bar{\mathbf{V}}}_{M-r} \;\bar{\bar{\boldsymbol{\eta}}}\; \bar{\bar{\mathbf{V}}}_{M-r}^T\right) \bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{S}}}_r \bar{\bar{\mathbf{U}}}_r^T \left(\bar{\bar{\mathbf{U}}}_r \;\bar{\bar{\boldsymbol{\gamma}}}\; \bar{\bar{\mathbf{U}}}_r^T\right) \bar{\mathbf{y}}^0$$

$$\left(\bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{S}}}_r \;\bar{\bar{\boldsymbol{\gamma}}}\; \bar{\bar{\mathbf{S}}}_r \bar{\bar{\mathbf{V}}}_r^T + \lambda \bar{\bar{\mathbf{V}}}_{M-r} \;\bar{\bar{\boldsymbol{\eta}}}\; \bar{\bar{\mathbf{V}}}_{M-r}^T\right) \bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{S}}}_r \bar{\bar{\boldsymbol{\gamma}}} \;\bar{\bar{\mathbf{U}}}_r^T\; \bar{\mathbf{y}}^0$$

Vamos agora pré-multiplicar os dois lados da equação acima por $\bar{\bar{\mathbf{V}}}_r^T$

$$\bar{\bar{\mathbf{V}}}_r^T \left(\bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{S}}}_r \;\bar{\bar{\boldsymbol{\gamma}}}\; \bar{\bar{\mathbf{S}}}_r \bar{\bar{\mathbf{V}}}_r^T + \lambda \bar{\bar{\mathbf{V}}}_{M-r} \;\bar{\bar{\boldsymbol{\eta}}}\; \bar{\bar{\mathbf{V}}}_{M-r}^T\right) \bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{V}}}_r^T \bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{S}}}_r \bar{\bar{\boldsymbol{\gamma}}} \;\bar{\bar{\mathbf{U}}}_r^T\; \bar{\mathbf{y}}^0$$

$$\left(\bar{\bar{\mathbf{V}}}_r^T \bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{S}}}_r \;\bar{\bar{\boldsymbol{\gamma}}}\; \bar{\bar{\mathbf{S}}}_r \bar{\bar{\mathbf{V}}}_r^T + \lambda \bar{\bar{\mathbf{V}}}_r^T \bar{\bar{\mathbf{V}}}_{M-r} \;\bar{\bar{\boldsymbol{\eta}}}\; \bar{\bar{\mathbf{V}}}_{M-r}^T\right) \bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{V}}}_r^T \bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{S}}}_r \bar{\bar{\boldsymbol{\gamma}}} \;\bar{\bar{\mathbf{U}}}_r^T\; \bar{\mathbf{y}}^0$$

Como $\bar{\bar{\mathbf{V}}}_r^T \bar{\bar{\mathbf{V}}}_{M-r} = \bar{\mathbf{0}}$ porque os autovetores são ortogonais então temos

$$\left(\bar{\bar{\mathbf{S}}}_r \;\bar{\bar{\boldsymbol{\gamma}}}\; \bar{\bar{\mathbf{S}}}_r \bar{\bar{\mathbf{V}}}_r^T\right) \bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{S}}}_r \bar{\bar{\boldsymbol{\gamma}}} \;\bar{\bar{\mathbf{U}}}_r^T\; \bar{\mathbf{y}}^0 \qquad (16)$$

Como as matrizes $\bar{\bar{\boldsymbol{\gamma}}}$ e $\bar{\bar{\mathbf{S}}}_r$ são matrizes não singulares podemos escrever

$$\left(\bar{\bar{\mathbf{V}}}_r^T\right) \bar{\mathbf{p}}^+ = \left(\bar{\bar{\mathbf{S}}}_r \;\bar{\bar{\boldsymbol{\gamma}}}\; \bar{\bar{\mathbf{S}}}_r\right)^{-1} \bar{\bar{\mathbf{S}}}_r \bar{\bar{\boldsymbol{\gamma}}} \;\bar{\bar{\mathbf{U}}}_r^T\; \bar{\mathbf{y}}^0$$

$$\bar{\bar{\mathbf{V}}}_r^T \bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{S}}}_r^{-1} \;\bar{\bar{\boldsymbol{\gamma}}}^{-1}\; \bar{\bar{\mathbf{S}}}_r^{-1} \;\bar{\bar{\mathbf{S}}}_r \;\bar{\bar{\boldsymbol{\gamma}}}\; \bar{\bar{\mathbf{U}}}_r^T\; \bar{\mathbf{y}}^0$$

$$\bar{\bar{\mathbf{V}}}_r^T \bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{S}}}_r^{-1} \;\bar{\bar{\mathbf{U}}}_r^T\; \bar{\mathbf{y}}^0 \qquad (17)$$

Se r = M a solução $\overline{\mathbf{p}}^{+}$ que satisfaz a equação (17) é única porque, se multiplicarmos os dois lados da equação (17) por $\overline{\overline{\mathbf{V}}}_{r}$ teremos, neste caso $\overline{\overline{\mathbf{V}}}_{r}\overline{\overline{\mathbf{V}}}_{r}^{T} = \overline{\overline{\mathbf{I}}}_{M}$ logo a equação (17) é dada por

$$\overline{\mathbf{p}}^{+} = \overline{\overline{\mathbf{V}}}_{r} \; \overline{\overline{\mathbf{S}}}_{r}^{-1} \; \overline{\overline{\mathbf{U}}}_{r}^{T} \; \overline{\mathbf{y}}^{0} \qquad \text{se r = M} \qquad (18)$$

O estimador acima $\overline{\mathbf{p}}^{+}$ é chamado como **estimador IG (Inversa Generalizada).**

Por outro lado, se r < M, não há uma solução única que satisfaz a equação (17). Isto porque $\overline{\overline{\mathbf{V}}}_{r}$ é singular. Então se multiplicarmos os dois lados da equação (17) por $\overline{\overline{\mathbf{V}}}_{r}$ teremos

$$\overline{\overline{\mathbf{V}}}_{r} \; \overline{\overline{\mathbf{V}}}_{r}^{T} \overline{\mathbf{p}}^{+} = \overline{\overline{\mathbf{V}}}_{r} \; \overline{\overline{\mathbf{S}}}_{r}^{-1} \; \overline{\overline{\mathbf{U}}}_{r}^{T} \; \overline{\mathbf{y}}^{0} \qquad \text{se r < M} \qquad (19)$$

Note, neste caso em que r< M, não há uma solução única para $\overline{\mathbf{p}}^{+}$ isto porque $\overline{\overline{\mathbf{V}}}_{r}\overline{\overline{\mathbf{V}}}_{r}^{T}$ é uma matriz é singular[2].

No entanto, observe que apesar de não existir uma solução única da equação (19) podemos afirmar que o estimador IG [equação (18)] é uma das possíveis soluções isto porque se substituirmos a equação (18) na equação (19) temos

$$\overline{\overline{\mathbf{V}}}_{r} \; \overline{\overline{\mathbf{V}}}_{r}^{T} \left(\overline{\overline{\mathbf{V}}}_{r} \; \overline{\overline{\mathbf{S}}}_{r}^{-1} \; \overline{\overline{\mathbf{U}}}_{r}^{T} \; \overline{\mathbf{y}}^{0} \right) = \overline{\overline{\mathbf{V}}}_{r} \; \overline{\overline{\mathbf{S}}}_{r}^{-1} \; \overline{\overline{\mathbf{U}}}_{r}^{T} \; \overline{\mathbf{y}}^{0}$$

$$\overline{\overline{\mathbf{V}}}_{r} \; \overline{\overline{\mathbf{V}}}_{r}^{T} \overline{\overline{\mathbf{V}}}_{r} \; \overline{\overline{\mathbf{S}}}_{r}^{-1} \; \overline{\overline{\mathbf{U}}}_{r}^{T} \; \overline{\mathbf{y}}^{0} = \overline{\overline{\mathbf{V}}}_{r} \; \overline{\overline{\mathbf{S}}}_{r}^{-1} \; \overline{\overline{\mathbf{U}}}_{r}^{T} \; \overline{\mathbf{y}}^{0}$$

[2] Lembre-se que $\overline{\overline{\mathbf{V}}} \; \overline{\overline{\mathbf{V}}}^{T} = \overline{\overline{\mathbf{V}}}_{\mathrm{r}}\overline{\overline{\mathbf{V}}}_{\mathrm{r}}^{T} + \overline{\overline{\mathbf{V}}}_{\mathrm{M-r}}\overline{\overline{\mathbf{V}}}_{\mathrm{M-r}}^{T} = \overline{\overline{\mathbf{I}}}_{M}$

$$\overline{\overline{\mathbf{V}}}_r \; \overline{\overline{\mathbf{S}}}_r^{-1} \; \overline{\overline{\mathbf{U}}}_r^T \; \overline{\mathbf{y}}^0 \equiv \overline{\overline{\mathbf{V}}}_r \; \overline{\overline{\mathbf{S}}}_r^{-1} \; \overline{\overline{\mathbf{U}}}_r^T \; \overline{\mathbf{y}}^0$$

Análise da Informação a priori introduzida pelo estimador Inversa Generalizada

Introduzimos a informação a priori da solução estimada

mínimar $\phi\ (\overline{\mathbf{p}}) = \left\| \overline{\overline{\mathbf{W}}}_{\mathbf{p}}^{1/2}\ \overline{\mathbf{p}} \right\|_2^2$

i.e.,

minimizar $\phi\ (\overline{\mathbf{p}}) = \overline{\mathbf{p}}^{\mathrm{T}}\ \overline{\overline{\mathbf{W}}}_{\mathbf{p}}\ \overline{\mathbf{p}}$

em que $\overline{\overline{\mathbf{W}}}_{\mathbf{p}} = \overline{\overline{\mathbf{V}}}_{M-r}\ \overline{\overline{\boldsymbol{\eta}}}\ \overline{\overline{\mathbf{V}}}_{M-r}^T$

Portanto, minimizamos a função

$$\phi\ (\overline{\mathbf{p}}) = \overline{\mathbf{p}}^{\mathrm{T}}\ \overline{\overline{\mathbf{V}}}_{M-r}\ \overline{\overline{\boldsymbol{\eta}}}\ \overline{\overline{\mathbf{V}}}_{M-r}^T\ \overline{\mathbf{p}} \qquad (20)$$

Considerando que o vetor $\overline{\boldsymbol{\alpha}}$ é a projeção dos parâmetros $\overline{\mathbf{p}}$ nos autovetores da matriz $\overline{\overline{\mathbf{V}}}$ (na base que gera o espaço dos parâmetros), i.e.,

$$\overline{\overline{\mathbf{V}}}^T\ \overline{\mathbf{p}} = \overline{\boldsymbol{\alpha}}\ .$$

Se r = m pré-multiplicando os dois lados da equação por $\overline{\overline{\mathbf{V}}}$ temos que

$$\overline{\overline{\mathbf{V}}}\ \overline{\overline{\mathbf{V}}}^T\ \overline{\mathbf{p}} = \overline{\overline{\mathbf{V}}}\ \overline{\boldsymbol{\alpha}}$$

$$\overline{\mathbf{p}} = \overline{\overline{\mathbf{V}}}\ \overline{\boldsymbol{\alpha}}$$

Substituindo o vetor $\overline{\mathbf{p}} = \overline{\overline{\mathbf{V}}}\,\overline{\boldsymbol{\alpha}}$ na equação (20) temos:

$$\phi\ (\overline{\mathbf{p}}) = \overline{\boldsymbol{\alpha}}^{\mathrm{T}}\ \overline{\overline{\mathbf{V}}}^{\mathrm{T}}\ \overline{\overline{\mathbf{V}}}_{M-r}\ \overline{\overline{\boldsymbol{\eta}}}\ \overline{\overline{\mathbf{V}}}_{M-r}^T\ \overline{\overline{\mathbf{V}}}\ \overline{\boldsymbol{\alpha}}$$

$$\phi\ (\bar{\mathbf{p}}) = \underset{(1\times2)}{\begin{bmatrix} \bar{\boldsymbol{\alpha}}_{\mathrm{r}}^{\mathrm{T}} & \bar{\boldsymbol{\alpha}}_{M-r}^{\mathrm{T}} \end{bmatrix}} \underset{(2\times1)}{\begin{bmatrix} \bar{\bar{\mathbf{V}}}_{r}^{T} \\ \bar{\bar{\mathbf{V}}}_{M-r}^{T} \end{bmatrix}} \underset{(1\times2)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{V}}}_{M-r} \end{bmatrix}} \underset{(2\times2)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{0}}} & \bar{\bar{\boldsymbol{\eta}}} \end{bmatrix}} \underset{(2\times1)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{V}}}_{M-r}^{T} \end{bmatrix}} \underset{(1\times2)}{\begin{bmatrix} \bar{\bar{\mathbf{V}}}_{r} & \bar{\bar{\mathbf{V}}}_{M-r} \end{bmatrix}} \underset{(2\times1)}{\begin{bmatrix} \bar{\boldsymbol{\alpha}}_{\mathrm{r}} \\ \bar{\boldsymbol{\alpha}}_{\mathrm{M-r}} \end{bmatrix}}$$

$$\phi\ (\bar{\mathbf{p}}) = \underset{(1\times1)}{\left(\bar{\boldsymbol{\alpha}}_{\mathrm{r}}^{\mathrm{T}} \bar{\bar{\mathbf{V}}}_{r}^{T} + \bar{\boldsymbol{\alpha}}_{M-r}^{\mathrm{T}} \bar{\bar{\mathbf{V}}}_{M-r}^{T} \right)} \underset{(1\times2)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{V}}}_{M-r} \end{bmatrix}} \underset{(2\times2)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{0}}} & \bar{\bar{\boldsymbol{\eta}}} \end{bmatrix}} \underset{(2\times1)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{V}}}_{M-r}^{T} \end{bmatrix}} \underset{(1\times1)}{\left(\bar{\bar{\mathbf{V}}}_{r} \bar{\boldsymbol{\alpha}}_{\mathrm{r}} + \bar{\bar{\mathbf{V}}}_{M-r} \bar{\boldsymbol{\alpha}}_{\mathrm{M-r}} \right)}$$

$$\phi\ (\bar{\mathbf{p}}) = \underset{(1\times2)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\boldsymbol{\alpha}}_{M-r}^{\mathrm{T}} \bar{\bar{\mathbf{V}}}_{M-r}^{T} \bar{\bar{\mathbf{V}}}_{M-r} \end{bmatrix}} \underset{(2\times2)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{0}}} & \bar{\bar{\boldsymbol{\eta}}} \end{bmatrix}} \underset{(2\times1)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{V}}}_{M-r}^{T} \bar{\bar{\mathbf{V}}}_{M-r} \bar{\boldsymbol{\alpha}}_{\mathrm{M-r}} \end{bmatrix}}$$

Como $\bar{\bar{\mathbf{V}}}_{M-r}^{T} \bar{\bar{\mathbf{V}}}_{M-r} = \bar{\bar{\mathbf{I}}}_{M-r}$

$$\phi\ (\bar{\mathbf{p}}) = \underset{(1\times2)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\boldsymbol{\alpha}}_{M-r}^{\mathrm{T}} \bar{\bar{\mathbf{I}}}_{M-r} \end{bmatrix}} \underset{(2\times2)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{0}}} & \bar{\bar{\boldsymbol{\eta}}} \end{bmatrix}} \underset{(2\times1)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} \\ \bar{\boldsymbol{\alpha}}_{\mathrm{M-r}} \bar{\bar{\mathbf{I}}}_{M-r} \end{bmatrix}}$$

$$\phi\ (\bar{\mathbf{p}}) = \underset{(1\times2)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\boldsymbol{\alpha}}_{M-r}^{\mathrm{T}} \bar{\bar{\boldsymbol{\eta}}} \end{bmatrix}} \underset{(2\times1)}{\begin{bmatrix} \bar{\bar{\mathbf{0}}} \\ \bar{\boldsymbol{\alpha}}_{\mathrm{M-r}} \bar{\bar{\mathbf{I}}}_{M-r} \end{bmatrix}}$$

$$\phi\ (\bar{\mathbf{p}}) = \bar{\boldsymbol{\alpha}}_{M-r}^{\mathrm{T}} \bar{\bar{\boldsymbol{\eta}}}\ \bar{\boldsymbol{\alpha}}_{\mathrm{M-r}} \qquad (21)$$

A informação a priori introduzida pelo estimador IG é de minimizar o funcional

$$\phi\ (\bar{\mathbf{p}}) = \bar{\mathbf{p}}^{\mathrm{T}} \bar{\bar{\mathbf{W}}}_{\mathbf{p}} \bar{\mathbf{p}}$$

Mostramos que a minimização deste funcional acima é equivalente a minimizar o funcional (21)

$$\phi\ (\bar{\mathbf{p}}) = \bar{\boldsymbol{\alpha}}_{M-r}^{\mathrm{T}} \bar{\bar{\boldsymbol{\eta}}}\ \bar{\boldsymbol{\alpha}}_{\mathrm{M-r}}$$

ou ainda que

$$\phi\ (\overline{\mathbf{p}}) = \sum_{i=r+1}^{M} \alpha_i^2 \eta_i$$

i.e., a informação introduzida pelo estimador IG é equivalente a minimizar a seguinte soma

$$\phi\ (\overline{\mathbf{p}}) = \alpha_{r+1}^2 \eta_1 + \alpha_{r+2}^2 \eta_2 + \ldots + \alpha_M^2 \eta_{M-r}\,. \qquad (22)$$

Note, para que o funcional $\phi\ (\overline{\mathbf{p}}) = \overline{\mathbf{p}}^T \overline{\overline{\mathbf{W}}}_{\mathbf{p}}\, \overline{\mathbf{p}}$ seja mínimo há duas possibilidades:

(1) Estabelecer valores infinitamente grandes para os elementos $\eta_i, \quad i = 1,2,\ldots,\ M - r$. Neste caso a matriz de pesos $\overline{\overline{\mathbf{W}}}_{\mathbf{p}}$ assumiria um padrão particular, ou seja

$$\overline{\overline{\mathbf{W}}}_{\mathbf{p}} = \overline{\overline{\mathbf{V}}} \begin{bmatrix} \overline{\overline{\mathbf{0}}} & \overline{\overline{\mathbf{0}}} \\ \overline{\overline{\mathbf{0}}} & \overline{\overline{\boldsymbol{\eta}}} \end{bmatrix} \overline{\overline{\mathbf{V}}}^T = \overline{\overline{\mathbf{V}}} \begin{bmatrix} \overline{\overline{\mathbf{0}}} & \overline{\overline{\mathbf{0}}} \\ \overline{\overline{\mathbf{0}}} & \overline{\overline{\infty}} \end{bmatrix} \overline{\overline{\mathbf{V}}}^T$$

Note que esta primeira possibilidade seria um caso particular, ao contrário queremos que $\phi\ (\overline{\mathbf{p}}) = \overline{\mathbf{p}}^T \overline{\overline{\mathbf{W}}}_{\mathbf{p}}\, \overline{\mathbf{p}}$ seja mínimo para quaisquer valores de $\eta_i, \quad i = 1,2,\ldots,\ M - r$

(2) A segunda possibilidade para que o funcional $\phi\ (\overline{\mathbf{p}}) = \overline{\mathbf{p}}^T \overline{\overline{\mathbf{W}}}_{\mathbf{p}}\, \overline{\mathbf{p}}$ seja mínimo é estabelecer que os $\alpha_i, \quad i = r + 1,\ r + 2,\ldots,\ M$ devem ser nulos

Diante destas duas possibilidade concluímos que o estimador IG utiliza a informação a priori que o vetor $\overline{\boldsymbol{\alpha}}_{M\ -\ r} = \overline{\mathbf{0}}$ uma vez que este estimador (caso r = M) é dado por

$$\bar{\mathbf{p}}^{+} = \bar{\bar{\mathbf{V}}}_{r} \; \bar{\bar{\mathbf{S}}}_{r}^{-1} \; \bar{\bar{\mathbf{U}}}_{r}^{T} \; \bar{\mathbf{y}}^{0}$$

que não utiliza os valores da matriz diagonal $\bar{\bar{\boldsymbol{\eta}}}$.

Em resumo, o estimador IG introduz a informação a priori que a projeção do vetor de parâmetros nas bases que geram o espaço nulo é o vetor nulo. Em outras palavras, o estimador IG introduz a informação a priori que

$$\bar{\boldsymbol{\alpha}}_{\mathrm{M-r}} = \bar{\mathbf{0}} .$$

Como o vetor $\bar{\boldsymbol{\alpha}}_{M-r}$ é a projeção da solução nula na base do espaço nulo do espaço dos parâmetros, i.e.,

$$\bar{\mathbf{p}}^{null} = \bar{\bar{\mathbf{V}}}_{M-r} \; \bar{\boldsymbol{\alpha}}_{M-r} ,$$

então podemos dizer que o estimador IG introduz a informação a priori que a solução nula é igual ao vetor nulo

$$\bar{\mathbf{p}}^{null} = \bar{\mathbf{0}}$$

Deduzindo o estimador Inversa Generalizada via minimização de Q e a introdução direta que $\bar{\boldsymbol{\alpha}}_{\mathrm{M-r}} = \bar{\mathbf{0}}$

Já vimos no Tópico 15 que

$$Q = \| \bar{\boldsymbol{\varepsilon}} \|_2^2 = \left\| \bar{\boldsymbol{\beta}}_r - \bar{\bar{\mathbf{S}}}_r \; \bar{\boldsymbol{\alpha}}_r \right\|_2^2 + \left\| \bar{\boldsymbol{\beta}}_{M-r} - \bar{\bar{\mathbf{S}}}_{M-r} \; \bar{\boldsymbol{\alpha}}_{M-r} \right\|_2^2 + \left\| \bar{\boldsymbol{\beta}}_{N-M} \right\|_2^2$$

considerando que no IG $\bar{\boldsymbol{\alpha}}_{\mathrm{M-r}} = \bar{\mathbf{0}}$ temos

$$Q = \left\| \bar{\boldsymbol{\beta}}_r - \bar{\bar{\mathbf{S}}}_r \ \bar{\boldsymbol{\alpha}}_r \right\|_2^2 + \left\| \bar{\boldsymbol{\beta}}_{M-r} \right\|_2^2 + \left\| \bar{\boldsymbol{\beta}}_{N-M} \right\|_2^2$$

$$Q = (\bar{\boldsymbol{\beta}}_r - \bar{\bar{\mathbf{S}}}_r \ \bar{\boldsymbol{\alpha}}_r)^T (\bar{\boldsymbol{\beta}}_r - \bar{\bar{\mathbf{S}}}_r \ \bar{\boldsymbol{\alpha}}_r) + \bar{\boldsymbol{\beta}}^T_{M-r} \bar{\boldsymbol{\beta}}_{M-r} + \bar{\boldsymbol{\beta}}^T_{N-M} \ \bar{\boldsymbol{\beta}}_{N-M}$$

A condição para que Q seja mínimo é:

$$\bar{\nabla}_{\bar{\boldsymbol{\alpha}}_r} \{Q\} = \bar{\mathbf{0}}$$

$$\bar{\nabla}_{\bar{\boldsymbol{\alpha}}_r} \{Q\} = 2 \ \bar{\nabla}_{\bar{\boldsymbol{\alpha}}_r} \left\{ (\bar{\boldsymbol{\beta}}_r - \bar{\bar{\mathbf{S}}}_r \ \bar{\boldsymbol{\alpha}}_r)^T \right\} (\bar{\boldsymbol{\beta}}_r - \bar{\bar{\mathbf{S}}}_r \ \bar{\boldsymbol{\alpha}}_r)$$

$$\bar{\nabla}_{\bar{\boldsymbol{\alpha}}_r} \{Q\} = -2 \ \bar{\bar{\mathbf{S}}}_r^{\mathrm{T}} \ (\bar{\boldsymbol{\beta}}_r - \bar{\bar{\mathbf{S}}}_r \ \bar{\boldsymbol{\alpha}}_r)$$

$$\bar{\nabla}_{\bar{\boldsymbol{\alpha}}_r} \{Q\} = -2 \ \bar{\bar{\mathbf{S}}}_r^{\mathrm{T}} \ \bar{\boldsymbol{\beta}}_r + 2 \ \bar{\bar{\mathbf{S}}}_r^{\mathrm{T}} \bar{\bar{\mathbf{S}}}_r \ \bar{\boldsymbol{\alpha}}_r$$

Como $\bar{\bar{\mathbf{S}}}_r$ é uma matriz diagonal então

$$\bar{\nabla}_{\bar{\boldsymbol{\alpha}}_r} \{Q\} = -2 \ \bar{\bar{\mathbf{S}}}_r \ \bar{\boldsymbol{\beta}}_r + 2 \ \bar{\bar{\mathbf{S}}}_r^{2} \ \bar{\boldsymbol{\alpha}}_r$$

Pela condição de mínimo i.e. $\bar{\nabla}_{\bar{\boldsymbol{\alpha}}_r} \{Q\} = \bar{\mathbf{0}}$ temos

$$-2 \ \bar{\bar{\mathbf{S}}}_r \ \bar{\boldsymbol{\beta}}_r + 2 \ \bar{\bar{\mathbf{S}}}_r^{2} \ \bar{\boldsymbol{\alpha}}^*_r = \bar{\mathbf{0}}$$

$$2 \ \bar{\bar{\mathbf{S}}}_r^{2} \ \bar{\boldsymbol{\alpha}}^*_r = 2 \ \bar{\bar{\mathbf{S}}}_r \ \bar{\boldsymbol{\beta}}_r$$

$$\bar{\boldsymbol{\alpha}}^*_r = \bar{\bar{\mathbf{S}}}_r^{-1} \ \bar{\boldsymbol{\beta}}_r \qquad (23)$$

Como o vetor $\bar{\boldsymbol{\alpha}}^*$ é a projeção do vetor de parâmetros nos vetores base que geram o espaço de parâmetros (i.e., vetores colunas da matriz $\bar{\bar{\mathbf{V}}}$ de autovetores) então temos que o estimador IG é dado por

$$\bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{V}}}_r\,\bar{\boldsymbol{\alpha}}^*_r + \bar{\bar{\mathbf{V}}}_{M-r}\,\bar{\boldsymbol{\alpha}}^*_{M-r}$$

Como o estimador IG introduz a informação a priori que $\bar{\boldsymbol{\alpha}}_{M-r} = \bar{\mathbf{0}}$ temos

$$\bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{V}}}_r\,\bar{\boldsymbol{\alpha}}^*_r$$

Substituindo a equação (23) na equação acima temos:

$$\bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{V}}}_r\,\bar{\bar{\mathbf{S}}}_r^{-1}\,\bar{\boldsymbol{\beta}}_r$$

Lembrando que o vetor $\bar{\boldsymbol{\beta}}$ é a projeção do vetor dos dados observados nos vetores base que geram o espaço das observações i.e. $\bar{\bar{\mathbf{U}}}^T\bar{\mathbf{y}}^o = \bar{\boldsymbol{\beta}}$. Então temos que

$$\begin{bmatrix} \bar{\boldsymbol{\beta}}_r \\ \bar{\boldsymbol{\beta}}_{M-r} \\ \bar{\boldsymbol{\beta}}_{N-M} \end{bmatrix} = \begin{bmatrix} \bar{\bar{\mathbf{U}}}_r \\ \bar{\bar{\mathbf{U}}}_{M-r} \\ \bar{\bar{\mathbf{U}}}_{N-M} \end{bmatrix} \bar{\mathbf{y}}^o$$

logo

$$\bar{\boldsymbol{\beta}}_r = \bar{\bar{\mathbf{U}}}_r^T\,\bar{\mathbf{y}}^o$$

$$\bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{V}}}_r\,\bar{\bar{\mathbf{S}}}_r^{-1}\,\bar{\bar{\mathbf{U}}}_r^T\,\bar{\mathbf{y}}^o \qquad (24)$$

O estimador IG impõe que $\bar{\boldsymbol{\alpha}}_{M-r} = \bar{\mathbf{0}}$. Portanto o IG impõe que a solução estimada $\bar{\mathbf{p}}^{+}$ seja ortogonal ao espaço nulo.. Em outras palavras, o estimador IG introduz a informação a priori que a projeção da solução estimada $\bar{\mathbf{p}}^{+}$ no espaço nulo seja zero.

<u>Analisando o estimador Inversa Generalizada para dados com ruído em comparação ao estimador MQ</u>

Vimos que o estimador IG é definido como

$$\bar{\mathbf{p}}^{+} = \bar{\bar{\mathbf{V}}}_r \; \bar{\bar{\mathbf{S}}}_r^{-1} \; \bar{\bar{\mathbf{U}}}_r^{T} \, \bar{\mathbf{y}}^{o}$$

ou ainda

$$\bar{\mathbf{p}}^{+} = \bar{\bar{\mathbf{V}}}_r \; \bar{\bar{\mathbf{S}}}_r^{-1} \; \bar{\boldsymbol{\beta}}_r$$

Se, em uma situação hipotética, pudéssemos decompor os dados observados $\bar{\mathbf{y}}^{o}$ em duas parcelas sendo que uma delas é o dado livre de ruído $\bar{\mathbf{y}}^{Exato}$ e a segunda parcela é o ruído aditivo $\bar{\boldsymbol{\varepsilon}}$, então teríamos $\bar{\mathbf{y}}^{o} = \bar{\mathbf{y}}^{Exato} + \bar{\boldsymbol{\varepsilon}}$. Nesta situação hipotética o vetor $\bar{\boldsymbol{\beta}}$ é dado por

$$\bar{\boldsymbol{\beta}}_r = \bar{\bar{\mathbf{U}}}_r^{T} \, \bar{\mathbf{y}}^{o}$$

$$\bar{\boldsymbol{\beta}}_r = \bar{\bar{\mathbf{U}}}_r^{T} \left(\bar{\mathbf{y}}^{Exato} + \bar{\boldsymbol{\varepsilon}} \right)$$

$$\bar{\boldsymbol{\beta}}_r = \bar{\bar{\mathbf{U}}}_r^{T} \, \bar{\mathbf{y}}^{Exato} + \bar{\bar{\mathbf{U}}}_r^{T} \, \bar{\boldsymbol{\varepsilon}}$$

$$\bar{\boldsymbol{\beta}}_r = \bar{\boldsymbol{\beta}}_r^{Exato} + \bar{\boldsymbol{\beta}}_r^{Ruido}$$

Portanto o vetor $\overline{\boldsymbol{\beta}}_r$ foi dividido em duas parcelas em que a primeira parcela ($\overline{\boldsymbol{\beta}}_r^{\,Exato}$) é a projeção dos dados hipotéticos livre de ruído $\overline{\mathbf{y}}^{Exato}$ nos autovetores da matriz $\overline{\overline{\mathbf{U}}}_r$ e a segunda parcela ($\overline{\boldsymbol{\beta}}_r^{\,Ruido}$) é a projeção o ruído $\overline{\boldsymbol{\varepsilon}}$ nos autovetores da matriz $\overline{\overline{\mathbf{U}}}_r$. Substituindo a expressão acima no estimador IG temos que

$$\overline{\mathbf{p}}^{+} = \overline{\overline{\mathbf{V}}}_r \; \overline{\overline{\mathbf{S}}}_r^{-1} \left(\overline{\boldsymbol{\beta}}_r^{\,Exato} + \overline{\boldsymbol{\beta}}_r^{\,Ruido}\right)$$

$$\overline{\mathbf{p}}^{+} = \overline{\overline{\mathbf{V}}}_r \; \overline{\overline{\mathbf{S}}}_r^{-1} \overline{\boldsymbol{\beta}}_r^{\,Exato} + \overline{\overline{\mathbf{V}}}_r \; \overline{\overline{\mathbf{S}}}_r^{-1} \overline{\boldsymbol{\beta}}_r^{\,Ruido}$$

$$\overline{\mathbf{p}}^{+} = \sum_{i=1}^{r} \frac{\beta_i^{\,Exato}}{S_i} \; \overline{\mathbf{v}}_i + \sum_{i=1}^{r} \frac{\beta_i^{\,Ruido}}{S_i} \; \overline{\mathbf{v}}_i \qquad (25)$$

Vamos agora comparar o estimador IG [equação (25)] com o estimador MQ usando a mesma situação hipotética em que os dados observados $\overline{\mathbf{y}}^{o}$ foram decompostos em duas parcelas sendo que uma delas é o dado livre de ruído $\overline{\mathbf{y}}^{Exato}$ e a segunda parcela é o ruído aditivo $\overline{\boldsymbol{\varepsilon}}$.

Vimos no Tópico 15 que o estimador MQ é expresso como

$$\hat{\overline{\mathbf{p}}} = \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \; \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1} \; \overline{\boldsymbol{\beta}}$$

Vimos que nesta situação hipotética o i-ésimo elemento do vetor N-dimensional $\overline{\boldsymbol{\beta}}$ é dado

$$\beta_i = \beta_i^{\,Exato} + \beta_i^{\,Ruido}$$

Substituindo a expressão acima no estimador MQ temos

$$\hat{\mathbf{p}} = \sum_{i=1}^{M} \frac{\beta_i^{Exato}}{S_i} \bar{\mathbf{v}}_i + \sum_{i=1}^{M} \frac{\beta_i^{Ruido}}{S_i} \bar{\mathbf{v}}_i \qquad (26)$$

Comparando o estimador MQ (equação 26) com o estimador estimador IG (equação 25) vemos que estes estimadores são exatamente iguais se r = M. Veja também que a segunda parcela do estimador MQ, i.e., o segundo termo da equação 26

$$\sum_{i=1}^{M} \frac{\beta_i^{Ruido}}{S_i} \bar{\mathbf{v}}_i$$

pode crescer infinitamente se $S_i \approx 0$. Portanto a solução estimada via MQ com dados com ruído pode estar infinitamente distante da solução estimada com dados livres de ruído (dados exatos) devido a segunda parcela do somatório da equação (26) em que cada termo do somatório pode ser muito grande se $S_i \rightarrow 0$ amplificando o ruído $\bar{\boldsymbol{\varepsilon}}$ contido no vetor $\bar{\boldsymbol{\beta}}^{Ruido} = \bar{\bar{\mathbf{U}}}^T \bar{\boldsymbol{\varepsilon}}$ (projeção do ruído nos autovetores do espaço das observações).

Diferentemente, no caso do estimador IG equação 25) a segunda parcela é um somatório envolvendo apenas r valores singulares e não os M valores singulares como no estimador MQ. Portanto o estimador IG tem a flexibilidade de não utilizar os M valores singulares do sistema, mas sim os r (r $\leq M$) valores singulares convenientes de modo a tornar o problema estável. Para alcançar esta estabilidade da solução estimada via IG, basta estipular um valor r < M de modo a eliminar aqueles valores singulares próximos a zero. Veja então que eliminando todos os valores singulares próximos a zero a segunda parcela do estimador IG, i.e., $\sum_{i=1}^{r} \frac{\beta_i^{Ruido}}{S_i} \bar{\mathbf{v}}_i$, não promoverá a amplificação indesejável do numerador desta parcela e com isto não ocorrerá a amplificação do ruído dos dados observados.

Lembrando que os valores singulares são ordenandos em ordem decrescente, i.e., $S_1 \geq S_2 \geq S_3 \geq \ldots \geq S_r \geq S_{r+1} \geq \ldots \geq S_M$, então quanto MENOR o valor atribuído a r MAIOR a estabilidade introduzida na determinação da solução via IG

TRADE OFF : Estabilidade X Tendenciosidade

Vimos então que há uma relação entre o valor de r e a estabilidade da solução estimada via IG. O preço pago por esta estabilização é a introdução de uma tendenciosidade na estimativa dos parâmetros via IG. Veja que se os dados forem exatos (não contaminados por ruído) a estimativa dos parâmetros via MQ é dada por

$$\hat{\bar{\mathbf{p}}} = \sum_{i=1}^{M} \frac{\beta_i}{S_i} \bar{\mathbf{v}}_i$$

Porem, sob esta mesma condição o estimador IG é expresso

$$\bar{\mathbf{p}}^{+} = \sum_{i=1}^{r} \frac{\beta_i}{S_i} \bar{\mathbf{v}}_i$$

Note que a diferença entre $\hat{\bar{\mathbf{p}}}$ e $\bar{\mathbf{p}}^{+}$ será tanto maior quanto menor for o valor de r

$$\hat{\bar{\mathbf{p}}} - \bar{\mathbf{p}}^{+} = \sum_{i=1}^{M} \frac{\beta_i}{S_i} - \sum_{i=1}^{r} \frac{\beta_i}{S_i} \bar{\mathbf{v}}_i$$

$$\hat{\bar{\mathbf{p}}} - \bar{\mathbf{p}}^{+} = \sum_{i=r+1}^{M} \frac{\beta_i}{S_i} \bar{\mathbf{v}}_i$$

Em notação matricial esta diferença é dada por

$$\hat{\bar{\mathbf{p}}} - \bar{\mathbf{p}}^{+} = \bar{\bar{\mathbf{V}}}_M \, \bar{\bar{\mathbf{S}}}_M^{-1} \, \bar{\boldsymbol{\beta}}_M - \bar{\bar{\mathbf{V}}}_r \, \bar{\bar{\mathbf{S}}}_r^{-1} \, \bar{\boldsymbol{\beta}}_r$$

$$\hat{\bar{\mathbf{p}}} - \bar{\mathbf{p}}^{+} = \begin{bmatrix} \bar{\bar{\mathbf{V}}}_r & \bar{\bar{\mathbf{V}}}_{M-r} \end{bmatrix} \begin{bmatrix} \bar{\bar{\mathbf{S}}}_r & \\ & \bar{\bar{\mathbf{S}}}_{M-r} \end{bmatrix}^{-1} \begin{bmatrix} \bar{\boldsymbol{\beta}}_r \\ \bar{\boldsymbol{\beta}}_{N-r} \end{bmatrix} - \begin{bmatrix} \bar{\bar{\mathbf{V}}}_r & \bar{\bar{0}} \end{bmatrix} \begin{bmatrix} \bar{\bar{\mathbf{S}}}_r & \\ & \bar{\bar{0}} \end{bmatrix}^{-1} \begin{bmatrix} \bar{\boldsymbol{\beta}}_r \\ \bar{0} \end{bmatrix}$$

$$\hat{\mathbf{p}} - \bar{\mathbf{p}}^{+} = \bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{S}}}_r^{-1} \bar{\boldsymbol{\beta}}_r + \bar{\bar{\mathbf{V}}}_{M-r} \bar{\bar{\mathbf{S}}}_{M-r}^{-1} \bar{\boldsymbol{\beta}}_{N-r} - \bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{S}}}_r^{-1} \bar{\boldsymbol{\beta}}_r \, .$$

$$\hat{\mathbf{p}} - \bar{\mathbf{p}}^{+} = \underset{(M \times M-r)}{\bar{\bar{\mathbf{V}}}_{M-r}} \; \underset{(M-r \times N-r)}{\bar{\bar{\mathbf{S}}}_{M-r}^{-1}} \; \underset{(N-r \times 1)}{\bar{\boldsymbol{\beta}}_{N-r}}$$

Note que considerando dados exatos, a diferença entre o vetor de parâmetros exatos via MQ ($\hat{\mathbf{p}}$) e o o vetor de parâmetros exatos via IG ($\bar{\mathbf{p}}^{+}$) envolve apenas a informação contida no espaço M-r i.e., espaço não iluminado.

<u>Analisando o estimador Inversa Generalizada para dados com ruído em comparação ao estimador Ridge Regression (Tikhonov de ordem zero)</u>

Já definimos que o estimador IG (equação 25) é dado por

$$\bar{\mathbf{p}}^{+} = \sum_{\mathrm{i}=1}^{\mathrm{r}} \frac{\beta_i^{Exato}}{S_i} \bar{\mathbf{v}}_i + \sum_{\mathrm{i}=1}^{\mathrm{r}} \frac{\beta_i^{Ruido}}{S_i} \bar{\mathbf{v}}_i \qquad (27)$$

Vimos no tópico 18 que o estimador do Ridge Regression (Regularizador de Tikhonov de ordem zero) é dado por

$$\bar{\mathbf{p}}^{*} = \sum_{\mathrm{i}=1}^{\mathrm{M}} \frac{S_i}{S_i^2 + k} \beta_i^{Exato} \bar{\mathbf{v}}_i + \sum_{\mathrm{i}=1}^{\mathrm{M}} \frac{S_i}{S_i^2 + k} \beta_i^{Ruido} \bar{\mathbf{v}}_i \qquad (28)$$

Note que no estimador IG (equação 27) os somatórios envolvem o espaço r dimensional, portanto o espaço M-r dimensional é totalmente eliminado. Já no estimador o Ridge Regression (Regularizador de Tikhonov de ordem zero) [equação (28)] os somatórios envolvem todo o espaço M dimensional.

No IG os valores singulares pertencentes ao espaço M-r são todos eliminados. Portanto, para alcançar a estabilidade da solução estimada há a eliminação de M-r valores singulares próximos de zero. Já no estimador o Ridge Regression (Regularizador de Tikhonov de ordem zero) todos os M valores singulares são utilizados porém há uma distorção de todos os valores singulares (tanto daqueles valores singulares pequenos, pertencentes ao espaço M-r dimensional, quanto dos valores singulares grandes, pertencentes ao espaço r dimensional)

Em resumo, o estimador IG elimina todos os valores singulares pequenos, pertencentes ao espaço M-r dimensional (espaço não iluminado) e preserva todos os valores singulares pertencentes ao espaço r dimensional (espaço iluminado). Ao contrário o estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) distorce (modifica) todos os M valores singulares através do uso de um valor positivo para a variável k [equação (28)]. Esta distorção dos valores singulares é pequena para os valores singulares grandes (pertencentes ao espaço iluminado), por outro lado esta distorção é muito grande para os valores singulares pequenos o que garante a estabilidade da solução.

Significado da informação a priori introduzida pelo estimador IG

A informação a priori introduzida pelo estimador IG é:

$$\bar{\boldsymbol{\alpha}}_{M-r} = \bar{\mathbf{0}}$$

Portanto o IG introduz a informação que a projeção dos elementos do vetor de parâmetros que pertencem ao espaço nulo (M-r) no novo sistema de referência em que os eixos formados pelos autovetores da matriz $\underset{(M \times M-r)}{\bar{\bar{\mathbf{V}}}_{M-r}}$

Em outras palavras o IG introduz a informação que a solução pertencente ao espaço nulo é o vetor nulo

$$\bar{\mathbf{p}}^{null} = \bar{\bar{\mathbf{V}}}_{M-r}\,\bar{\boldsymbol{\alpha}}_{M-r}$$

$$\bar{\mathbf{p}}^{null} = \bar{\mathbf{0}}$$

$$\bar{\bar{\mathbf{V}}}^{T}_{M-r}\;\bar{\mathbf{p}} = \bar{\mathbf{0}}$$

Então o IG impõe que a solução $\bar{\mathbf{p}}^{+}$ é ortogonal ao espaço nulo, uma vez que sua projeção no espaço nulo (espaço M-r) é nulo ($\bar{\boldsymbol{\alpha}}_{\mathrm{M-r}} = \bar{\mathbf{0}}$). Lembre-se que $\bar{\boldsymbol{\alpha}}_{\mathrm{M-r}}$ é a projeção da solução nula na base do espaço nulo.

Critério para a determinação do posto efetivo r* do estimador IG

Vimos que o estimador IG introduz uma tendenciosidade a solução estimada para obtenção de uma solução estável. O posto ótimo **r*** deve ser um equilíbrio entre a tendenciosidade e a estabilidade da solução estimada.

Vamos apresentar alguns critérios para determinação do posto ótimo **r***.

(1) Limite Superior para a Variância do k-ésimo parâmetro

Sob as premissas estatísticas estabelecidas na Figura 2, a variância do k-ésimo parâmetro estimado via IG é

$$\mathrm{var}\left\{ \mathbf{p}^{+}{}_{\mathrm{k}} \right\} = \sigma^{2} \sum_{j=1}^{r} \frac{v_{kj}^{2}}{S_{j}^{2}} \qquad k = 1, \ldots M$$

O posto ótimo **r*** será o maior inteiro que satisfaz a seguinte desigualdade

$$\sigma^{2} \sum_{j=1}^{r} \frac{v_{kj}^{2}}{S_{j}^{2}} \leq \sigma_{\mathrm{sup}}^{2}(k) \qquad (33)$$

em que $\sigma_{\mathrm{sup}}^{2}(k)$ é a máxima variância tolerada para o k-ésimo parâmetro.

Se escolhermos o maior r que satisfaz a desigualdade (33) então isto implicará um menor espaço nulo (M-r) e, conseqüentemente, implicará um menor espaço para ser incorporado informação a priori $\overline{\boldsymbol{\alpha}}_{\mathrm{M-r}} = \overline{\mathbf{0}}$ logo menor será a tendenciosidade.

(1) Limite superior para a razão $\dfrac{S_r}{S_{r+1}}$

Vamos escolher o r como o maior inteiro satisfazendo

$$\frac{s_r}{s_{r+1}} < c$$

em que c é um limite, usualmente, no entorno de 10. Então verificamos se a razão entre os valores singulares sucessivos é menor que c.

(2) Análise da norma do resíduo

$$Q = \left\| \bar{\boldsymbol{\varepsilon}} \right\|_2^2 = \left\| \bar{\mathbf{y}}^{\mathbf{o}} - \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}}^{+} \right\|_2^2$$

Substituindo o estimador IG $\bar{\mathbf{p}}^{+} = \bar{\bar{\mathbf{V}}}_r\, \bar{\bar{\mathbf{S}}}_r^{-1}\, \bar{\bar{\mathbf{U}}}_r^T\, \bar{\mathbf{y}}^o$ na equação acima temos que a norma Euclideana do resíduo é

$$Q = \left\| \bar{\boldsymbol{\varepsilon}} \right\|_2^2 = \left\| \bar{\mathbf{y}}^{\mathbf{o}} - \bar{\bar{\mathbf{A}}}\,\bar{\bar{\mathbf{V}}}_r\, \bar{\bar{\mathbf{S}}}_r^{-1}\, \bar{\bar{\mathbf{U}}}_r^T\, \bar{\mathbf{y}}^o \right\|_2^2$$

Usando a decomposição em valores singulares da matriz $\bar{\bar{\mathbf{A}}} = \bar{\bar{\mathbf{U}}}_{\mathrm{r}}\, \bar{\bar{\mathbf{S}}}_{\mathrm{r}}\, \bar{\bar{\mathbf{V}}}_{\mathrm{r}}^{\mathrm{T}}$, temos

$$Q = \left\| \bar{\boldsymbol{\varepsilon}} \right\|_2^2 = \left\| \bar{\mathbf{y}}^{\mathbf{o}} - \bar{\bar{\mathbf{U}}}_{\mathrm{r}}\, \bar{\bar{\mathbf{S}}}_{\mathrm{r}}\, \bar{\bar{\mathbf{V}}}_{\mathrm{r}}^{\mathrm{T}}\, \bar{\bar{\mathbf{V}}}_r\, \bar{\bar{\mathbf{S}}}_r^{-1}\, \bar{\bar{\mathbf{U}}}_r^T\, \bar{\mathbf{y}}^o \right\|_2^2$$

Como $\bar{\bar{\mathbf{V}}}_{\mathrm{r}}^{\mathrm{T}}\bar{\bar{\mathbf{V}}}_r = \bar{\bar{\mathbf{I}}}_{(r\times r)}$ e $\bar{\bar{\mathbf{S}}}_r^{-1}\, \bar{\bar{\mathbf{S}}}_{\mathrm{r}} = \bar{\bar{\mathbf{I}}}_{(r\times r)}$ temos que

$$Q = \left\| \bar{\boldsymbol{\varepsilon}} \right\|_2^2 = \left\| \bar{\mathbf{y}}^{\mathbf{o}} - \bar{\bar{\mathbf{U}}}_{\mathrm{r}}\, \bar{\bar{\mathbf{U}}}_r^T\, \bar{\mathbf{y}}^o \right\|_2^2$$

$$Q = \left\| \bar{\boldsymbol{\varepsilon}} \right\|_2^2 = \left\| \left(\underset{(N\times N)}{\bar{\bar{\mathbf{I}}}} - \bar{\bar{\mathbf{U}}}_{\mathrm{r}}\, \bar{\bar{\mathbf{U}}}_r^T \right) \bar{\mathbf{y}}^o \right\|_2^2$$

Como $\bar{\bar{\mathbf{U}}}_r\, \bar{\bar{\mathbf{U}}}_r^{\mathbf{T}} + \bar{\bar{\mathbf{U}}}_{N-r}\, \bar{\bar{\mathbf{U}}}_{N-r}^{\mathbf{T}} = \bar{\bar{\mathbf{I}}}_{(N\times N)}$ então temos que $\bar{\bar{\mathbf{U}}}_{N-r}\, \bar{\bar{\mathbf{U}}}_{N-r}^{\mathbf{T}} = \underset{(N\times N)}{\bar{\bar{\mathbf{I}}}} - \bar{\bar{\mathbf{U}}}_r\, \bar{\bar{\mathbf{U}}}_r^{\mathbf{T}}$ logo

$$Q = \left\| \overline{\boldsymbol{\varepsilon}} \right\|_2^2 = \left\| \overline{\overline{\mathbf{U}}}_{\mathrm{N-r}} \; \overline{\overline{\mathbf{U}}}_{N-r}^{T} \; \overline{\mathbf{y}}^{o} \right\|_2^2$$

Como $\overline{\overline{\mathbf{U}}}_{N-r}^{T} \; \overline{\mathbf{y}}^{o} = \overline{\boldsymbol{\beta}}_{N-r}$ então temos que

$$Q = \left\| \overline{\boldsymbol{\varepsilon}} \right\|_2^2 = \left\| \overline{\overline{\mathbf{U}}}_{\mathrm{N-r}} \; \overline{\boldsymbol{\beta}}_{N-r} \right\|_2^2 \tag{33}$$

Veja então que o vetor de resíduos do estimador IG pode ser escrito como

$$\overline{\boldsymbol{\varepsilon}} = \overline{\overline{\mathbf{U}}}_{\mathrm{N-r}} \; \overline{\overline{\mathbf{U}}}_{\mathrm{N-r}}^{\mathrm{T}} \; \overline{\mathbf{y}}^{\mathbf{o}}$$

Logo os resíduos do estimador IG é a projeção dos dados observados no espaço N-r das observações. Matematicamente minimizar a função-objeto (33) significa impor que a projeção dos dados observados no espaço N-r das observações, i.e. $\overline{\boldsymbol{\beta}}_{N-r}$, deve ser aproximadamente zero. Em outras palavras, a minimização de

$$Q = \left\| \overline{\boldsymbol{\varepsilon}} \right\|_2^2 = \left\| \overline{\overline{\mathbf{U}}}_{\mathrm{N-r}} \; \overline{\boldsymbol{\beta}}_{N-r} \right\|_2^2$$

implica $\overline{\boldsymbol{\beta}}_i \approx \overline{0}, i = r+1, r+2, ..., N$, logo estes coeficientes não contribuirão significantemente para o ajuste IG. Então podemos dizer que o ajuste do estimador IG é controlado apenas pela projeção dos dados observados no espaço r das observações, i.e. $\overline{\boldsymbol{\beta}}_r$.

Na prática faremos um gráfico do ajuste Q versus o posto. Assim o posto ótimo **r*** será o primeiro inteiro a partir do qual o resíduo Q não cai substancialmente

Dedução e análise do estimador Inversa Generalizada (IG) via Estimador Genérico:

Vimos que o estimador IG é deduzido do problema matemático:

$$\text{Minimize} \quad \bar{\mathbf{p}}^{\mathrm{T}}\ \bar{\bar{\mathbf{W}}}_{\mathbf{p}}\ \bar{\mathbf{p}}$$

$$\text{Sujeito a} \left(\bar{\bar{\mathbf{A}}}\bar{\mathbf{p}} - \bar{\mathbf{y}}^0\right)^{\mathrm{T}} \bar{\bar{\mathbf{W}}}_{\mathbf{e}} \left(\bar{\bar{\mathbf{A}}}\bar{\mathbf{p}} - \bar{\mathbf{y}}^0\right) = \delta$$

Façamos

$$\bar{\bar{\mathbf{W}}}_{\mathbf{p}} = \bar{\bar{\mathbf{V}}} \begin{bmatrix} \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{0}}} & \bar{\bar{\boldsymbol{\eta}}} \end{bmatrix} \bar{\bar{\mathbf{V}}}^{\mathrm{T}}$$

em que $\bar{\bar{\boldsymbol{\eta}}}$ é uma matriz diagonal, não singular com dimensão $(\mathrm{M} - \mathrm{r} \times \mathrm{M} - \mathrm{r})$ logo

$$\bar{\bar{\mathbf{W}}}_{\mathbf{p}} = \bar{\bar{\mathbf{V}}}_{M-r}\ \bar{\bar{\boldsymbol{\eta}}}\ \bar{\bar{\mathbf{V}}}_{M-r}^{T}$$

e

$$\bar{\bar{\mathbf{W}}}_{\mathbf{e}} = \bar{\bar{\mathbf{U}}} \begin{bmatrix} \bar{\bar{\boldsymbol{\gamma}}} & \bar{\bar{\mathbf{0}}} \\ \bar{\bar{\mathbf{0}}} & \bar{\bar{\mathbf{0}}} \end{bmatrix} \bar{\bar{\mathbf{U}}}^{\mathbf{T}}$$

em que $\bar{\bar{\boldsymbol{\gamma}}}$ é uma matriz diagonal, não singular com dimensão $(\mathrm{r} \times \mathrm{r})$ logo

$$\bar{\bar{\mathbf{W}}}_{\mathbf{e}} = \bar{\bar{\mathbf{U}}}_{r}\ \bar{\bar{\boldsymbol{\gamma}}}\ \bar{\bar{\mathbf{U}}}_{r}^{T}$$

Vimos anteriormente que se r = M o estimador Inversa Generalizada é dado pela equação

$$\bar{\mathbf{p}}^{+} = \bar{\bar{\mathbf{V}}}_{r}\ \bar{\bar{\mathbf{S}}}_{r}^{-1}\ \bar{\bar{\mathbf{U}}}_{r}^{T}\ \bar{\mathbf{y}}^{0} \qquad (3.1)$$

Comparando a equação (3.1) com a equação (3) concluímos que no estimador Inversa Generalizada

$$\overline{\overline{\mathbf{H}}} = \overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{S}}}_r^{-1} \overline{\overline{\mathbf{U}}}_{\mathrm{r}}^{\mathrm{T}} \tag{3.2}$$

$$\overline{\mathbf{p}}^{\mathbf{o}} = \overline{\mathbf{0}} \tag{3.3}$$

$$\overline{\overline{\mathbf{W}}}_y = \overline{\overline{\mathbf{U}}}_r \ \overline{\overline{\boldsymbol{\gamma}}} \ \overline{\overline{\mathbf{U}}}_r^T \tag{3.4}$$

$$\overline{\overline{\mathbf{W}}}_{\mathbf{p}} = \overline{\overline{\mathbf{V}}}_{M-r} \ \overline{\overline{\boldsymbol{\eta}}} \ \overline{\overline{\mathbf{V}}}_{M-r}^T \tag{3.5}$$

$$\mu = 0 \tag{3.6}$$

Logo no estimador Inversa Generalizada temos que

$$\hat{\overline{\mathbf{p}}} = \overline{\overline{\mathbf{H}}} \ \overline{\mathbf{y}}^{\mathbf{o}} \tag{3.7}$$

3.1) Análise Estatística do Estimador Inversa Generalizada Via Estimador Genérico

3.1.1) A esperança de $\overline{\mathbf{p}}^{+}$ sob as premissas estatísticas 1 1 — — — — 1 1

Premissas estatísticas sobre o ruído que contamina os dados geofísicos e outras variáveis para calcularmos:
a Esperança do estimador IG

Esperança :

Premissas estatísticas 1 1 — — — — 1 1

Erro aditivo

Erro com média nula

variáveis independentes sem erros

parâmetros não são aleatórios

Vimos na equação (6) que

$$E\left\{\tilde{\bar{\mathbf{p}}}\right\} = \overline{\overline{\mathbf{H}\,\mathbf{A}}}\,\bar{\mathbf{p}} + \left(\bar{\bar{\mathbf{I}}} - \overline{\overline{\mathbf{H}\,\mathbf{A}}}\right)\bar{\mathbf{p}}^{\mathbf{o}}$$

Como no estimador Inversa Generalizada $\bar{\mathbf{p}}^{\mathbf{o}} = \bar{\mathbf{0}}$ e $\bar{\bar{\mathbf{H}}} = \bar{\bar{\mathbf{V}}}_r\,\bar{\bar{\mathbf{S}}}_r^{-1}\,\bar{\bar{\mathbf{U}}}_r^T$ temos que a equação acima é dada por

$$E\{\bar{\mathbf{p}}^+\} = \bar{\bar{\mathbf{V}}}_r\,\bar{\bar{\mathbf{S}}}_r^{-1}\,\bar{\bar{\mathbf{U}}}_r^T \quad \bar{\bar{\mathbf{A}}}\;\bar{\mathbf{p}}$$

Usando a decomposição em valores singulares da matriz de sensibilidade é igual a $\bar{\bar{\mathbf{A}}} = \bar{\bar{\mathbf{U}}}_{\mathbf{r}}\;\bar{\bar{\mathbf{S}}}_{\mathbf{r}}\;\bar{\bar{\mathbf{V}}}_{\mathbf{r}}^{\mathbf{T}}$ temos que

$$E\{\bar{\mathbf{p}}^+\} = \bar{\bar{\mathbf{V}}}_r\,\bar{\bar{\mathbf{S}}}_r^{-1}\,\bar{\bar{\mathbf{U}}}_r^T \quad \bar{\bar{\mathbf{U}}}_r\;\bar{\bar{\mathbf{S}}}_r\;\bar{\bar{\mathbf{V}}}_r^T\;\bar{\mathbf{p}}$$

$$\boxed{E\{\bar{\mathbf{p}}^+\} = \bar{\bar{\mathbf{V}}}_r\,\bar{\bar{\mathbf{V}}}_r^T\;\bar{\mathbf{p}}} \qquad (3.8)$$

Portanto o estimador Inversa Generalizada é um estimador tendencioso.

Veja que, somente se r = M temos que $\bar{\bar{\mathbf{V}}}_r\,\bar{\bar{\mathbf{V}}}_r^T = \bar{\bar{\mathbf{I}}}$, este caso temos o estimador MQ sobredeterminado em que $E\{\hat{\bar{\mathbf{p}}}\} = \bar{\mathbf{p}}$

3.1.2) A Matriz de Covariância de $\bar{\mathbf{p}}^+$

Sob as premissas estatísticas 1 1 1 1 — — 1 1

Premissas estatísticas sobre o ruído que contamina os dados geofísicos e outras variáveis para calcularmos a

Matriz de covariância do estimador IG

Premissas estatísticas 1 1 1 1 — — 1 1

Erro aditivo

Erro com média nula

Erros com variância constante =σ^2

Erros não correlacionáveis

variáveis independentes sem erros

parâmetros não são aleatórios

Vimos na equação (7) que

$$\mathrm{COV}\left\{\tilde{\mathbf{p}}\right\} = \overline{\overline{\mathbf{H}}}\ \mathrm{cov}\left\{\boldsymbol{\varepsilon}\right\}\ \overline{\overline{\mathbf{H}}}^{\mathbf{T}}$$

Como no estimador Inversa Generalizada e $\overline{\overline{\mathbf{H}}} = \overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{S}}}_r^{-1} \overline{\overline{\mathbf{U}}}_{\mathrm{r}}^{\mathrm{T}}$ temos que a equação acima é

$$\mathrm{COV}\left\{\overline{\mathbf{p}}^{+}\right\} = \overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{S}}}_r^{-1} \overline{\overline{\mathbf{U}}}_{\mathrm{r}}^{\mathrm{T}}\ \mathrm{cov}\left\{\overline{\boldsymbol{\varepsilon}}\right\}\ \overline{\overline{\mathbf{U}}}_r\ \overline{\overline{\mathbf{S}}}_r^{-1}\ \overline{\overline{\mathbf{V}}}_r^{T}$$

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

$$\mathbf{COV}\left\{\overline{\boldsymbol{\varepsilon}}\right\} = \sigma^{2}\,\overline{\overline{\mathbf{I}}}$$

$$\mathrm{COV}\left\{\overline{\mathbf{p}}^{+}\right\} = \sigma^{2}\,\overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{S}}}_r^{-1} \overline{\overline{\mathbf{U}}}_{\mathrm{r}}^{\mathrm{T}}\ \overline{\overline{\mathbf{U}}}_r\ \overline{\overline{\mathbf{S}}}_r^{-1}\ \overline{\overline{\mathbf{V}}}_r^{T}$$

$$\boxed{\mathrm{COV}\left\{\overline{\mathbf{p}}^{+}\right\} = \sigma^{2}\,\overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{S}}}_r^{-2}\ \overline{\overline{\mathbf{V}}}_r^{T}} \qquad (3.9)$$

Veja que a covariância do kj-ésimo elemento da matriz $\mathrm{cov}\left\{\hat{\mathbf{p}}\right\}$ é

$$\mathrm{cov}\left\{\mathbf{p}^{+}{}_{\mathrm{kj}}\right\} = \sigma^{2}\sum_{j=1}^{r}\frac{v_{kj}\,v_{kj}}{S_j^{2}}, \qquad k = 1,\ldots, M$$

Como a variância de $\overline{\mathbf{p}}^{+}$ são os elementos da diagonal da matriz de covariância ($\mathrm{cov}\{\overline{\mathbf{p}}^{+}\}$) então temos que a variância do k-ésimo parâmetro é

$$\mathrm{var}\left\{\hat{\mathbf{p}}_{\mathrm{k}}\right\} = \sigma^{2}\sum_{j=1}^{r}\frac{v_{kj}^{2}}{S_j^{2}} \qquad (3.10)$$

Note que no estimador IG temos também variâncias inversamente proporcionais aos valores singulares da matriz de sensibilidade $\overline{\overline{\mathbf{A}}}$. No entanto, diferentemente do estimadores MQ sobre e MQ sub, vemos que o somatório na equação (3.10) envolve apenas r valores singulares. Então, o estimador Inversa Generalizada permite utilizar apenas os r maiores valores singulares (aqueles r valores singulares que não são próximos a zero).

Esta é a grande vantagem do estimador IG quando comparado aos estimadores:
i) MQ SOBRE cuja matriz de covariância é dada por

$$\text{COV}\ \left\{ \hat{\mathbf{p}} \right\} = \sigma^2 \overline{\overline{\mathbf{V}}}_M \overline{\overline{\mathbf{S}}}_M^{-2} \overline{\overline{\mathbf{V}}}_M^T \qquad \textbf{MQ SOBRE}$$

sendo, portanto, a variância do k-ésimo parâmetro dada por

$$\text{var}\left\{ \hat{\mathbf{p}}_k \right\} = \sigma^2 \sum_{j=1}^{M} \frac{v_{kj}^2}{S_j^{\,2}} \qquad \textbf{MQ SOBRE}$$

e
ii) MQ SUB cuja matriz de covariância é dada por

$$\text{COV}\ \left\{ \hat{\mathbf{p}} \right\} = \sigma^2 \overline{\overline{\mathbf{V}}}_N \overline{\overline{\mathbf{S}}}_N^{-2} \overline{\overline{\mathbf{V}}}_N^T \qquad \textbf{MQ SUB}$$

sendo, portanto, a variância do k-ésimo parâmetro dada por

$$\text{var}\left\{ \hat{\mathbf{p}}_k \right\} = \sigma^2 \sum_{j=1}^{N} \frac{v_{kj}^2}{S_j^{\,2}} \qquad \textbf{MQ SUB}$$

Notamos que no MQ sobre e MQ Sub, obrigatoriamente, as variâncias dos parâmetros envolvem, respectivamente, os M e N valores singulares. Ao contrário, no estimador IG o somatório na equação (3.10) envolve apenas r valores singulares em que r pode ser menor que M (e/ou menor que N). Isto possibilita excluir do somatório da equação (3.10) aqueles M-r valores singulares próximos de zero.

Então quanto menor o valor atribuído a r , maior será a estabilidade introduzida na determinação da solução via IG. O preço pago por isto é a introdução da tendenciosidade na estimativa dos parâmetros que pode ser atestada matematicamente por

$$E\{\overline{\mathbf{p}}^+\} = \overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{V}}}_r^T\ \overline{\mathbf{p}}$$

se $r \neq M$ temos que $\overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{V}}}_r^T \neq \overline{\overline{\mathbf{I}}}$ logo o estimador IG é tendencioso.

2.1.3) A esperança da distância $\left(\overline{\mathbf{p}}^+ - \overline{\mathbf{p}}\right)^{\mathbf{T}} \left(\overline{\mathbf{p}}^+ - \overline{\mathbf{p}}\right)$

Vimos na equação (8) que

$$E\left\{\left\|\tilde{\overline{\mathbf{p}}}-\overline{\mathbf{p}}\right\|_2^2\right\}=tr\{\mathrm{cov}(\tilde{\overline{\mathbf{p}}})\}+\left\|\mathrm{E}\left\{\ \tilde{\overline{\mathbf{p}}}\ \right\}-\overline{\mathbf{p}}\right\|_2^2$$

Na equação (3.9) deduzimos que no estimador MQ sobre

$$\mathrm{COV}\ \left\{\ \overline{\mathbf{p}}^{+}\ \right\}=\sigma^{2}\ \overline{\overline{\mathbf{V}}}_r\,\overline{\overline{\mathbf{S}}}_r^{\,-2}\ \overline{\overline{\mathbf{V}}}_r^{T}$$

e n a equação (3.8) deduzimos que

$$E\{\overline{\mathbf{p}}^{+}\}=\overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{V}}}_r^{T}\ \overline{\mathbf{p}}$$

Logo temos que no estimador Inversa Generalizada a esperança da distância $\left(\overline{\mathbf{p}}^{+}-\overline{\mathbf{p}}\right)^{\mathbf{T}}\left(\overline{\mathbf{p}}^{+}-\overline{\mathbf{p}}\right)$ é dada por

$$E\{[\overline{\mathbf{p}}^{+}-\overline{\mathbf{p}}]^{\mathrm{T}}[\overline{\mathbf{p}}^{+}-\overline{\mathbf{p}}]\}=\mathrm{tr}\left\{\sigma^{2}\ \overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{S}}}_r^{\,-2}\ \overline{\overline{\mathbf{V}}}_r^{T}\ \right\}+\left\|\overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{V}}}_r^{T}\ \overline{\mathbf{p}}-\overline{\mathbf{p}}\right\|_2^2$$

$$E\{[\overline{\mathbf{p}}^{+}-\overline{\mathbf{p}}]^{\mathrm{T}}[\overline{\mathbf{p}}^{+}-\overline{\mathbf{p}}]\}=\mathrm{tr}\left\{\sigma^{2}\ \overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{S}}}_r^{\,-2}\ \overline{\overline{\mathbf{V}}}_r^{T}\ \right\}+\left\|\left(\overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{V}}}_r^{T}\ \text{-}\ \overline{\overline{\mathbf{I}}}\right)\overline{\mathbf{p}}\right\|_2^2$$

Lembrando da propriedade do traço do produto de matrizes

$$tr\left\{\overline{\overline{\mathbf{C}}}\,\overline{\overline{\mathbf{B}}}\right\}=tr\left\{\overline{\overline{\mathbf{B}}}\,\overline{\overline{\mathbf{C}}}\right\}$$

Então podemos escrever que

$$E\{[\overline{\mathbf{p}}^{+}-\overline{\mathbf{p}}]^{\mathrm{T}}[\overline{\mathbf{p}}^{+}-\overline{\mathbf{p}}]\}=\mathrm{tr}\left\{\sigma^{2}\ \overline{\overline{\mathbf{S}}}_r^{\,-2}\ \overline{\overline{\mathbf{V}}}_r^{T}\ \overline{\overline{\mathbf{V}}}_r\right\}+\left\|\left(\overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{V}}}_r^{T}\ \text{-}\ \overline{\overline{\mathbf{I}}}\right)\overline{\mathbf{p}}\right\|_2^2$$

$$E\{[\overline{\mathbf{p}}^{+}-\overline{\mathbf{p}}]^{\mathrm{T}}[\overline{\mathbf{p}}^{+}-\overline{\mathbf{p}}]\}=\mathrm{tr}\left\{\sigma^{2}\ \overline{\overline{\mathbf{S}}}_r^{\,-2}\ \right\}+\left\|\left(\overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{V}}}_r^{T}\ \text{-}\ \overline{\overline{\mathbf{I}}}\right)\overline{\mathbf{p}}\right\|_2^2$$

Como no caso Inversa Generalizada r < M então temos que

$$\overline{\overline{\mathbf{V}}}_{\mathrm{M}}\overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}=\overline{\overline{\mathbf{V}}}_{\mathrm{r}}\overline{\overline{\mathbf{V}}}_{\mathrm{r}}^{\mathrm{T}}+\overline{\overline{\mathbf{V}}}_{\mathrm{M-r}}\overline{\overline{\mathbf{V}}}_{\mathrm{M-r}}^{\mathrm{T}}=\overline{\overline{\mathbf{I}}}_{M}$$

então temos que $\overline{\overline{\mathbf{V}}}_{\mathrm{r}}\overline{\overline{\mathbf{V}}}_{\mathrm{r}}^{\mathrm{T}}-\overline{\overline{\mathbf{I}}}_{M}=-\overline{\overline{\mathbf{V}}}_{\mathrm{M-r}}\overline{\overline{\mathbf{V}}}_{\mathrm{M-r}}^{\mathrm{T}}$

$$E\{[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]^{\mathrm{T}}[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]\} = \mathrm{tr}\left\{\sigma^2\, \overline{\overline{\mathbf{S}}}_r^{\,-2}\right\} + \left\| \left(-\overline{\overline{\mathbf{V}}}_{\text{M-r}} \overline{\overline{\mathbf{V}}}_{\text{M-r}}^{\mathrm{T}} \right) \bar{\mathbf{p}} \right\|_2^2$$

$$E\{[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]^{\mathrm{T}}[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]\} = \mathrm{tr}\left\{\sigma^2\, \overline{\overline{\mathbf{S}}}_r^{\,-2}\right\} + \bar{\mathbf{p}}^T \left(-\overline{\overline{\mathbf{V}}}_{\text{M-r}} \overline{\overline{\mathbf{V}}}_{\text{M-r}}^{\mathrm{T}} \right)^{\mathrm{T}} \left(-\overline{\overline{\mathbf{V}}}_{\text{M-r}} \overline{\overline{\mathbf{V}}}_{\text{M-r}}^{\mathrm{T}} \right) \bar{\mathbf{p}}$$

$$E\{[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]^{\mathrm{T}}[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]\} = \mathrm{tr}\left\{\sigma^2\, \overline{\overline{\mathbf{S}}}_r^{\,-2}\right\} + \bar{\mathbf{p}}^T \left(-\overline{\overline{\mathbf{V}}}_{\text{M-r}} \overline{\overline{\mathbf{V}}}_{\text{M-r}}^{\mathrm{T}} \right) \left(-\overline{\overline{\mathbf{V}}}_{\text{M-r}} \overline{\overline{\mathbf{V}}}_{\text{M-r}}^{\mathrm{T}} \right) \bar{\mathbf{p}}$$

$$E\{[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]^{\mathrm{T}}[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]\} = \mathrm{tr}\left\{\sigma^2\, \overline{\overline{\mathbf{S}}}_r^{\,-2}\right\} + \bar{\mathbf{p}}^T\, \overline{\overline{\mathbf{V}}}_{\text{M-r}} \overline{\overline{\mathbf{V}}}_{\text{M-r}}^{\mathrm{T}} \overline{\overline{\mathbf{V}}}_{\text{M-r}} \overline{\overline{\mathbf{V}}}_{\text{M-r}}^{\mathrm{T}}\, \bar{\mathbf{p}}$$

$$E\{[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]^{\mathrm{T}}[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]\} = \mathrm{tr}\left\{\sigma^2\, \overline{\overline{\mathbf{S}}}_r^{\,-2}\right\} + \bar{\mathbf{p}}^T\, \overline{\overline{\mathbf{V}}}_{\text{M-r}} \overline{\overline{\mathbf{V}}}_{\text{M-r}}^{\mathrm{T}}\, \bar{\mathbf{p}}$$

Chamando $\overline{\overline{\mathbf{V}}}_{\text{M - r}}^{\mathrm{T}} \bar{\mathbf{p}} = \bar{\boldsymbol{\alpha}}_{M-r}$ temos que

$$E\{[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]^{\mathrm{T}}[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]\} = \sigma^2 \;\mathrm{tr}\left\{\overline{\overline{\mathbf{S}}}_r^{\,-2}\right\} + \bar{\boldsymbol{\alpha}}_{M-r}^T\, \bar{\boldsymbol{\alpha}}_{M-r}$$

Com a decomposição em valores singulares da matriz de sensibilidade é igual a $\overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{U}}}_{\mathbf{r}}\ \overline{\overline{\mathbf{S}}}_{\mathbf{r}}\ \overline{\overline{\mathbf{V}}}_{\mathbf{r}}^{\mathbf{T}}$ em que r = N, temos que esperança da distância $\left(\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}\right)^{\mathbf{T}}\left(\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}\right)$ do estimador Inversa Generalizada é dado por

$$\boxed{E\{[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]^{\mathrm{T}}[\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}]\} = \sigma^2 \sum_{\mathrm{i}=1}^{\mathrm{r}} \frac{1}{S_i^2} + \sum_{\mathrm{i}=r+1}^{\mathrm{M}} \alpha_{\mathrm{i}}^2} \qquad (3.11)$$

$$E\left\{\left\|\bar{\mathbf{p}}^+ - \bar{\mathbf{p}}\right\|_2^2\right\} = \sigma^2 \frac{1}{S_1^2} + \sigma^2 \frac{1}{S_2^2} + \ldots + \sigma^2 \frac{1}{S_r^2} + \alpha_{\mathrm{r}+1}^2 + \alpha_{\mathrm{r}+2}^2 + \ldots + \alpha_{\mathrm{M}}^2$$

em que $S_i, i = 1,2,\ldots,r$ são os r maiores valores singulares da matriz de sensibilidade $\overline{\overline{\mathbf{A}}}$.

Note que quanto menor o valor de r maior é o espaço M – r e conseqüentemente maior é a dimensão do vetor $\bar{\boldsymbol{\alpha}}_{M-r}$. Se aumentarmos o valor de r menor será o espaço M-r e menor a dimensão do vetor $\bar{\boldsymbol{\alpha}}_{M-r}$.

Então veja o primeiro termo da Esperança da distância $\left\|\bar{\mathbf{p}}^{+} - \bar{\mathbf{p}}\right\|_2^2$, equação (3.11), é formada por duas parcelas:

1) $tr\,\{\mathrm{cov}(\ \bar{\mathbf{p}}^{+})\} = \sigma^2 \sum_{i=1}^{r} \frac{1}{S_i^2}$ é a primeira parcela e representa uma medida de covariância do estimador , ou seja, uma medida da estabilidade da solução estimada. Esta parcela é uma função crescente monotônica de r, i.e., quanto maior r maior será o valor desta primeira parcela e, por outro lado, quanto menor r menor será o valor desta primeira parcela.

2) $\left\| \mathrm{E}\left\{\bar{\mathbf{p}}^{+}\right\} - \bar{\mathbf{p}}\right\|_2^2 = \sum_{i=r+1}^{M} \alpha_i^2$ é a segunda parcela representa uma medida da distância da esperança do estimador ($\bar{\mathbf{p}}^{+}$) e o vetor de parâmetros ($\bar{\mathbf{p}}$). Veja portanto que esta segunda parcela é uma medida da tendensiosidade do estimador $\bar{\mathbf{p}}^{+}$. Esta parcela é uma função decrescente monotônica de r, i.e., quanto maior r menor será o valor desta segunda parcela e, por outro lado, quanto menor r maior será o valor desta segunda parcela.

Graficamente a esperança da distância $\left(\bar{\mathbf{p}}^{+} - \bar{\mathbf{p}}\right)^{\mathbf{T}}\left(\bar{\mathbf{p}}^{+} - \bar{\mathbf{p}}\right)$ do estimador Inversa Generalizada pode ser simplificada pela Figura 1 em que r* é o posto “ótimo”, i.e., um trade-off entre a tendenciosidade e a estabilidade da solução. Note que se r = M temos a $E\left\{\left\|\bar{\mathbf{p}}^{+} - \bar{\mathbf{p}}\right\|_2^2\right\}$ do estimador de MQ SOBRE (i.e., solução se tendeciosidade porém com variância máxima).

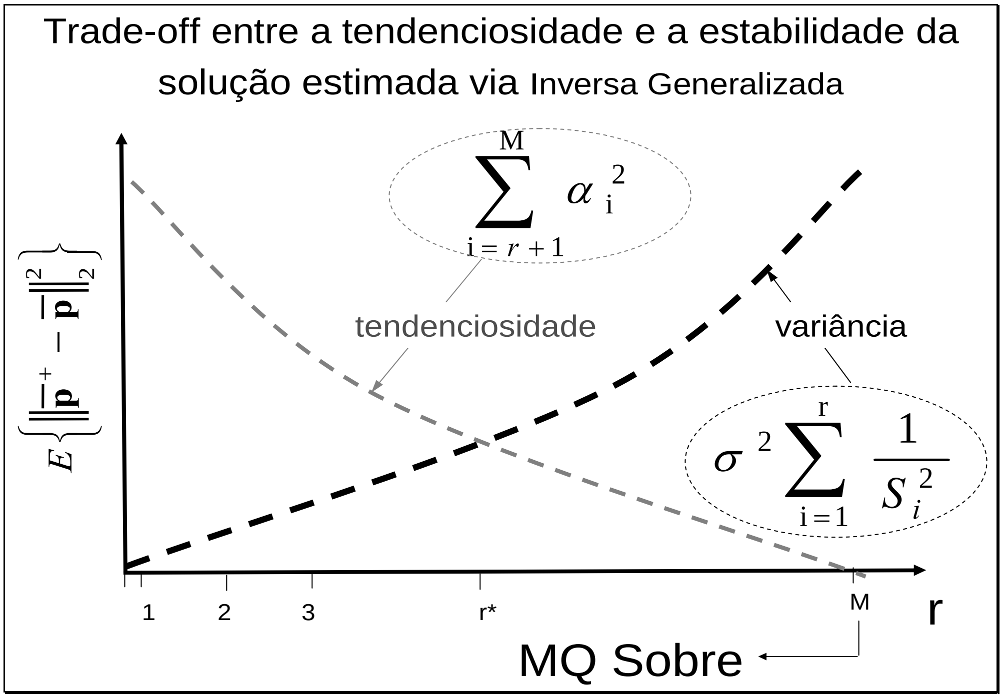

Figura 3.1

1.1.4) A matriz de resolução associada a $\overline{\mathbf{p}}^{+}$

Vimos na equação (9) que

$$\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{H}}}\;\overline{\overline{\mathbf{A}}}$$

Como no estimador Inversa Generalizada $\overline{\overline{\mathbf{H}}} = \overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{S}}}_r^{-1}\overline{\overline{\mathbf{U}}}_r^{\mathrm{T}}$ temos que a equação acima é dada por

$$\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{S}}}_r^{-1}\overline{\overline{\mathbf{U}}}_r^{T}\;\overline{\overline{\mathbf{A}}}$$

Usando a decomposição em valores singulares da matriz de sensibilidade é igual a $\overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{U}}}_{\mathbf{r}}\;\overline{\overline{\mathbf{S}}}_{\mathbf{r}}\;\overline{\overline{\mathbf{V}}}_{\mathbf{r}}^{\mathbf{T}}$ temos que

$$\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{S}}}_r^{-1}\overline{\overline{\mathbf{U}}}_r^{T}\;\overline{\overline{\mathbf{U}}}_r\;\overline{\overline{\mathbf{S}}}_r\;\overline{\overline{\mathbf{V}}}_r^{T}$$

$$\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_r\overline{\overline{\mathbf{V}}}_r^{T} \tag{3.12}$$

Veja que a matriz de resolução estimador Inversa Generalizada NÃO é igual a matriz identidade, portanto não é a máxima resolução a menos que r = M porque, neste caso, $\overline{\overline{\mathbf{V}}}_{\mathrm{M}}\overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} = \overline{\overline{\mathbf{I}}}_M$ e $\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{I}}}_M$. Como em geral r < M, o produto $\overline{\overline{\mathbf{V}}}_{\mathrm{r}}\overline{\overline{\mathbf{V}}}_{\mathrm{r}}^{\mathrm{T}} \neq \overline{\overline{\mathbf{I}}}_M$.

Isto significa dizer que cada elemento NÃO é estimado de modo ÚNICO (i.e., $p_j^+, j = 1,2,...,M$ NÃO são estimados independentemente). Dizemos que cada elemento de $\overline{\mathbf{p}}^+$ NÃO é perfeitamente resolvido. Em outras palavras, se a matriz de resolução estimador Inversa Generalizada é diferente da a matriz identidade, isto significa dizer que o j-ésimo parâmetro estimado via IG é obtido como uma combinação linear de TODOS os parâmetros. Logo no estimador Inversa Generalizada a resolução não é máxima.

Em resumo no estimador IG em geral r < M então $\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{V}}}_r^T \neq \overline{\overline{\mathbf{I}}}_M$ então o aspecto da i-ésima linha da matriz de resolução passa de uma delta de Dirac [caso em que r = M e $\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_M \overline{\overline{\mathbf{V}}}_M^T = \overline{\overline{\mathbf{I}}}_M$, Figura 3.2a] para uma função sinc discreta (Figura 3.2b)

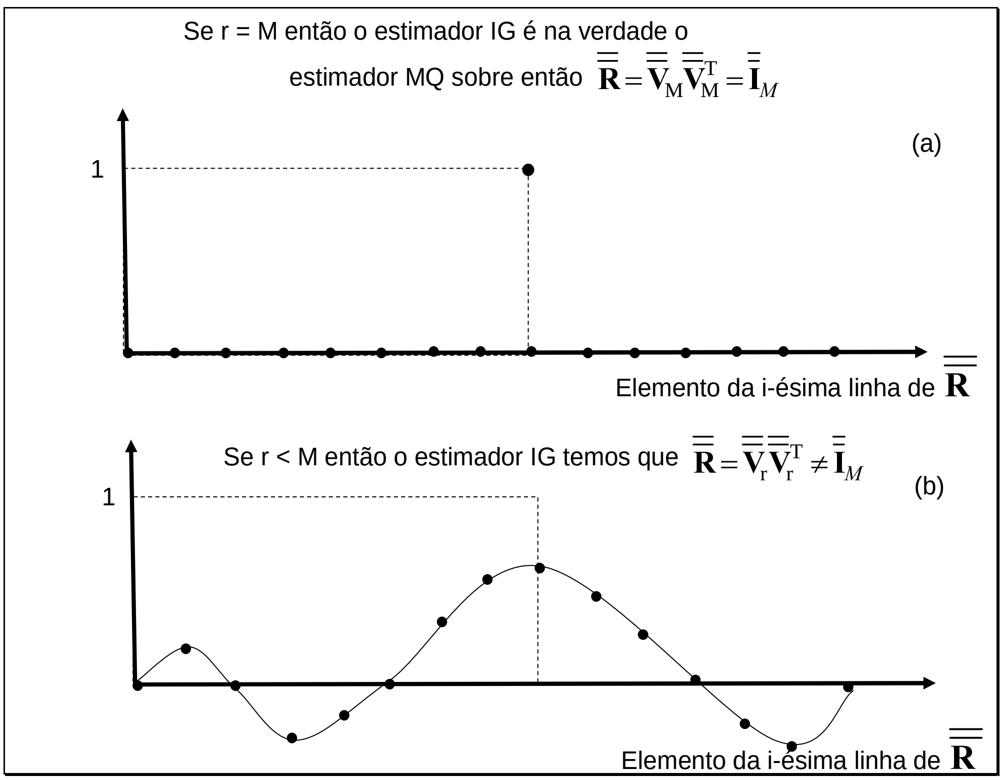

Figura 3.2

Assim $p_i^+ = \overline{\mathbf{r}}_i^T \, \overline{\mathbf{p}}$ mostra que neste caso em que r < M a estimativa IG do i-ésimo parâmetro é obtida pela combinação linear de todos os parâmetros com uma maior contribuição do i-ésimo parâmetro verdadeiro. Quanto menor for o valor de r em relação a M, mais achatada será a função sinc e maior a contribuição relativa de outros parâmetros em relação ao i-ésimo parâmetro verdadeiro estimado p_i^+

Note que adicionalmente que

$$\bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{R}}}\,\bar{\mathbf{p}}$$

$$\bar{\mathbf{p}}^+ = \bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{V}}}_r^T \bar{\mathbf{p}}$$

Como temos que $\bar{\bar{\mathbf{V}}}_M \bar{\bar{\mathbf{V}}}_M^T = \bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{V}}}_r^T + \bar{\bar{\mathbf{V}}}_{M-r} \bar{\bar{\mathbf{V}}}_{M-r}^T = \bar{\bar{\mathbf{I}}}_M$ então temos que

$$\bar{\bar{\mathbf{V}}}_r \bar{\bar{\mathbf{V}}}_r^T = \bar{\bar{\mathbf{I}}}_M - \bar{\bar{\mathbf{V}}}_{M-r} \bar{\bar{\mathbf{V}}}_{M-r}^T$$

logo

$$\bar{\mathbf{p}}^+ = \left(\bar{\bar{\mathbf{I}}}_M - \bar{\bar{\mathbf{V}}}_{M-r} \bar{\bar{\mathbf{V}}}_{M-r}^T \right) \bar{\mathbf{p}}$$

$$\bar{\mathbf{p}}^+ = \bar{\mathbf{p}} - \bar{\bar{\mathbf{V}}}_{M-r} \bar{\bar{\mathbf{V}}}_{M-r}^T \bar{\mathbf{p}}$$

Chamando $\bar{\bar{\mathbf{V}}}_{M-r}^T \bar{\mathbf{p}} = \bar{\boldsymbol{\alpha}}_{M-r}$ temos que

$$\bar{\mathbf{p}}^+ = \bar{\mathbf{p}} - \bar{\bar{\mathbf{V}}}_{M-r} \bar{\boldsymbol{\alpha}}_{M-r}$$

Considerando que a solução IG excluiu todos os M – r valores singulares iguais a zero, então $\bar{\bar{\mathbf{V}}}_{M-r} \bar{\boldsymbol{\alpha}}_{M-r} = \bar{\mathbf{p}}^{null}$ expresso como

$$\bar{\mathbf{p}}^+ = \bar{\mathbf{p}} - \bar{\mathbf{p}}^{Null} \tag{3.13}$$

O resultado mostrado pela equação (3.13) diz que o estimador Inversa Generalizada, se forem excluídos os M-r valores singulares iguais a zero, é uma solução que nunca contém qualquer solução nula ($\bar{\mathbf{p}}^{null}$)

3.1.5) A matriz densidade de informação associada a $\bar{\mathbf{p}}^{+}$

Vimos na equação (10) que

$$\underset{(N \times N)}{\overline{\overline{\mathbf{F}}}} = \overline{\overline{\mathbf{A}}}\ \overline{\overline{\mathbf{H}}}$$

Como no estimador Inversa Generalizada $\overline{\overline{\mathbf{H}}} = \overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{S}}}_r^{-1} \overline{\overline{\mathbf{U}}}_r^{\mathrm{T}}$ temos que a equação acima é dada por

$$\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{A}}}\ \overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{S}}}_r^{-1} \overline{\overline{\mathbf{U}}}_r^{\mathrm{T}}$$

Usando a decomposição em valores singulares da matriz de sensibilidade é igual a $\overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{U}}}_{\mathbf{r}}\ \overline{\overline{\mathbf{S}}}_{\mathbf{r}}\ \overline{\overline{\mathbf{V}}}_{\mathbf{r}}^{\mathbf{T}}$ temos que

$$\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{U}}}_{\mathrm{r}}\ \overline{\overline{\mathbf{S}}}_r\ \overline{\overline{\mathbf{V}}}_r^{T}\ \overline{\overline{\mathbf{V}}}_r \overline{\overline{\mathbf{S}}}_r^{-1} \overline{\overline{\mathbf{U}}}_{\mathrm{r}}^{\mathrm{T}}$$

$$\boxed{\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{U}}}_{\mathrm{r}}\ \overline{\overline{\mathbf{U}}}_{\mathrm{r}}^{\mathrm{T}}} \qquad (3.14)$$

A equação (3.14) mostra que a matriz densidade de informação do estimador Inversa Generalizada só será igual a matriz identidade se r = N. Se r < N a matriz $\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{U}}}_{\mathrm{r}}\ \overline{\overline{\mathbf{U}}}_{\mathrm{r}}^{\mathrm{T}} \neq \overline{\overline{\mathbf{I}}}_N$. Isto quer dizer que se r < N os dados observados e os dados ajustados via estimador Inversa Generalizada é diferente de zero ($\bar{\boldsymbol{\varepsilon}} \neq \bar{\mathbf{0}}$). Conclusão, no estimador Inversa Generalizada o ajuste em geral NÃO EXATO (se r < N).