ANÁLISE DO ESTIMADOR RIDGE REGRESSION (REGULARIZADOR DE TIKHONOV DE ORDEM ZERO) VIA ESTIMADOR GENÉRICO

Apresentaremos neste tópico a dedução dos estimador REGULARIZADOR DE TIKHONOV DE ORDEM ZERO através do estimador genérico (Tópico 20). Em seguida apresentaremos as medidas de eficiências, i.e. Esperança, Covariância. Matriz de Resolução, Matriz Dendidade de Informação, deste estimador a partir das equações deduzidas no tópico 21.

REVISÃO:

Obtemos, no Tópico 20 as duas equações dos estimadores gerais $\tilde{\bar{\mathbf{p}}}$ que chamamos de **estimador genérico** são:

$$\tilde{\bar{\mathbf{p}}} = \bar{\mathbf{p}}^{\mathbf{o}} + \left(\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{W}}}_y\bar{\bar{\mathbf{A}}} + \mu(\delta)\bar{\bar{\mathbf{W}}}_p\right)^{-1}\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{W}}}_y\left(\bar{\mathbf{y}}^{\mathbf{o}} - \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}}^{\mathbf{o}}\right) \tag{1}$$

e

$$\tilde{\bar{\mathbf{p}}} = \bar{\mathbf{p}}^{\mathbf{o}} + \bar{\bar{\mathbf{W}}}_p^{-1}\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\left[\bar{\bar{\mathbf{A}}}\bar{\bar{\mathbf{W}}}_p^{-1}\bar{\bar{\mathbf{A}}}^{\mathrm{T}} + \mu(\delta)\bar{\bar{\mathbf{W}}}_y^{-1}\right]^{-1}\left(\bar{\mathbf{y}}^{\mathbf{o}} - \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}}^{\mathbf{o}}\right) \tag{2}$$

Note que os estimadores genérico (1) e (2) podem ser escritos na forma geral

$$\boxed{\tilde{\bar{\mathbf{p}}} = \bar{\mathbf{p}}^{\mathbf{o}} + \bar{\bar{\mathbf{H}}}\left(\bar{\mathbf{y}}^{\mathbf{o}} - \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}}^{\mathbf{o}}\right)} \tag{3}$$

com

$$\bar{\bar{\mathbf{H}}} = \left(\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{W}}}_y\bar{\bar{\mathbf{A}}} + \mu(\delta)\bar{\bar{\mathbf{W}}}_p\right)^{-1}\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{W}}}_y \tag{4}$$

no caso do estimador (1) e com

$$\bar{\bar{\mathbf{H}}} = \bar{\bar{\mathbf{W}}}_p^{-1}\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\left[\bar{\bar{\mathbf{A}}}\bar{\bar{\mathbf{W}}}_p^{-1}\bar{\bar{\mathbf{A}}}^{\mathrm{T}} + \mu(\delta)\bar{\bar{\mathbf{W}}}_y^{-1}\right]^{-1} \tag{5}$$

no caso do estimador (2) .

Posteriormente, deduzimos no Tópico 21 expressões genéricas `dos medidores de eficiência e desempenho dos estimadores $\tilde{\bar{\mathbf{p}}}$.

Especificamente estes medidores gerais são

1) A esperança de $\tilde{\bar{\mathbf{p}}}$;

$$\boxed{E\left\{\tilde{\bar{\mathbf{p}}}\right\} = \bar{\bar{\mathbf{H}}}\,\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} + \left(\bar{\bar{\mathbf{I}}} - \bar{\bar{\mathbf{H}}}\,\bar{\bar{\mathbf{A}}}\right)\bar{\mathbf{p}}^{\mathbf{o}}} \qquad (6)$$

2) A matriz de covariância de $\tilde{\bar{\mathbf{p}}}$;

$$\boxed{\mathrm{COV}\left\{\tilde{\bar{\mathbf{p}}}\right\} = \bar{\bar{\mathbf{H}}}\ \mathrm{cov}\left\{\bar{\boldsymbol{\varepsilon}}\right\}\ \bar{\bar{\mathbf{H}}}^{\mathbf{T}}} \qquad (7)$$

3) A esperança da distância $\left(\tilde{\bar{\mathbf{p}}} - \bar{\mathbf{p}}\right)^{\mathbf{T}}\left(\tilde{\bar{\mathbf{p}}} - \bar{\mathbf{p}}\right)$

$$\boxed{E\left\{\left\|\tilde{\bar{\mathbf{p}}} - \bar{\mathbf{p}}\right\|_2^2\right\} = tr\left\{\mathrm{cov}(\tilde{\bar{\mathbf{p}}})\right\} + \left\|E\left\{\tilde{\bar{\mathbf{p}}}\right\} - \bar{\mathbf{p}}\right\|_2^2} \qquad (8)$$

4) A matriz de resolução associada a $\tilde{\bar{\mathbf{p}}}$; e

$$\boxed{\underset{(M\ \times\ M)}{\bar{\bar{\mathbf{R}}}} = \bar{\bar{\mathbf{H}}}\ \bar{\bar{\mathbf{A}}}} \qquad (9)$$

5) A matriz de densidade de informação associada a $\tilde{\bar{\mathbf{p}}}$;

$$\boxed{\underset{(N \times N)}{\bar{\bar{\mathbf{F}}}} = \bar{\bar{\mathbf{A}}}\ \bar{\bar{\mathbf{H}}}} \qquad (10)$$

ANÁLISE DO ESTIMADOR RIDGE REGRESSION (REGULARIZADOR DE TIKHONOV DE ORDEM ZERO) VIA ESTIMADOR GENÉRICO

Vimos anteriormente que o estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) é formulado como:

$$\begin{cases} \min \ \left\| \bar{\mathbf{p}} \right\|_2^2 \\ \text{sujeito a}: \ \left\| \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} - \bar{\mathbf{y}}^{\mathbf{o}} \right\|_2^2 = \delta. \end{cases} \tag{4.1}$$

em que δ é o erro médio quadrático (EMQ) da seqüência de realizações da variável aleatória que contamina as observações.

A solução do problema vinculado (4.1) é o **Estimador Ridge Regression (Regularizador de Tikhonov de Ordem Zero)**

$$\bar{\mathbf{p}}^* = \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{A}}} + k\,\bar{\bar{\mathbf{I}}} \right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\mathbf{y}}^0 \tag{4.2}$$

Comparando a equação (4.2) com a equação (3) concluímos que no estimador Ridge Regression (Regularizador de Tikhonov de ordem zero)

$$\bar{\bar{\mathbf{H}}} = \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{A}}} + k\,\bar{\bar{\mathbf{I}}} \right)^{-1} \bar{\bar{\mathbf{A}}}^T \tag{4.3}$$

$$\bar{\mathbf{p}}^{\mathbf{o}} = \bar{\mathbf{0}} \tag{4.4}$$

$$\bar{\bar{\mathbf{W}}}_y = \bar{\bar{\mathbf{I}}} \tag{4.5}$$

$$\bar{\bar{\mathbf{W}}}_p = \bar{\bar{\mathbf{I}}} \tag{4.6}$$

$$\mu = \mathrm{k} \tag{4.7}$$

Logo no estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) temos que

$$\bar{\mathbf{p}}^* = \bar{\bar{\mathbf{H}}}\ \bar{\mathbf{y}}^{\mathbf{o}} \tag{4.8}$$

4.1) Análise Estatística do Estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) Via Estimador Genérico

4.1.1) A esperança de $\bar{\mathbf{p}}^*$ sob as premissas estatísticas 1 1 — — — — 1 1

Vimos na equação (6) que

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

Como no estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) $\bar{\mathbf{p}}^{\mathbf{o}} = \bar{\mathbf{0}}$ e $\bar{\bar{\mathbf{H}}} = \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{A}}} + k\bar{\bar{\mathbf{I}}} \right)^{-1} \bar{\bar{\mathbf{A}}}^T$ temos que a equação acima é dada por

$$\boxed{\mathrm{E}\left\{ \bar{\mathbf{p}}^* \right\} = \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{A}}} + k\bar{\bar{\mathbf{I}}} \right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}}} \qquad (4.9)$$

Portanto o estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) é um estimador tendencioso. O que era de se esperar porque este estimador introduz a tendenciosidade através da minimização do funcional $\phi\,(\bar{\mathbf{p}}) = \left\| \bar{\mathbf{p}} \right\|_2^2$. Portanto, matematicamente, o estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) introduz o "vício" de mínima norma Euclideana do vetor de parâmetros.

Caso Particular: Note que no caso particular em que k = 0 o estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) é torna-se o estimador MQ SOBRE.

Sob a condição M = r, então fazendo a decomposição em valores singulares da matriz de sensibilidade temos que

$$\bar{\bar{\mathbf{A}}}^{\mathrm{T}} \bar{\bar{\mathbf{A}}} = \bar{\bar{\mathbf{V}}}_{\mathrm{M}}\, \bar{\bar{\mathbf{S}}}_{\mathrm{M}}\, \bar{\bar{\mathbf{U}}}_{\mathrm{M}}^{\mathrm{T}}\, \bar{\bar{\mathbf{U}}}_{\mathrm{M}}\, \bar{\bar{\mathbf{S}}}_{\mathrm{M}}\, \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} = \bar{\bar{\mathbf{V}}}_{\mathrm{M}}\, \bar{\bar{\mathbf{S}}}_{\mathrm{M}}^{2}\, \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

Usando então a decomposição em valores singulares da matriz de sensibilidade temos que a esperança do estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) é

$$E\left\{\,\overline{\mathbf{p}}^{*}\right\}=\left(\overline{\overline{\mathbf{A}}}^{T}\overline{\overline{\mathbf{A}}}+k\overline{\overline{\mathbf{I}}}\right)^{-1}\overline{\overline{\mathbf{A}}}^{T}\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}}$$

$$E\left\{\,\overline{\mathbf{p}}^{*}\right\}=\left(\overline{\overline{\mathbf{V}}}_{M}\overline{\overline{\mathbf{S}}}_{M}^{2}\overline{\overline{\mathbf{V}}}_{M}^{\mathrm{T}}+k\overline{\overline{\mathbf{I}}}\right)^{-1}\overline{\overline{\mathbf{V}}}_{M}\overline{\overline{\mathbf{S}}}_{M}^{2}\overline{\overline{\mathbf{V}}}_{M}^{\mathrm{T}}\overline{\mathbf{p}}$$

Se o posto da matriz $\overline{\overline{\mathbf{A}}}$ é igual ao número de parâmetros (i.e., se r = M) temos que $\overline{\overline{\mathbf{V}}}_{M}^{-1}=\overline{\overline{\mathbf{V}}}_{M}^{T}$. Como a inversa do termo $\left(\overline{\overline{\mathbf{V}}}_{M}\overline{\overline{\mathbf{S}}}_{M}^{2}\right)^{-1}=\overline{\overline{\mathbf{S}}}_{M}^{-2}\overline{\overline{\mathbf{V}}}_{M}^{T}$ então este termo pode ser introduzindo pré-multiplicando a matriz a ser invertida da equação acima o que resulta na expressão

$$E\left\{\overline{\mathbf{p}}^{*}\right\}=\left[\left(\overline{\overline{\mathbf{S}}}^{-2}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\right)\left(\overline{\overline{\mathbf{V}}}\,\overline{\overline{\mathbf{S}}}^{2}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}+k\overline{\overline{\mathbf{I}}}\right)\right]^{-1}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\overline{\mathbf{p}}$$

$$E\left\{\overline{\mathbf{p}}^{*}\right\}=\left[\overline{\overline{\mathbf{S}}}^{-2}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\overline{\overline{\mathbf{V}}}\,\overline{\overline{\mathbf{S}}}^{2}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}+k\overline{\overline{\mathbf{S}}}^{-2}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\right]^{-1}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\overline{\mathbf{p}}$$

$$E\left\{\overline{\mathbf{p}}^{*}\right\}=\left[\overline{\overline{\mathbf{S}}}^{-2}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\overline{\overline{\mathbf{V}}}\,\overline{\overline{\mathbf{S}}}^{2}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}+k\overline{\overline{\mathbf{S}}}^{-2}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\right]^{-1}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\overline{\mathbf{p}}$$

$$E\left\{\overline{\mathbf{p}}^{*}\right\}=\left[\overline{\overline{\mathbf{V}}}^{\mathrm{T}}+k\overline{\overline{\mathbf{S}}}^{-2}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\right]^{-1}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\overline{\mathbf{p}}$$

$$E\left\{\overline{\mathbf{p}}^{*}\right\}=\left[\left(\overline{\overline{\mathbf{I}}}+k\overline{\overline{\mathbf{S}}}^{-2}\right)\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\right]^{-1}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\overline{\mathbf{p}}$$

Como o posto da matriz $\overline{\overline{\mathbf{A}}}$ é igual ao número de parâmetros (i.e., se r = M) temos que $\overline{\overline{\mathbf{V}}}^{-1}=\overline{\overline{\mathbf{V}}}^{T}$ então $\left(\overline{\overline{\mathbf{V}}}^{T}\right)^{-1}=\overline{\overline{\mathbf{V}}}$ a equação acima pode se escrita como

$$\boxed{E\left\{\overline{\mathbf{p}}^{*}\right\}=\overline{\overline{\mathbf{V}}}\left[\overline{\overline{\mathbf{I}}}+k\overline{\overline{\mathbf{S}}}^{-2}\right]^{-1}\overline{\overline{\mathbf{V}}}^{\mathrm{T}}\overline{\mathbf{p}}}\qquad(4.10)$$

4.1.2) A Matriz de Covariância de $\bar{\mathbf{p}}^*$ sob as premissas estatísticas 1 1 1 1 — — 1 1

Vimos na equação (7) que

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

Como no estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) e $\bar{\bar{\mathbf{H}}} = \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{A}}} + k\bar{\bar{\mathbf{I}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T$ temos que a equação acima é dada por

$$\mathrm{COV}\left\{\bar{\mathbf{p}}*\right\} = \left(\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{A}}} + k\bar{\bar{\mathbf{I}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^{\mathbf{T}} COV\left\{\bar{\boldsymbol{\varepsilon}}\right\} \bar{\bar{\mathbf{A}}}\left(\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{A}}} + k\bar{\bar{\mathbf{I}}}\right)^{-1}$$

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

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

$$\mathrm{COV}\left\{\bar{\mathbf{p}}*\right\} = \sigma^2\left(\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{A}}} + k\bar{\bar{\mathbf{I}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^{\mathbf{T}} \bar{\bar{\mathbf{A}}}\left(\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{A}}} + k\bar{\bar{\mathbf{I}}}\right)^{-1} \qquad (4.11)$$

Sob as mesmas condições M = r, então usando a decomposição em valores singulares da matriz de sensibilidade temos que

$$\bar{\bar{\mathbf{A}}}^{\mathrm{T}}\bar{\bar{\mathbf{A}}} = \bar{\bar{\mathbf{V}}}_{\mathrm{M}} \bar{\bar{\mathbf{S}}}_{\mathrm{M}} \bar{\bar{\mathbf{U}}}_{\mathrm{M}}^{\mathrm{T}} \bar{\bar{\mathbf{U}}}_{\mathrm{M}} \bar{\bar{\mathbf{S}}}_{\mathrm{M}} \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} = \bar{\bar{\mathbf{V}}}_{\mathrm{M}} \bar{\bar{\mathbf{S}}}_{\mathrm{M}}^2 \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

Substituindo na expressão (4.11) temos que

$$\mathrm{COV}\left\{\bar{\mathbf{p}}*\right\} = \sigma^2\left(\bar{\bar{\mathbf{V}}}_{\mathrm{M}} \bar{\bar{\mathbf{S}}}_{\mathrm{M}}^2 \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} + k\bar{\bar{\mathbf{I}}}\right)^{-1} \bar{\bar{\mathbf{V}}}_{\mathrm{M}} \bar{\bar{\mathbf{S}}}_{\mathrm{M}}^2 \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} \left(\bar{\bar{\mathbf{V}}}_{\mathrm{M}} \bar{\bar{\mathbf{S}}}_{\mathrm{M}}^2 \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} + k\bar{\bar{\mathbf{I}}}\right)^{-1}$$

Como o posto da matriz $\bar{\bar{\mathbf{A}}}$ é igual ao número de parâmetros (i.e., se r = M) temos que $\bar{\bar{\mathbf{V}}}_{\mathrm{M}} \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} = \bar{\bar{\mathbf{I}}}$, então podemos reescrever a equação acima como:

$$\mathrm{COV}\left\{\bar{\mathbf{p}}*\right\} = \sigma^2\left[\bar{\bar{\mathbf{V}}}_{\mathrm{M}}\left(\bar{\bar{\mathbf{S}}}_{\mathrm{M}}^2 + k\right)\bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}\right]^{-1} \bar{\bar{\mathbf{V}}}_{\mathrm{M}} \bar{\bar{\mathbf{S}}}_{\mathrm{M}}^2 \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} \left[\bar{\bar{\mathbf{V}}}_{\mathrm{M}}\left(\bar{\bar{\mathbf{S}}}_{\mathrm{M}}^2 + k\right)\bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}\right]^{-1}$$

Como o posto da matriz $\bar{\bar{\mathbf{A}}}$ é igual ao número de parâmetros (i.e., se r = M) temos que $\overline{\overline{\mathbf{V}}}_M^{\;-1} = \overline{\overline{\mathbf{V}}}_M^{\;T}$ então $\left(\overline{\overline{\mathbf{V}}}_M^{\;T}\right)^{-1} = \overline{\overline{\mathbf{V}}}_M$ e a equação acima pode se escrita como

$$\mathrm{COV}\left\{\bar{\mathbf{p}}^*\right\} = \sigma^2 \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 + k\right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 + k\right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

$$\mathrm{COV}\left\{\bar{\mathbf{p}}^*\right\} = \sigma^2 \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 + k\right]^{-1} \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 + k\right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

Como $\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 = \overline{\overline{\mathbf{S}}}_{\mathrm{M}} \overline{\overline{\mathbf{S}}}_{\mathrm{M}}$ a equação acima pode se escrita como

$$\mathrm{COV}\left\{\bar{\mathbf{p}}^*\right\} = \sigma^2 \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 + k\right]^{-1} \overline{\overline{\mathbf{S}}}_{\mathrm{M}} \overline{\overline{\mathbf{S}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 + k\right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

$$\mathrm{COV}\left\{\bar{\mathbf{p}}^*\right\} = \sigma^2 \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\left(\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 + k\right)\right]^{-1} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\left(\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^2 + k\right)\right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

$$\mathrm{COV}\left\{\bar{\mathbf{p}}^*\right\} = \sigma^2 \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-1} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

$$\boxed{\mathrm{COV}\left\{\bar{\mathbf{p}}^*\right\} = \sigma^2 \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}} \qquad (4.12)$$

Vamos chamar a matriz diagonal da equação (4.12) como

$$\left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2} = \overline{\overline{\mathbf{B}}}$$

em que o j-ésimo elemento é dado por

$$B_{jj} = \frac{1}{\left(S_j + \dfrac{k}{S_j}\right)^2} = \frac{1}{\left(\dfrac{S_j^{\,2} + k}{S_j}\right)^2}$$

logo

$$B_{jj} = \frac{S_j^{\,2}}{\left(S_j^{\,2} + k\right)^2}$$

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

$$\mathrm{var}\{\mathbf{p}*_{\mathrm{k}}\} = \sigma^2 \sum_{j=1}^{M} v_{kj}^2 B_{jj}$$

Substituindo B_{jj} na expressão acima temos

$$\mathrm{var}\{\mathbf{p}*_{\mathrm{k}}\} = \sigma^2 \sum_{j=1}^{M} v_{kj}^2 \frac{S_j^{\,2}}{\left(S_j^{\,2} + k\right)^2} \qquad (4.13)$$

Note que a variância do k-ésimo parâmetro estimado via estimador RIDGE REGRESSION (REGULARIZADOR DE TIKHONOV DE ORDEM ZERO) (N≥M= r) envolve também os M valores singulares, tal como no estimador MQ SOBRE. Portanto, a matriz de covariância do estimador Ridge inclue tanto os valores singulares altos como os baixos. Lembramos que os valores singulares baixos são os responsáveis pelos valores elevados para a variância do estimador MQ SOBRE implicando na instabilidade da solução estimada, isto porque a variância do k-ésimo parâmetro estimado via MQ SOBRE,

$$\mathrm{var}\{\hat{\mathbf{p}}_{\mathrm{k}}\} = \sigma^2 \sum_{j=1}^{M} \frac{v_{kj}^2}{S_j^{\,2}}, \qquad \text{MQ SOBRE}$$

é inversamente proporcional aos valores singulares da matriz de sensibilidade $\bar{\bar{\mathbf{A}}}$. Ao contrário do MQ SOBRE, a variância do k-ésimo parâmetro estimado via RIDGE

REGRESSION (REGULARIZADOR DE TIKHONOV DE ORDEM ZERO) é impedida de crescer infinitamente devido ao parâmetro de regularização (multiplicador de Lagrange) K que não deixa o denominador da equação (4.13) assumir valores próximos a zero.

Veja que a estabilidade da estimativa via RIDGE REGRESSION (REGULARIZADOR DE TIKHONOV DE ORDEM ZERO) é assegurada pois podemos utilizar um valor suficientemente grande de modo a elevar os valores singulares baixos que são responsáveis pelos valores elevados da variância dos parâmetros estimados.

Veja que o parâmetro k causa uma distorção nos M valores singulares. Desta forma o parâmetro k modifica tanto os valores singulares baixos como os altos.

4.1.3) A esperança da distância $(\bar{\mathbf{p}}* - \bar{\mathbf{p}})^{\mathrm{T}}(\bar{\mathbf{p}}* - \bar{\mathbf{p}})$

Vimos na equação (8) que

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

Na equação (4.12) deduzimos que no estimador Ridge Regression (Regularizador de Tikhonov de ordem zero)

$$\mathrm{COV}\left\{\bar{\mathbf{p}}*\right\} = \sigma^2 \bar{\bar{\mathbf{V}}}_{\mathrm{M}} \left[\bar{\bar{\mathbf{S}}}_{\mathrm{M}} + k\bar{\bar{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2} \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

e n a equação (4.10) deduzimos que

$$E\{\bar{\mathbf{p}}^*\} = \bar{\bar{\mathbf{V}}}_M \left[\bar{\bar{\mathbf{I}}} + k\bar{\bar{\mathbf{S}}}_M^{-2}\right]^{-1} \bar{\bar{\mathbf{V}}}_M^{\mathrm{T}} \bar{\mathbf{p}}$$

Logo temos que a esperança da distância $(\bar{\mathbf{p}}* - \bar{\mathbf{p}})^{\mathrm{T}}(\bar{\mathbf{p}}* - \bar{\mathbf{p}})$ é dada por

$$E\{[\bar{\mathbf{p}}* - \bar{\mathbf{p}}]^{\mathrm{T}}[\bar{\mathbf{p}}* - \bar{\mathbf{p}}]\} = \mathrm{tr}\{\sigma^2 \bar{\bar{\mathbf{V}}}_{\mathrm{M}} \left[\bar{\bar{\mathbf{S}}}_{\mathrm{M}} + k\bar{\bar{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2} \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}\} + \left\|\bar{\bar{\mathbf{V}}}_M \left[\bar{\bar{\mathbf{I}}} + k\bar{\bar{\mathbf{S}}}_M^{-2}\right]^{-1} \bar{\bar{\mathbf{V}}}_M^{\mathrm{T}} \bar{\mathbf{p}} - \bar{\mathbf{p}}\right\|_2^2$$

$$E\{[\bar{\mathbf{p}}* - \bar{\mathbf{p}}]^{\mathrm{T}}[\bar{\mathbf{p}}* - \bar{\mathbf{p}}]\} = \mathrm{tr}\{\sigma^2 \bar{\bar{\mathbf{V}}}_{\mathrm{M}} \left[\bar{\bar{\mathbf{S}}}_{\mathrm{M}} + k\bar{\bar{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2} \bar{\bar{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}\} + \left\|\left(\bar{\bar{\mathbf{V}}}_M \left[\bar{\bar{\mathbf{I}}} + k\bar{\bar{\mathbf{S}}}_M^{-2}\right]^{-1} \bar{\bar{\mathbf{V}}}_M^{\mathrm{T}} - \bar{\bar{\mathbf{I}}}\right)\bar{\mathbf{p}}\right\|_2^2$$

Lembrando da propriedade do traço do produto de matrizes

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

$$E\{[\overline{\mathbf{p}}*-\overline{\mathbf{p}}]^{\mathrm{T}}[\overline{\mathbf{p}}*-\overline{\mathbf{p}}]\} = \mathrm{tr}\{\sigma^2\left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2}\overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}\,\overline{\overline{\mathbf{V}}}_{\mathrm{M}}\} + \left\|\left(\overline{\overline{\mathbf{V}}}_M\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1}\overline{\overline{\mathbf{V}}}_M^{\mathrm{T}} - \overline{\overline{\mathbf{I}}}\right)\overline{\mathbf{p}}\right\|_2^2$$

Como $\overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}\,\overline{\overline{\mathbf{V}}}_{\mathrm{M}} = \overline{\overline{\mathbf{I}}}$ temos

$$E\{[\overline{\mathbf{p}}*-\overline{\mathbf{p}}]^{\mathrm{T}}[\overline{\mathbf{p}}*-\overline{\mathbf{p}}]\} = \mathrm{tr}\{\sigma^2\left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2}\} + \left\|\left(\overline{\overline{\mathbf{V}}}_M\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1}\overline{\overline{\mathbf{V}}}_M^{\mathrm{T}} - \overline{\overline{\mathbf{I}}}\right)\overline{\mathbf{p}}\right\|_2^2$$

Como o posto da matriz $\overline{\overline{\mathbf{A}}}$ é igual ao número de parâmetros (i.e., se r = M) temos que $\overline{\overline{\mathbf{V}}}_{\mathrm{M}}\,\overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} = \overline{\overline{\mathbf{I}}}$, então podemos reescrever a equação acima como:

$$E\{[\overline{\mathbf{p}}*-\overline{\mathbf{p}}]^{\mathrm{T}}[\overline{\mathbf{p}}*-\overline{\mathbf{p}}]\} = \mathrm{tr}\{\sigma^2\left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2}\} + \left\|\overline{\overline{\mathbf{V}}}_M\left(\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1} - \overline{\overline{\mathbf{I}}}\right)\overline{\overline{\mathbf{V}}}_M^{\mathrm{T}}\,\overline{\mathbf{p}}\right\|_2^2$$

Chamando $\overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}\ \overline{\mathbf{p}} = \overline{\boldsymbol{\alpha}}_M$ temos que

$$E\{[\overline{\mathbf{p}}*-\overline{\mathbf{p}}]^{\mathrm{T}}[\overline{\mathbf{p}}*-\overline{\mathbf{p}}]\} = \mathrm{tr}\{\sigma^2\left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2}\} + \left\|\overline{\overline{\mathbf{V}}}_M\left(\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1} - \overline{\overline{\mathbf{I}}}\right)\overline{\boldsymbol{\alpha}}_M\right\|_2^2$$

$$E\{\|\overline{\mathbf{p}}*-\overline{\mathbf{p}}\|_2^2\} = \mathrm{tr}\{\sigma^2\left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2}\} +$$
$$\left[\overline{\overline{\mathbf{V}}}_M\left(\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1} - \overline{\overline{\mathbf{I}}}\right)\overline{\boldsymbol{\alpha}}_M\right]^T\left[\overline{\overline{\mathbf{V}}}_M\left(\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1} - \overline{\overline{\mathbf{I}}}\right)\overline{\boldsymbol{\alpha}}_M\right]$$

$$E\{\|\overline{\mathbf{p}}*-\overline{\mathbf{p}}\|_2^2\} = \mathrm{tr}\{\sigma^2\left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2}\} +$$
$$\overline{\boldsymbol{\alpha}}_M^T\left(\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1} - \overline{\overline{\mathbf{I}}}\right)\overline{\overline{\mathbf{V}}}_M^{\mathrm{T}}\,\overline{\overline{\mathbf{V}}}_M\left(\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1} - \overline{\overline{\mathbf{I}}}\right)\overline{\boldsymbol{\alpha}}_M$$

$$E\{\|\overline{\mathbf{p}}*-\overline{\mathbf{p}}\|_2^2\} = \mathrm{tr}\{\sigma^2\left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2}\} + \overline{\boldsymbol{\alpha}}_M^T \left(\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1} - \overline{\overline{\mathbf{I}}}\right)\left(\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1} - \overline{\overline{\mathbf{I}}}\right)\overline{\boldsymbol{\alpha}}_M$$

$$E\{\|\overline{\mathbf{p}}*-\overline{\mathbf{p}}\|_2^2\} = \mathrm{tr}\{\sigma^2\left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}} + k\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-1}\right]^{-2}\} + \overline{\boldsymbol{\alpha}}_M^T \left(\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1} - \overline{\overline{\mathbf{I}}}\right)^2 \overline{\boldsymbol{\alpha}}_M \qquad (4.14)$$

Vamos chamar a matriz diagonal do segundo termo como

$$\left(\left[\overline{\overline{\mathbf{I}}} + k\overline{\overline{\mathbf{S}}}_M^{-2}\right]^{-1} - \overline{\overline{\mathbf{I}}}\right)^2 = \overline{\overline{\mathbf{C}}}$$

em que o j-ésimo elemento é dado por

$$C_{jj} = \left[\left(1+\frac{k}{S_j^2}\right)^{-1} - 1\right]^2 = \left[\left(\frac{S_j^2+k}{S_j^2}\right)^{-1} - 1\right]^2 = \left[\frac{S_j^2}{S_j^2+k} - 1\right]^2 = \left[\frac{S_j^2 - S_j^2 + k}{S_j^2+k}\right]^2 = \left[\frac{k}{S_j^2+k}\right]^2$$

$$C_{jj} = \frac{k^2}{\left(S_j^2+k\right)^2}$$

Vamos então expressar a equação (4.14) em termos de somatório

$$E\{\|\overline{\mathbf{p}}*-\overline{\mathbf{p}}\|_2^2\} = \sigma^2\sum_{\mathrm{j=1}}^{\mathrm{M}} B_{jj} + \sum_{\mathrm{j=1}}^{\mathrm{M}} \alpha_j^2 C_{jj}$$

Substituindo B_{jj} e C_{jj} na expressão acima temos

$$\boxed{E\{\|\overline{\mathbf{p}}*-\overline{\mathbf{p}}\|_2^2\} = \sigma^2\sum_{\mathrm{j=1}}^{\mathrm{M}} \frac{S_j^{\,2}}{\left(S_j^{\,2}+k\right)^2} + \sum_{\mathrm{j=1}}^{\mathrm{M}} \alpha_j^2 \frac{k^2}{\left(S_j^2+k\right)^2}} \qquad (4.15)$$

Então veja o primeiro termo da Esperança da distância $\|\overline{\mathbf{p}}*-\overline{\mathbf{p}}\|_2^2$ equação (4.15) é formada por duas parcelas:

1) $tr\,\{\text{cov}(\bar{\mathbf{p}}*)\} = \sigma^2 \sum_{j=1}^{M} \frac{S_j^2}{\left(S_j^2 + k\right)^2}$ é a primeira parcela e representa uma medida de covariância do estimador Ridge Regression (Regularizador de Tikhonov de ordem zero). Portanto, esta parcela é uma medida da estabilidade da solução estimada.

2) $\left\| \text{E}\left\{\bar{\mathbf{p}}^{+}\right\} - \bar{\mathbf{p}} \right\|_2^2 = \sum_{j=1}^{M} \alpha_j^2 \frac{k^2}{\left(S_j^2 + k\right)^2}$ 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 tendenciosidade do estimador $\bar{\mathbf{p}}*$.

Graficamente a equação (4.15) que representa a esperança da distância $\left(\bar{\mathbf{p}}* - \bar{\mathbf{p}}\right)^{\mathbf{T}}\left(\bar{\mathbf{p}}* - \bar{\mathbf{p}}\right)$ do estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) pode ser simplificada pela Figura 4.1 em que k é o parâmetro de regularização. No appêndice (31.4) deduzimos, matematicamente, o comportamento distinto destas duas parcelas que compõem a equação (4.15), que pode ser resumido como:

1) a primeira parcela, $\sigma^2 \sum_{j=1}^{M} \frac{S_j^2}{\left(S_j^2 + k\right)^2}$, é uma função decrescente monotônica de k, i.e., quanto maior k menor será o valor desta primeira parcela e, por outro lado, quanto menor k maior será o valor desta primeira parcela.

2) a segunda parcela, $\sum_{j=1}^{M} \alpha_j^2 \frac{k^2}{\left(S_j^2 + k\right)^2}$, é uma função crescente monotônica de k i.e., quanto maior k maior será o valor desta segunda parcela e, por outro lado, quanto menor r menor será o valor desta segunda parcela.

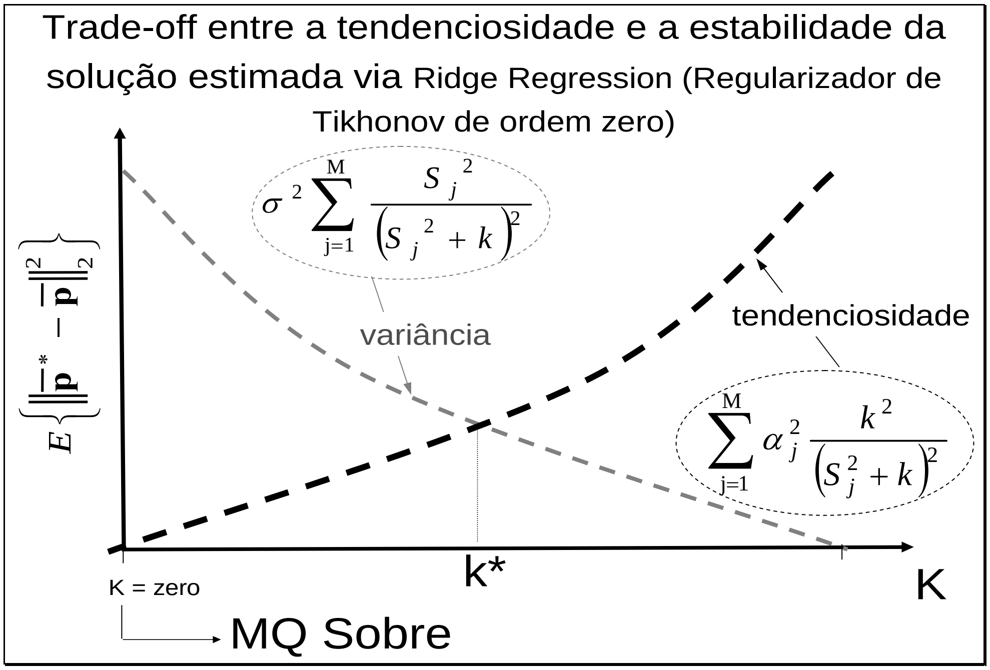

Figura 4.1

Note que se o valor de k for igual a zero a

$$E\{\|\bar{\mathbf{p}}^* - \bar{\mathbf{p}}\|_2^2\} = \sigma^2 \sum_{j=1}^{M} \frac{S_j^{\,2}}{\left(S_j^{\,2}\right)^2} = \sigma^2 \sum_{j=1}^{M} \frac{S_j^{\,2}}{S_j^{\,4}} = \sigma^2 \sum_{j=1}^{M} \frac{1}{S_j^{\,2}}$$

i.e., se k = 0 temos que a

$$E\{\|\bar{\mathbf{p}}^* - \bar{\mathbf{p}}\|_2^2\} = \sigma^2 \sum_{j=1}^{M} \frac{1}{S_j^{\,2}}$$

que corresponde a, $E\{\|\hat{\bar{\mathbf{p}}} - \bar{\mathbf{p}}\|_2^2\}$, i.e., esperânça da distância entre os parâmetros estimados via MQ SOBRE ($\hat{\bar{\mathbf{p}}}$) e os parâmetros $\bar{\mathbf{p}}$, que explicam os dados geofísicos exatos (dados sem ruído)

$$E\{\|\bar{\mathbf{p}}^* - \bar{\mathbf{p}}\|_2^2\} = E\{\|\hat{\bar{\mathbf{p}}} - \bar{\mathbf{p}}\|_2^2\} = \sigma^2 \sum_{j=1}^{M} \frac{1}{S_j^{\,2}} \qquad \text{se k} = 0$$

Então se k = 0, temos o estimador MQ SOBRE que é um estimador NÃO tendencioso, porem um estimador que produz soluções instáveis (com variância máxima dos parâmetros).

A medida que k aumenta o estimador $\bar{\mathbf{p}}^*$ se distancia do estimador $\hat{\bar{\mathbf{p}}}$ (MQ SOBRE). O resultado é que quanto maior o valor de K maior é a tendeciosidade da solução estimada via Ridge Regression (Regularizador de Tikhonov de ordem zero), ou seja, mais informação a priori estamos introduzindo ao problema. Por outro lado, quanto maior o valor de K menor é a variância dos parâmetros estimados via Ridge Regression (Regularizador de Tikhonov de ordem zero). Portanto, quanto maior o valor de K a solução estimada via Ridge Regression (Regularizador de Tikhonov de ordem zero) é mais tendenciosa porém mais estável.

Isto significa que há um trade-off entre a tendenciosidade e a estabilidade da solução estimada via Ridge Regression (Regularizador de Tikhonov de ordem zero). O parâmetro "ótimo" de regularização, k*, será aquele parâmentro que equilibra a tendenciosidade com a variância da solução estimada (Figura 4.1).

4.1.4) A matriz de resolução associada a $\bar{\mathbf{p}}^*$

Vimos na equação (9) que

$$\bar{\bar{\mathbf{R}}} = \bar{\bar{\mathbf{H}}}\ \bar{\bar{\mathbf{A}}}$$

Como no estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) $\bar{\bar{\mathbf{H}}} = \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{A}}} + k\bar{\bar{\mathbf{I}}} \right)^{-1} \bar{\bar{\mathbf{A}}}^T$ temos que a equação acima é dada por

$$\bar{\bar{\mathbf{R}}} = \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{A}}} + k\bar{\bar{\mathbf{I}}} \right)^{-1} \bar{\bar{\mathbf{A}}}^T\ \bar{\bar{\mathbf{A}}} \qquad (4.16)$$

Veja que a matriz de resolução estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) (N ≥ M = r) não é igual a matriz identidade, logo não é máxima.

Sob a condição M = r, então fazendo a decomposição em valores singulares da matriz de sensibilidade temos que

$$\overline{\overline{\mathbf{A}}}^{\mathrm{T}} \overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \overline{\overline{\mathbf{S}}}_{\mathrm{M}} \overline{\overline{\mathbf{U}}}_{\mathrm{M}}^{\mathrm{T}} \overline{\overline{\mathbf{U}}}_{\mathrm{M}} \overline{\overline{\mathbf{S}}}_{\mathrm{M}} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} = \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

Usando o resultado acima a matriz de Resolução do estimador Ridge Regression (Regularizador de Tikhonov de ordem zero),

$$\overline{\overline{\mathbf{R}}} = \left(\overline{\overline{\mathbf{A}}}^{T} \overline{\overline{\mathbf{A}}} + k \overline{\overline{\mathbf{I}}} \right)^{-1} \overline{\overline{\mathbf{A}}}^{T} \overline{\overline{\mathbf{A}}}$$

pode ser reescrita como

$$\overline{\overline{\mathbf{R}}} = \left(\overline{\overline{\mathbf{V}}}_{\mathrm{M}} \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} + k \overline{\overline{\mathbf{I}}} \right)^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

Como o posto da matriz $\overline{\overline{\mathbf{A}}}$ é igual ao número de parâmetros (i.e., se r = M) temos que $\overline{\overline{\mathbf{V}}}_{\mathrm{M}} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} = \overline{\overline{\mathbf{I}}}$, então podemos reescrever a equação acima como:

$$\overline{\overline{\mathbf{R}}} = \left[\overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left(\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} + k \right) \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} \right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

Como o posto da matriz $\overline{\overline{\mathbf{A}}}$ é igual ao número de parâmetros (i.e., se r = M) temos que $\overline{\overline{\mathbf{V}}}_{M}^{-1} = \overline{\overline{\mathbf{V}}}_{M}^{T}$ então $\left(\overline{\overline{\mathbf{V}}}_{M}^{T} \right)^{-1} = \overline{\overline{\mathbf{V}}}_{M}$ e a equação acima pode se escrita como

$$\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\left(\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} + k \right) \right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

$$\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left(\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} + k \right)^{-1} \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

$$\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-2} \left(\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} + k \right) \right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

$$\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-2} \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{2} + k \, \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-2} \right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}$$

$$\boxed{\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \left[\overline{\overline{\mathbf{I}}} + k \, \overline{\overline{\mathbf{S}}}_{\mathrm{M}}^{-2} \right]^{-1} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}}} \qquad (4.17)$$

Veja que a matriz de resolução estimador Ridge Regression (Regularizador de Tikhonov de ordem zero), NÃO é igual a matriz identidade, portanto não é a máxima resolução a menos que k = 0 porque, neste caso teremos que , $\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_{\mathrm{M}} \overline{\overline{\mathbf{V}}}_{\mathrm{M}}^{\mathrm{T}} = \overline{\overline{\mathbf{I}}}_{M}$, pois r = M.

Veja que o i-ésimo elemento da diagonal da matriz de resolução estimador Ridge Regression (Regularizador de Tikhonov de ordem zero), é dado por

$$r_{ii} = \sum_{j=1}^{M} \frac{v_{ij}^{2}}{\left(1 + \dfrac{k}{S_{j}^{2}} \right)} = \sum_{j=1}^{M} \frac{v_{ij}^{2}}{\left(\dfrac{S_{j}^{2} + k}{S_{j}^{2}} \right)} = \sum_{j=1}^{M} \frac{v_{ij}^{2} \; S_{i}^{2}}{(S_{i}^{2} + k)}$$

$$r_{ii} = \sum_{j=1}^{M} v_{ij}^{2} \frac{S_{i}^{2}}{(S_{i}^{2} + k)} \qquad (4.18)$$

Note que $\sum_{j=1}^{M} v_{ij}^{2} = 1$ e o termo $\dfrac{S_{i}^{2}}{(S_{i}^{2} + k)} \leq 1$. Então o i-ésimo elemento da matriz de resolução, $\overline{\overline{\mathbf{R}}}$, estimador Ridge Regression (Regularizador de Tikhonov de ordem zero), será sempre um valor menor ou igual a 1 $(r_{ii} \leq 1)$. Isto significa dizer que quanto maior o valor de k maior será a perda de resolução

Dizemos que cada elemento de $\overline{\mathbf{p}}^{*}$ NÃO é perfeitamente resolvido. Em outras palavras, se a matriz de resolução estimador Ridge Regression (Regularizador de

Tikhonov de ordem zero), é diferente da a matriz identidade, isto significa dizer que o j-ésimo parâmetro estimado é obtido como uma combinação linear de TODOS os parâmetros. Em resumo no estimador Ridge Regression (Regularizador de Tikhonov de ordem zero), em que I $k \neq 0$ então $\overline{\overline{\mathbf{R}}} \neq \overline{\overline{\mathbf{I}}}_M$ então o aspecto da i-ésima linha da matriz de resolução passa de uma delta de Dirac [caso em que k = 0 e $\overline{\overline{\mathbf{R}}} = \overline{\overline{\mathbf{V}}}_M \overline{\overline{\mathbf{V}}}_M^T = \overline{\overline{\mathbf{I}}}_M$, Figura 4.2a] para uma função sinc discreta (Figura 4.2b) em que usamos um valor de $k = k_1$. Se usarmos um $k = k_2$ em que k_2 é muito maior que $k_1 \left(k_2 >>> k_1\right)$, então o aspecto da i-ésima linha da matriz de resolução será também uma função sinc discreta (Figura 4.2c) porém mais achatada que a função sinc mostrada na Figura 4.2b, isto porque menor será o termo $\frac{S_i^2}{(S_i^2 + k)}$ da equção (4.18).

Como

$$\overline{\mathbf{p}}^* = \overline{\overline{\mathbf{R}}}\,\overline{\mathbf{p}} \quad ,$$

então podemos dizer que a estimativa Ridge Regression (Regularizador de Tikhonov de ordem zero) do i-ésimo parâmetro é dada pelo produto escalar da i-ésima linha da matriz de resolução pelo vetor de parâmetros, i.e.,

$$p_i^* = \overline{\mathbf{r}}_i^T \, \overline{\mathbf{p}}$$

Já vimos que quanto maior for o valor de k, mais achatada será a função sinc, isto porque menor será o termo $\frac{S_i^2}{(S_i^2 + k)}$ da equção (4.18), então maior será a contribuição relativa de outros parâmetros em relação ao i-ésimo parâmetro verdadeiro estimado p_i^* . Isto porque a estimativa Ridge Regression (Regularizador de Tikhonov de ordem zero) do i-ésimo parâmetro, p_i^* , é uma combinação linear de todos os parâmetros com uma maior contribuição do i-ésimo parâmetro verdadeiro.

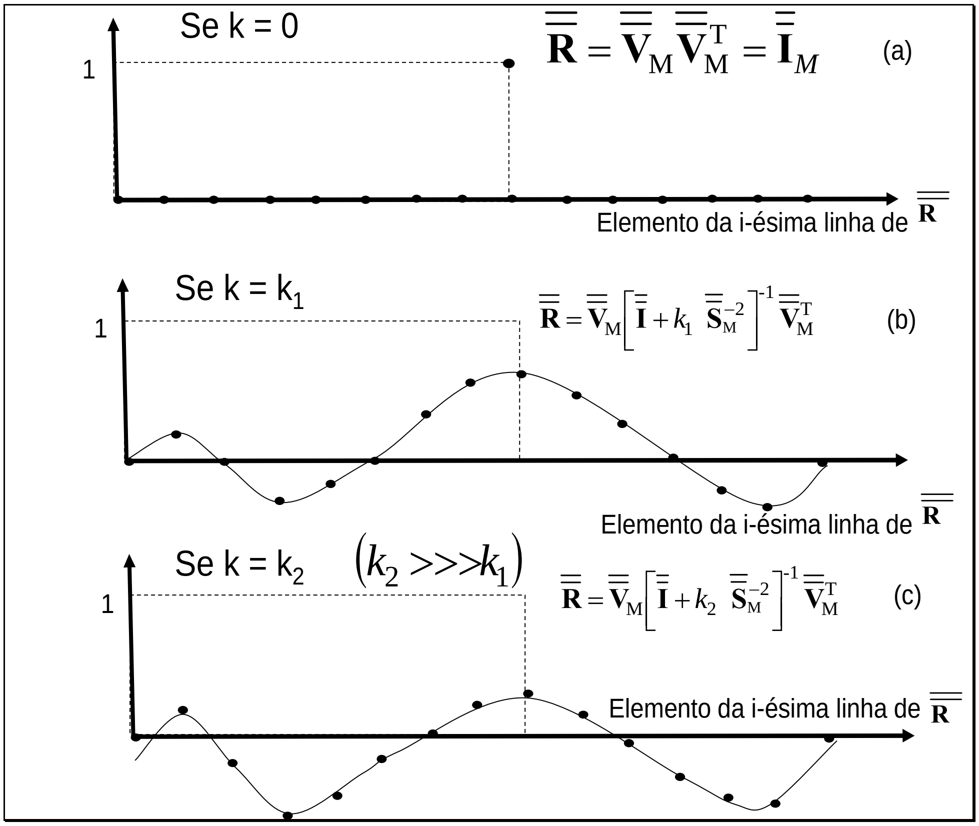

Figura 4.2

4.1.5) A matriz densidade de informação associada a $\overline{\mathbf{p}}^*$

Vimos na equação (10) que

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

Como no estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) $\overline{\overline{\mathbf{H}}} = \left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}} + k \overline{\overline{\mathbf{I}}} \right)^{-1} \overline{\overline{\mathbf{A}}}^T$ temos que a equação acima é dada por

$$\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{A}}} \left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}} + k \overline{\overline{\mathbf{I}}} \right)^{-1} \overline{\overline{\mathbf{A}}}^T \tag{4.19}$$

Fazendo a decomposição em valores singulares da matriz de sensibilidade sob a condição r = M, temos que $\overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{U}}}_M \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{V}}}_M^T$ portanto temos que

$$\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{V}}}_M \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{U}}}_M^T \overline{\overline{\mathbf{U}}}_M \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{V}}}_M^T = \overline{\overline{\mathbf{V}}}_M \overline{\overline{\mathbf{S}}}_M^2 \overline{\overline{\mathbf{V}}}_M^T$$

Usando os resultados acima temos que equação (4.19) é dada por

$$\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{U}}}_M \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{V}}}_M^T \left(\overline{\overline{\mathbf{V}}}_M \overline{\overline{\mathbf{S}}}_M^2 \overline{\overline{\mathbf{V}}}_M^T + k \overline{\overline{\mathbf{I}}} \right)^{-1} \overline{\overline{\mathbf{V}}}_M \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{U}}}_M^T$$

Como o posto da matriz $\overline{\overline{\mathbf{A}}}$ é igual ao número de parâmetros (i.e., se r = M) temos que $\overline{\overline{\mathbf{V}}}_M \overline{\overline{\mathbf{V}}}_M^T = \overline{\overline{\mathbf{I}}}$, então podemos reescrever a equação acima como:

$$\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{U}}}_M \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{V}}}_M^T \left[\overline{\overline{\mathbf{V}}}_M \left(\overline{\overline{\mathbf{S}}}_M^2 + k \right) \overline{\overline{\mathbf{V}}}_M^T \right]^{-1} \overline{\overline{\mathbf{V}}}_M \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{U}}}_M^T$$

como r = m temos que $\overline{\overline{\mathbf{V}}}_M^{-1} = \overline{\overline{\mathbf{V}}}_M^{\;T}$ então

$$\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{U}}}_M \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{V}}}_M^T \overline{\overline{\mathbf{V}}}_M \left[\overline{\overline{\mathbf{S}}}_M^2 + k \right]^{-1} \overline{\overline{\mathbf{V}}}_M^T \overline{\overline{\mathbf{V}}}_M \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{U}}}_M^T$$

$$\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{U}}}_M \overline{\overline{\mathbf{S}}}_M \left[\overline{\overline{\mathbf{S}}}_M^2 + k \right]^{-1} \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{U}}}_M^T$$

Colocando em evidência a matriz $\overline{\overline{\mathbf{S}}}_M^2 = \overline{\overline{\mathbf{S}}}_M^1 \overline{\overline{\mathbf{S}}}_M^1$ no termo a ser invertido temos

$$\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{U}}}_M \overline{\overline{\mathbf{S}}}_M \left[\overline{\overline{\mathbf{S}}}_M \left(\overline{\overline{\mathbf{I}}} + k \overline{\overline{\mathbf{S}}}_M^{-2} \right) \overline{\overline{\mathbf{S}}}_M \right]^{-1} \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{U}}}_M^T$$

$$\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{U}}}_M \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{S}}}_M^{-1} \left[\overline{\overline{\mathbf{I}}} + k \overline{\overline{\mathbf{S}}}_M^{-2} \right]^{-1} \overline{\overline{\mathbf{S}}}_M^{-1} \overline{\overline{\mathbf{S}}}_M \overline{\overline{\mathbf{U}}}_M^T$$

$$\boxed{\overline{\overline{\mathbf{F}}} = \overline{\overline{\mathbf{U}}}_M \left[\overline{\overline{\mathbf{I}}} + k \overline{\overline{\mathbf{S}}}_M^{-2} \right]^{-1} \overline{\overline{\mathbf{U}}}_M^T} \qquad (4.20)$$

A equação (4.20) mostra que a matriz densidade de informação do estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) é diferente da matriz identidade. Isto quer dizer que os dados observados e os dados ajustados via estimador Ridge Regression (Regularizador de Tikhonov de ordem zero) é diferente de zero ($\bar{\boldsymbol{\varepsilon}} \neq \bar{\mathbf{0}}$).

Veja que a matriz densidade de informação do estimador Ridge Regression (Regularizador de Tikhonov de ordem zero), NÃO é igual a matriz identidade, portanto o ajuste não é exato.

Veja que o i-ésimo elemento da diagonal da matriz densidade de informação do estimador Ridge Regression (Regularizador de Tikhonov de ordem zero), é dado por

$$f_{ii} = \sum_{j=1}^{M} \frac{u_{ij}^2}{\left(1 + \frac{k}{S_j^2}\right)} = \sum_{j=1}^{M} \frac{u_{ij}^2}{\left(\frac{S_j^2 + k}{S_j^2}\right)} = \sum_{j=1}^{M} \frac{u_{ij}^2 \; S_i^2}{(S_i^2 + k)}$$

$$f_{ii} = \sum_{j=1}^{M} u_{ij}^2 \frac{S_i^2}{(S_i^2 + k)} \tag{4.21}$$

Note que $\sum_{j=1}^{M} u_{ij}^2 = 1$ e o termo $\frac{S_i^2}{(S_i^2 + k)} \leq 1$. Então o i-ésimo elemento da matriz densidade de informação, $\overline{\overline{\mathbf{F}}}$, do estimador Ridge Regression (Regularizador de Tikhonov de ordem zero), será sempre um valor menor ou igual a 1 $(f_{ii} \leq 1)$. Isto significa dizer que quanto maior o valor de k maior será o desajuste, i.e. maior o resíduo entre os dados observados e os dados ajustados via estimador Ridge Regression (Regularizador de Tikhonov de ordem zero).