UMA POSSÍVEL SOLUÇÃO PARA O PROBLEMA DE MAL CONDICIONAMENTO DA MATRIZ DO MODELO DE REGRESSÃO

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Srta. Reiko Aoki e aprovada pela comissão julgadora.

Campinas, 18 de dezembro de 1992.

Prof.Dr. Euclydes Custódio Lima Filho

Dissertação apresentada ao Instituto de Matemática , Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística.

Conteúdo

1 CAPÍTULO 1

1.1 Introdução

Seja o modelo de regressão linear simples:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

onde

$\mathbf{y}$ é o vetor $n \times 1$ de variáveis resposta

$\mathbf{X}$ é a matriz $n \times p$ de variáveis preditoras

β é o vetor $p \times 1$ de parâmetros do modelo

ϵ é o vetor erro n×1, onde $E(\ \epsilon\ ) = \mathbf{0}$ e $Var(\ \epsilon\ ) = \sigma^2\mathbf{I}$.

Este tipo de ajuste é bastante utilizado na prática e uma maneira de encontrarmos o vetor de parâmetros estimados do modelo é através da utilização do método dos quadrados mínimos.

O método dos quadrados mínimos foi primeiramente proposto como um procedimento algébrico por Legendre em 1805 e mais tarde como um procedemento estatístico por Gauss em 1809, a partir disto seu uso foi difundido nas áreas de análise de dados estatísticos e hoje é um dos métodos mais conhecidos e utilizados.

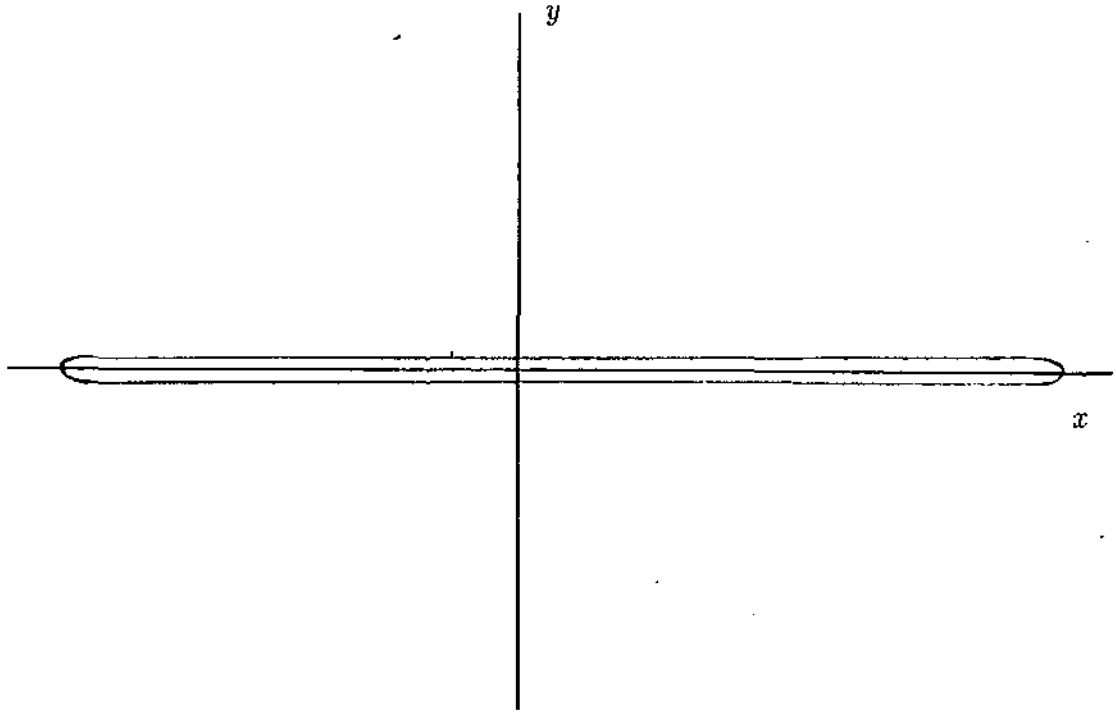
Quando o posto da matriz do modelo de regressão é completa, ou seja, o posto de $\mathbf{X}$ é igual p, com a utilização do método dos quadrados mínimos podemos encontrar um estimador linear não viciado de variância mínima entre todos os estimadores lineares não viciados. Mas quando o posto da matriz do modelo $\mathbf{X}$ é menor do que p, não é possível o cálculo de todos os coeficientes de regressão a partir das observações , ou quando a matriz do modelo de regressão $\mathbf{X}$ é mal condicionada pequenas mudanças nos valores desta matriz ou no vetor de variáveis resposta pode causar uma grande mudança no vetor de parâmetros estimados. Estas pequenas mudanças são frequentes na prática já que ao fazermos os cálculos necessários para a obtenção dos parâmetros estimados em microcomputadores estaremos sujeitos a erros de arredondamento e truncamento, e além destes tipos de erros

podemos ter também erros de medida ao coletar os dados.

Quando o posto da matriz do modelo de regressão é menor do que p, mostraremos que através da inclusão de determinadas linhas de observações se torna possível o cálculo de todos os coeficientes de regressão. Quando a matriz é mal condicionada, uma das maneiras de resolvermos este problema é através da seleção de variáveis a ser utilizada no modelo, ou seja, se tivermos alguma variável que não seja significante podemos retirá-la do modelo. Mas na prática existem variáveis que apesar do teste estatístico indicar que elas não são significantes, são variáveis importantes para o pesquisador e que não podem ser omitidos do modelo, ou seja , é necessário um procedimento que não omita estas variáveis e que ao mesmo tempo não haja grandes mudanças nos valores dos parâmetros estimados quando ocorre pequenas perturbações nas obsevaç oes.

Sabemos que quando uma matriz é mal condicionada o número de condição desta matriz é grande, e o número de condição depende da norma utilizada. Utilizaremos aqui a norma definida como: $\| \mathbf{X} \| = max_{\| \mathbf{y} \| = 1} \| \mathbf{Xy} \|$ para o cálculo do número de condição da matriz do modelo $\mathbf{X}$. Neste caso o número de condição vai ser grande quando o menor auto-valor de $\mathbf{X^T X}$ for pequeno em relação ao maior auto-valor de $\mathbf{X^T X}$ e mostramos que com a inclusão de determinadas linhas de observações podemos aumentar os valores dos auto-valores de $\mathbf{X^T X}$ que são pequenos em relação ao maior auto-valor de $\mathbf{X^T X}$ sem alterar o valor do maior auto-valor de $\mathbf{X^T X}$, fazendo com que diminua o número de condição de $\mathbf{X}$ e assim a matriz ficará bem condicionada. Mostramos também que a variância dos parâmetros estimados dependem dos auto-valores de $\mathbf{X^T X}$, e que quanto menor o auto-valor de $\mathbf{X^T X}$ maior pode ser a variância dos parâmetros estimados. Assim aumentando-se os valores dos auto-valores pequenos em relação ao maior auto-valor os valores das variâncias dos parâmetros estimados podem diminuir.

Geometricamente, se colocarmos a matriz $\mathbf{X^T X}$ na forma quadrática obteremos uma elipsóide e quando temos algum auto-valor pequeno em relação ao maior auto-valor significa que temos eixos pequenos em relação ao maior eixo. Suponha que estejamos trabalhando com duas variáveis, ou seja a elipse se encontra em um plano e que o primeiro auto-valor seja muito pequeno em relação ao segundo auto-valor. Teremos então uma elipse bastante achatada.



Neste caso seria como se estivéssemos trabalhando em uma reta ao invés de estarmos trabalhando em um plano. Assim queremos aumentar o tamanho do menor eixo, fazendo com que esta figura se aproxime de uma circunferência, ou seja, queremos aumentar o valor do primeiro auto-valor que é pequeno em relação ao segundo auto-valor.

Assim supondo que todas as variáveis são importantes, ou seja, não queremos descartar nenhuma das variáveis, estudamos uma maneira de melhorar a condição da matriz do modelo de regressão através da inclusão de novas linhas de observações.

Um resumo desta dissertação é o seguinte.

O capítulo 1 foi dividido em 3 seções, onde na primeira seção fizemos um revisão geral sobre o conceito de projetor e projetor ortogonal, a fim de que pudéssemos apresentar uma melhor visão e entendimento do problema e solução do ajuste por quadrados mínimos. Na segunda seção é apresentado o conceito de decomposição em valores singulares e algumas propriedades de matrizes com a utilização desta, relacionando-os com auto-valores e auto-vetores da matriz. Este conceito será bastante utilizado posteriormente a fim de que possamos estudar o problema de mal condicionamento da matriz do modelo de regressão $\mathbf{X}$, como isto afeta os estimadores de quadrados mínimos encontrados para o ajuste do modelo e encontrar uma possível solução para este problema. Na terceira seção, definiremos a inversa de Moore Penrose e algumas propriedades sobre esta serão apresentadas, sendo que esta inversa é única e nos apresenta a melhor solução de quadrados mínimos quando o posto de $\mathbf{X}$ é menor do que p.

No capítulo 2, discutiremos o problema e a solução por quadrados mínimos com a utilização dos conceito introduzidos no capítulo 1.

No capítulo 3, colocaremos o problema de mal condicionamento da matriz de variáveis preditoras $\mathbf{X}$, quando isto ocorre , a sua consequencia e como poderia ser resolvido este problema. Assim como o desenvolvimento de um método recursivo para a solução proposta e a apresentação de duas matrizes mal condicionadas para exemplificar o método proposto, sendo que a primeira matriz é constituído de dados fictícios e a segunda matriz é constituído de dados reais retirado do artigo "An appraisal of least squares programs for the electronic computer from the point of view of the user" de James W. Longley e Silver Spring o qual é bastante citado na literatura quando se trata de problemas relacionados com o mal condicionamento da matriz.

2 CAPÍTULO II

2.1 Projetores

Nesta seção desenvolveremos o conceito de operador projeção e como estes podem estar relacionados à decomposição de soma direta assim como a relação com matrizes idempotentes.

Dizemos que um espaço linear $\Im$ é a soma direta dos subespaços $\Im_1$ e $\Im_2$, e denotaremos $\Im = \Im_1 \oplus \Im_2$ se, e somente se, as seguintes condições forem satisfeitas.

(i) $\Im_1 \subseteq \Im$ e $\Im_2 \subseteq \Im$;

(ii) para cada $x \in \Im$, existe um $x_1 \in \Im_1$ e $x_2 \in \Im_2$ tal que :

$$x = x_1 + x_2 \quad e\ ;$$

(iii) se $x \in \Im_1$ e $x \in \Im_2$ então $x = 0$

Observe que a condição (iii) é análogo a dizermos que $\Im_1$ e $\Im_2$ são virtualmente disjuntas, ou seja, $\Im_1 \cap \Im_2$ consiste somente do vetor nulo, o que implicará que a resolução descrito em (ii) é única. Pois, dado que:

$$x \in \Im \quad e \quad x = x_1 + x_2 = y_1 + y_2 \ ,$$

onde,

$$x_1, y_1 \in \Im_1 \ e \ x_2, y_2 \in \Im_2,$$

então

$$x_1 - y_1 = y_2 - x_2 \ , \ e \ como \ x_1 - y_1 \in \Im_1 \ e \ y_2 - x_2 \in \Im_2 \ ,$$

por (iii) implica que :

$$y_1 = x_1 \ e \ y_2 = x_2 \ \square$$

Se $\Im = \Im_1 \oplus \Im_2$ diremos que $\Im_1$ e $\Im_2$ são complementares, ou $\Im_1$ é complemento de $\Im_2$, ou $\Im_2$ é complemento de $\Im_1$. Se tivermos que $< \mathbf{x}, \mathbf{y} > = 0$ (produto interno de $\mathbf{x}$ e $\mathbf{y}$) para todo $\mathbf{x} \in \Im_1$ e $\mathbf{y} \in \Im_2$ então dizemos que $\Im_1$ é complemento ortogonal de $\Im_2$, ou $\Im_2$ é complemento ortogonal de $\Im_1$.

DEFINIÇÃO 2.1 *Projetor : Considere um vetor arbitrário* $\mathbf{x} \in \Im = \Im_1 \oplus \Im_2$ *e seja* $\mathbf{x} = \mathbf{x_1} + \mathbf{x_2}$, *tal que* $\mathbf{x_1} \in \Im_1$ *e* $\mathbf{x_2} \in \Im_2$, *onde* $\mathbf{x_1}$ *e* $\mathbf{x_2}$ *são únicos. O mapeamento* $\mathbf{P} : \mathbf{x} \rightarrow \mathbf{x_1}$ *é chamado projetor de* $\Im$ *em* $\Im_1$ *na direção de* $\Im_2$

TEOREMA 2.1 *Um operador linear* $\mathbf{P}$ *é um projetor se e somente se* $\mathbf{P}$ *é idempotente, ou seja,* $\mathbf{P^2} = \mathbf{P}$.

Prova:

($\Rightarrow$) Seja $\Im$ um espaço linear, tal que $\Im = \Im_1 \oplus \Im_2$ e $\mathbf{P}$ um projetor de $\Im$ em $\Im_1$ na direção de $\Im_2$. Então para todo $\mathbf{x_1} \in \Im_1$ temos que $\mathbf{P}\ (\ \mathbf{x_1}\ ) = \mathbf{x_1}$ e para todo $\mathbf{x_2} \in \Im_2$ temos que $\mathbf{P}\ (\ \mathbf{x_2}\ ) = \mathbf{0}$. Então para cada $\mathbf{x} \in \Im$, $\mathbf{x} = \mathbf{x_1} + \mathbf{x_2}$ e por (iii) é único. E portanto:

$$\mathbf{P^2}\ (\ \mathbf{x}\ ) = \mathbf{P}\ (\ \mathbf{P}\ (\ \mathbf{x_1} + \mathbf{x_2}\ )\ ) = \mathbf{P}\ (\ \mathbf{P}\ (\ \mathbf{x_1}\ ) + \mathbf{P}\ (\ \mathbf{x_2}\ )\ ) =$$
$$= \mathbf{P}\ (\ \mathbf{x_1}\ ) = \mathbf{x_1} = \mathbf{P}\ (\ \mathbf{x}\ ).$$

($\Leftarrow$) Seja $\mathbf{x} \in \Im$ onde $\Im_1 = \mathrm{Im}(\ \mathbf{P}\ )$ (imagem de $\mathbf{P}$), $\Im_2 = \mathrm{Ker}(\mathbf{P})$ (Kernell de $\mathbf{P}$) e $\mathbf{x_1} = \mathbf{P}\ (\ \mathbf{x}\ )$. Então se $\mathbf{P^2} = \mathbf{P}$, temos que:

$$\mathbf{P}\ (\ \mathbf{x_1}\ ) = \mathbf{P}\ (\ \mathbf{P}\ (\ \mathbf{x}\ )\ ) = \mathbf{P^2}\ (\ \mathbf{x}\ ) = \mathbf{P}\ (\ \mathbf{x}\ ) = \mathbf{x_1}.$$

Seja $\mathbf{x_2} = \mathbf{x}$ - $\mathbf{x_1}$. Então :

$$\mathbf{P}\ (\ \mathbf{x_2}\ ) = \mathbf{P}\ (\ \mathbf{x} - \mathbf{x_1}\ ) = \mathbf{P}\ (\ \mathbf{x}\ ) - \mathbf{P}\ (\ \mathbf{x_1}\ ) = \mathbf{x_1} - \mathbf{x_1} = 0.$$

Então $\mathbf{x} = \mathbf{x_1} + \mathbf{x_2}$ onde $\mathbf{x_1} \in Im\ (\ \mathbf{P}\ )$ e $\mathbf{x_2} \in K\ (\ \mathbf{P}\ )$. Claramente $Im(\ \mathbf{P}\ ) \cap \mathrm{K}\ (\ \mathbf{P}\ ) = 0$ e portanto $\mathbf{P}$ é um projetor de $\Im$ em $\Im_1$ na direção de $\Im_2 \square$.

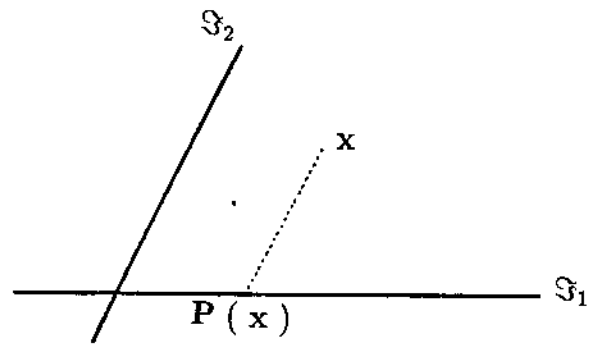


Fig 1.1

No caso onde $\Im$ é o plano real a Fig 1.1 indica a interpretação destas palavras.

COROLÁRIO 2.1 *Se* **P** *é o projetor de* $\Im$ *em* $\Im_1$ *na direção de* $\Im_2$ *, então* **I - P** *é o projetor de* $\Im$ *em* $\Im_2$ *na direção de* $\Im_1$.

Prova:
Como:

$$(\mathbf{I} - \mathbf{P})^2 = \mathbf{I} - 2\mathbf{P} + \mathbf{P}^2 \quad e \; \mathbf{P}^2 = \mathbf{P}$$

$$(\mathbf{I} - \mathbf{P})^2 = (\mathbf{I} - \mathbf{P})$$

e portanto **I - P** é idempotente e **I - P** é projetor.Agora se **P** é um projetor de $\Im$ em $\Im_1$ na direção de $\Im_2$, temos que:

$$\mathbf{P}x = x_1$$

e portanto

$$(\mathbf{I} - \mathbf{P})x = x - \mathbf{P}x = x_1 + x_2 - x_1 = x_2$$

ou seja (I - P) é um projetor de $\Im$ em $\Im_2$ na direção de $\Im_1$□.

TEOREMA 2.2 *Se* **P** *é idempotente, então:*

a)-Im(**I - P** *) = K (* **P** *)*
b)-K (**I - P** *) = Im(* **P** *)*

Prova:

a)-Se $\mathbf{x} \in$ Im(**I** - **P**) , então existe um $\mathbf{y} \in \mathcal{C}$ (onde $\mathcal{C}$ é um espaço *linear*) , tal que $\mathbf{x} =$ (**I** - **P**)**y**.Então:

$$\mathbf{Px} = \mathbf{P}(\mathbf{I} - \mathbf{P})\mathbf{y} = \mathbf{Py} - \mathbf{P}^2\mathbf{y} = \mathbf{Py} - \mathbf{Py} = \mathbf{0}$$

E portanto,

$$Im\ (\mathbf{I} - \mathbf{P}) \subseteq K\ (\mathbf{P})$$

Agora se $\mathbf{x} \in$ K (**P**), então $\mathbf{Px} = \mathbf{0}$ e portanto

$$(\mathbf{I} - \mathbf{P})\mathbf{x} = \mathbf{x} - \mathbf{Px} = \mathbf{x}$$

ou seja $\mathbf{x} \in$ Im(**I** - **P**), e temos que:

$$K\ (\mathbf{P}) \subseteq Im\ (\mathbf{I} - \mathbf{P})\ \square$$

b)- Se $\mathbf{x} \in$ K (**I** - **P**), então

$$(\mathbf{I} - \mathbf{P})\mathbf{x} = 0$$

e

$$\mathbf{x} = \mathbf{Px}$$

e portanto

$$K\ (\mathbf{I} - \mathbf{P}) \subseteq Im\ (\mathbf{P})$$

Agora se $\mathbf{x} \in$ Im(**P**), então existe um $\mathbf{y} \in \mathcal{C}$, tal que $\mathbf{Py} = \mathbf{x}$.
Então,

$$(\mathbf{I} - \mathbf{P})\mathbf{x} = (\mathbf{I} - \mathbf{P})\mathbf{Py} = \mathbf{Py} - \mathbf{P}^2\mathbf{y} = \mathbf{Py} - \mathbf{Py} = 0$$

E portanto,

$$Im\ (\mathbf{P}) \subseteq K\ (\mathbf{I} - \mathbf{P})\ \square$$

TEOREMA 2.3 *Se* **P** *é idempotente, então* $\mathcal{C} = K\ (\mathbf{P}) \oplus Im(\mathbf{P})$.

Prova:
O axioma (i) é certamente satisfeito. Agora para qualquer $\mathbf{x} \in \mathcal{C}$ podemos escrever $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ onde $\mathbf{x}_1 = (\mathbf{I} - \mathbf{P})\mathbf{x}$ e $\mathbf{x}_2 = \mathbf{Px}$, pois $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ $= (\mathbf{I} - \mathbf{P})\mathbf{x} + \mathbf{Px} = \mathbf{x} - \mathbf{Px} + \mathbf{Px} = \mathbf{x}$. Então $\mathbf{x}_1 \in$ Im(**I** - **P**) e pelo item (a) do Teorema 1.2, $\mathbf{x}_1 \in$ K (**P**) e $\mathbf{x}_2 \in$ Im(**P**) e portanto o axioma (ii) é satisfeito. Se $\mathbf{x} \in K($ **P**) e $\mathbf{x} \in$ Im(**P**) então pelo item (b) do teorema 1.2 (**I** - **P**)**x** = **0** e **Px** = **0**. E portanto $\mathbf{x} = \mathbf{0}$ e o axioma (iii) é satisfeito. $\square$

DEFINIÇÃO 2.2 *Seja* $\Im$ *um espaço linear com produto interno. Para um subespaço* $\Im_1$ *qualquer, denotaremos por* $\Im_1^{\perp}$ *o complemento ortogonal de* $\Im_1$, *então* $\Im = \Im_1 \oplus \Im_1^{\perp}$, *ou seja todo* $\mathbf{x} \in \Im$ *pode ser expresso unicamente como* $\mathbf{x} = \mathbf{x_1} + \mathbf{x_2}$, *onde* $\mathbf{x}_1 \in \Im_1$ *e* $\mathbf{x_2} \in \Im_1^{\perp}$. *A correspondência* $\mathbf{P} : \mathbf{x} \rightarrow \mathbf{x_1}$ *é uma transformação linear de* $\Im$, *o qual é chamado projeção ortogonal de* $\Im$ *sobre* $\Im_1$.

TEOREMA 2.4 *Se* $\mathbf{P}$ *é um projetor ortogonal de* $\Im$ *em* $\Im_1$, *então* ($\mathbf{I}$ - $\mathbf{P}$) *é um projetor ortogonal de* $\Im$ *em* $\Im_1^{\perp}$.

Prova: Análogo ao corolário 1.1.

TEOREMA 2.5 *Seja* $\mathbf{P}$ *um projetor ortogonal de* $\Im$ *em* $\Im_1$. *Então* $\mathbf{P^2} = \mathbf{P}$ *e* $\mathbf{P^T} = \mathbf{P}$, *onde* $\mathbf{P^T}$ *é o transposto de* $\mathbf{P}$. *Reciprocamente qualquer transformação linear* $\mathbf{P}$ *com estas propriedades é um projetor ortogonal em algum subespaço.*

Prova:

(I)- Se $\mathbf{P}$ é um projetor ortogonal de $\Im$ *em* $\Im_1$ então $\mathbf{P^2} = \mathbf{P}$ e $\mathbf{P^T} = \mathbf{P}$.

A primeira parte pode ser obtida através do teorema -2.1. Agora como $\mathbf{Px} \in \Im_1$ e $(\mathbf{I} - \mathbf{P})\mathbf{y} \in \Im_1^{\perp}$ $\forall\, \mathbf{x}, \mathbf{y} \in \Im$, temos que $< (\mathbf{I} - \mathbf{P})\mathbf{y}, \mathbf{Px} > = 0\ \forall\, \mathbf{x}, \mathbf{y} \in \Im$, ou seja, $< \mathbf{P^T}(\mathbf{I} - \mathbf{P})\mathbf{y}, \mathbf{x} > = 0\ \forall\, \mathbf{x}, \mathbf{y} \in \Im$. Então,

$$\mathbf{P^T}(\mathbf{I} - \mathbf{P})\mathbf{y} = 0, \qquad \forall\, \mathbf{y} \in \Im$$

$$\Rightarrow \quad \mathbf{P^T}(\mathbf{I} - \mathbf{P}) = 0$$

$$\Rightarrow \quad \mathbf{P^T} = \mathbf{P^T P}$$

e como,

$$\mathbf{P} = (\mathbf{P^T})^{\mathbf{T}} = (\mathbf{P^T P})^{\mathbf{T}} = \mathbf{P^T P} = \mathbf{P^T}$$

e portanto

$$\mathbf{P^T} = \mathbf{P}$$

(II)-Se $\mathbf{P} = \mathbf{P^2}$ e $\mathbf{P^T} = \mathbf{P}$ então $\mathbf{P}$ é um projetor ortogonal.

Seja $\Im_1$ o conjunto de todos os vetores $\mathbf{x_1}$, tal que $\mathbf{Px} = \mathbf{x_1}$ para algum $\mathbf{x} \in \Im$. Pelo Teorema -2.1 , se $\mathbf{P^2} = \mathbf{P}$,

$$\mathbf{Px_1} = \mathbf{x_1} \quad , \forall\ \mathbf{x_1} \in \Im_1$$

$$\mathbf{Px_2} = \mathbf{0} \quad , \forall\, \mathbf{x_2} = \mathbf{x} - \mathbf{x_1}$$

Queremos mostrar que para todo $\mathbf{x} \in \Im$, $\mathbf{x} = \mathbf{x_1} + \mathbf{x_2}$, onde $\mathbf{x_1} \in \Im_1$ e $\mathbf{x_2} \in \Im_1^{\perp}$. Como,

$$< \mathbf{Px}\ ,\ \mathbf{x_2}\ > = (\ \mathbf{Px}\ )^T\ \mathbf{x_2} = \mathbf{x^T P^T x_2}$$

e por suposição

$$\mathbf{P^T} = \mathbf{P}$$

$$< \mathbf{Px}\ ,\ \mathbf{x_2}\ > = \mathbf{x^T P x_2} = \mathbf{0}$$

$$\forall\, \mathbf{x} \in \Im\ , \quad \forall\, \mathbf{x_2} = \mathbf{x} - \mathbf{x_1}$$

e portanto,

$$\mathbf{x_2} \in \Im_1^{\perp} \quad \square$$

TEOREMA 2.6 *Seja* $\mathbf{P}$ *um projetor ortogonal de* $\Im$ *em* $\Im_1$. *Então,* $\|\ \mathbf{y}\ \|_2 \geq \|\ \mathbf{P}\ (\ \mathbf{y}\ )\ \|_2, \forall\, \mathbf{y} \in \Im$, *onde* $\|\ \ \|_2$ = *norma Euclidiana.(apêndice A, definição A.2)*

Prova:
Temos que:
(I) $\|\ \mathbf{Py}\ \|_2^2 \geq 0, \forall\, \mathbf{y} \in \Im$,

então

$$\|\ \mathbf{Py}\ \|_2^2 = < \mathbf{Py}\ ,\ \mathbf{Py}\ > = \mathbf{y^T P^T P y} =$$
$$= \mathbf{y^T P^2 y} = \mathbf{y^T P y} = < \mathbf{Py}\ ,\ \mathbf{y}\ > \geq 0$$

(II) $\|\ (\ \mathbf{I} - \mathbf{P}\ )\mathbf{y}\ \|_2^2 \geq 0, \forall\, \mathbf{y} \in \Im$,

então

$$\|\ (\ \mathbf{I} - \mathbf{P}\ )\mathbf{y}\ \|_2^2 = < (\ \mathbf{I} - \mathbf{P}\ )\mathbf{y}\ ,\ (\ \mathbf{I} - \mathbf{P}\ )\mathbf{y}\ > =$$

$$= \mathbf{y^T}\,(\,\mathbf{I}-\mathbf{P}\,)^T\,(\,\mathbf{I}-\mathbf{P}\,)\mathbf{y} = \mathbf{y^T}\,(\,\mathbf{I}-\mathbf{P}\,)^2\,\mathbf{y} =$$
$$= \mathbf{y^T}\,(\,\mathbf{I}-\mathbf{P}\,)\,\mathbf{y} = <\,(\,\mathbf{I}-\mathbf{P}\,)\mathbf{y}\,,\,\mathbf{y}\,> \geq 0$$

Agora ,

$$\|\ \mathbf{y}\ \|_2^2 - \|\ \mathbf{Py}\ \|_2^2 =$$
$$= <\,\mathbf{y}\,,\,\mathbf{y}\,> - <\,\mathbf{Py}\,,\,\mathbf{Py}\,> =^{(I)}$$
$$= <\,\mathbf{y}\,,\,\mathbf{y}\,> - <\,\mathbf{Py}\,,\,\mathbf{y}\,> \geq^{(II)} 0$$

e portanto

$$\|\ \mathbf{Py}\ \|_2 \leq \|\ \mathbf{y}\ \|_2\ \square$$

2.2 Decomposição em Valores Singulares

TEOREMA 2.7 *Decomposição em Valores Singulares: Seja* $\mathbf{A_{m\times n}}$ *uma matriz com m linhas e n colunas, e de posto r. Então existe uma matriz ortogonal* $\mathbf{P_{m\times m}}$ *(apêndice B) , uma matriz ortogonal* $\mathbf{Q_{n\times n}}$ *e uma matriz diagonal* $\mathbf{D_{m\times n}}$ *, tal que:*

$$\mathbf{P^TAQ} = \mathbf{D} \quad ou \quad \mathbf{A} = \mathbf{PDQ^T}$$

onde $\mathbf{D}$ *é uma matriz nula exceto na diagonal* $\sigma_1, \ldots, \sigma_r$ *que são os valores singulares de* $\mathbf{A}$.

Antes de provarmos o teorema, apresentaremos alguns resultados que serão utilizados.

1)-para $\forall$ matriz $\mathbf{A}$, $\mathbf{A^TA}$ é simétrica e positiva semidefinida.
Prova:Imediata.

2)-Os auto-valores de $\mathbf{A^TA}$ são não negativos.
Prova: Seja $\mathbf{A^TA} = \mathbf{B}$, então $\mathbf{Bx} = \lambda\mathbf{x}$ onde $\mathbf{B}$ é uma matriz real $n\times n$ e simétrica, λ é um escalar real chamado auto-valor, e $\mathbf{x}$ é um vetor $n\times 1$ chamado auto-vetor. Então,

$$\mathbf{Bx} = \lambda\mathbf{x}$$
$$\Rightarrow \quad \mathbf{x^TBx} = \lambda\mathbf{x^Tx}$$

e como $\mathbf{B}$ é semidefinida positiva $\mathbf{x^T B x} \geq 0 \; \forall \; \mathbf{x}$, e portanto $\lambda \geq 0$ pois $\mathbf{x^T x} \geq 0$.

3)-A raíz quadrada não negativa dos auto-valores de $\mathbf{A^T A}$ são chamados de valores singulares de $\mathbf{A}$.
Prova do Teorema :

Seja $\lambda_1, \ldots, \lambda_n$ e $\mathbf{x_1}, \ldots, \mathbf{x_n}$ os auto-valores e os correspondentes auto-vetores ortonormais de $\mathbf{A^T A}$, e seja $\sigma_i = \lambda_i^{1\backslash 2}$, i=1,...,n. Podemos supor que os $\lambda_i' s$, estejam ordenados tal que $\sigma_i > 0$ para i=1,...,r e $\sigma_i = 0$ para i=r+1,...,n.
Seja $\mathbf{y_i} = \frac{1}{\sigma_i}\mathbf{A x_i}$, i=1,...,r. Então:

$$\mathbf{y_i^T y_j} = \frac{1}{\sigma_i \sigma_j}\mathbf{x_i^T A^T A x_j} =$$

$$= \frac{\lambda_j}{\sigma_i \sigma_j}\mathbf{x_i^T x_j} = \delta_{ij} \qquad i,j = 1,...,r$$

$$onde, \quad \delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{caso contrário} \end{cases}$$

Nós podemos estender $\mathbf{y_1},...,\mathbf{y_r}$ para uma base ortonormal $\mathbf{y_1},...,\mathbf{y_m}$, e então:

$$\mathbf{P} = (\, \mathbf{y_1}, ..., \mathbf{y_m})$$

$$\mathbf{Q^T} = (\, \mathbf{x_1}, ..., \mathbf{x_n}\,)$$

as matrizes $\mathbf{P}$ e $\mathbf{Q}$ são ortogonais(apêndice B), e:

$$\mathbf{y_i^T A x_j} = \frac{1}{\sigma_i}\mathbf{x_i^T A^T A x_j} =$$

$$= \frac{1}{\sigma_i}\lambda_j \mathbf{x_i^T x_j} = \frac{1}{\sigma_i}\lambda_j \delta ij, \qquad i = 1,...,r, \; j = 1,...,n$$

Como $\mathbf{x_j^T A^T A x_j} = 0$ para j > r, temos que $\mathbf{A x_j} = 0$ para j > r e então:

$$\mathbf{y_i^T A x_j} = \begin{cases} 0 & \text{se } i > r, j > r \\ \sigma_j \mathbf{y_i^T y_j} & i > r, j \leq r \end{cases} \qquad \square$$

COROLÁRIO 2.2

$$\sigma_n \parallel \mathbf{x} \parallel_2 \leq \parallel \mathbf{Ax} \parallel_2 \leq \sigma_1 \parallel \mathbf{x} \parallel_2 \quad ,\forall\, \mathbf{x} \in \Re^n$$

onde σ_n é o menor valor singular e σ_1 é o maior valor singular.

Prova: temos que $\mathbf{Ax} = \mathbf{PDQ^T x}$ e como as matrizes ortogonais preservam a norma (apêndice B, corolário $B.1$) chamando-se $\mathbf{DQ^T x}$ de $\mathbf{y}$, temos:

$$\parallel \mathbf{Ax} \parallel_2 = \parallel \mathbf{PDQ^T x} \parallel_2 = \parallel \mathbf{Py} \parallel_2 = \parallel \mathbf{y} \parallel_2 = \parallel \mathbf{DQ^T x} \parallel_2$$

onde $\mathbf{y}$ é um vetor m×1. Chamando-se agora, $\mathbf{Q^T x}$ de $\mathbf{z}$, onde $\mathbf{z}$ é um vetor n×1 teremos que:

$$\sigma_n \parallel \mathbf{z} \parallel_2 \leq \parallel \mathbf{Dz} \parallel_2 \leq \sigma_1 \parallel \mathbf{z} \parallel_2 ,$$

pois $\mathbf{D}$ é uma matriz diagonal cujos elementos da diagonal são os valores singulares de $\mathbf{A^T A}$,

$$\sigma_n \parallel \mathbf{z} \parallel_2 \leq \parallel \mathbf{PDz} \parallel_2 \leq \sigma_1 \parallel \mathbf{z} \parallel_2$$

$$\sigma_n \parallel \mathbf{Q^T x} \parallel_2 \leq \parallel \mathbf{PDQ^T x} \parallel_2 \leq \parallel \mathbf{Q^T x} \parallel_2 \sigma_1$$

$$\sigma_n \parallel \mathbf{x} \parallel_2 \leq \parallel \mathbf{Ax} \parallel_2 \leq \parallel \mathbf{x} \parallel_2 \sigma_1. \qquad \square$$

TEOREMA 2.8 *Seja $\mathbf{A_{n\times n}}$ uma matriz simétrica e suponha que os seus auto-valores estejam ordenados. Então:*

$$\lambda_1 \mathbf{x^T x} \leq \mathbf{x^T Ax} \leq \lambda_n \mathbf{x^T x}$$

para todo $\mathbf{x} \in \Re^n$.

e

$$\lambda_{max} = \lambda_n = max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x^T Ax}}{\mathbf{x^T x}} = max_{\mathbf{x^T x}=1} \mathbf{x^T Ax}$$

$$\lambda_{min} = \lambda_1 = min_{\mathbf{x}\neq 0} \frac{\mathbf{x^T Ax}}{\mathbf{x^T x}} = min_{\mathbf{x^T x}=1} \mathbf{x^T Ax}$$

Prova:
Como a matriz $\mathbf{A}$ é simétrica, existe uma matriz $\mathbf{U_{n\times n}}$ ortogonal (Apêndice B) tal que $\mathbf{A = UDU^T}$ com D = diag(λ_1 , ... , λ_n) (Horn and Johnson , Cap 4, seção 1). Então , para $\forall\ \mathbf{x} \in \Re^n$:

$$\mathbf{x^T A x} = \mathbf{x^T U D U^T x} = (\ \mathbf{U^T x}\ )^T\ \mathbf{D}\ (\ \mathbf{U^T x}\ ) = \sum_{i=1}^{n}\ \lambda_i\ |\ (\ \mathbf{u_i^T x}\ )\ |^2$$

e portanto:

$$\lambda_1 \sum_{i=1}^{n} |\ (\ \mathbf{u_i^T x}\ )\ |^2 \leq \sum_{i=1}^{n}\ \lambda_i\ |\ (\ \mathbf{u_i^T x}\ )\ |^2 \leq\ \lambda_n\ \sum_{i=1}^{n} |\ (\ \mathbf{u_i^T x}\ )\ |^2$$

e

$$\lambda_1\ \sum_{i=1}^{n} |\ x_i\ |^2 \leq \sum_{i=1}^{n}\ \lambda_i\ |\ (\ \mathbf{u_i^T x}\ )\ |^2 \leq\ \lambda_n\ \sum_{i=1}^{n} |\ x_i\ |^2,$$

onde x_i é o i-ésimo elemento do vetor $\mathbf{x}$.
(obs:pois $\mathbf{U}$ é ortogonal),e então:

$$\lambda_1\ \mathbf{x^T x}\ \leq\ \mathbf{x^T A x}\ \leq\ \lambda_n\ \mathbf{x^T x}$$

Agora:

$$\frac{\mathbf{x^T A x}}{\mathbf{x^T x}} \leq\ \lambda_n \quad e \quad \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} \geq\ \lambda_1$$

e como se $\mathbf{x}$ for um auto-vetor de $\mathbf{A}$ correspondente ao auto-valor λ_i , $\mathbf{x^T A x}$ $= \mathbf{x^T}\lambda_i\mathbf{x} = \lambda_i\mathbf{x^T x}$, a desigualdade acima será igualdade quando :

$$max_{\mathbf{x}\neq\mathbf{0}}\ \ \frac{\mathbf{x^T A x}}{\mathbf{x^T x}}\ =\ \lambda_n\ \ e\ \ min_{\mathbf{x}\neq\mathbf{0}}\ \frac{\mathbf{x^T A x}}{\mathbf{x^T x}}\ =\ \lambda_1 \quad (1.8.1)$$

e como $\|\ \mathbf{x}\ \|_2^2 = \mathbf{x^T x}$ e,

$$\frac{\mathbf{x^T A x}}{\mathbf{x^T x}}\ =\ \left(\frac{\mathbf{x}}{(\ \mathbf{x^T x}\ )^{\frac{1}{2}}}\right)^T\ \mathbf{A}\ \left(\frac{\mathbf{x}}{(\ \mathbf{x^T x}\ )^{\frac{1}{2}}}\right)$$

as equações (1.8.1) são equivalentes à:

$$max_{\mathbf{x^T x}=1}\ \ \mathbf{x^T A x}\ =\ \lambda_n\ e\ min_{\mathbf{x^T x}\ =\ 1}\ \ \mathbf{x^T A x}\ =\ \lambda_1 \qquad \square$$

TEOREMA 2.9 *Seja* **A** *uma matriz n×n simétrica com autovalores* $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ *e seja k um dado inteiro, onde* $1 \leq k \leq n$. *Então:*

$$min_{\mathbf{w_1},\mathbf{w_2},...,\mathbf{w_{n-k}} \in \Re^n} \quad max_{\mathbf{x}\neq\mathbf{0},\ \mathbf{x}\in\Re^n,\ \mathbf{x}\perp\mathbf{w_1},\mathbf{w_2},...,\mathbf{w_{n-k}}} \quad \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} = \lambda_k \quad (1.9.1)$$

e

$$max_{\mathbf{w1},\mathbf{w2},...,\mathbf{w_{k-1}} \in \Re^n} \quad min_{\mathbf{x}\neq\mathbf{0},\ \mathbf{x}\in\Re^n,\ \mathbf{x}\perp\mathbf{w_1},\mathbf{w_2},...,\mathbf{w_{r-1}}} \quad \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} = \lambda_k \quad (1.9.2)$$

A demonstração deste teorema se encontra no apêndice C.

TEOREMA 2.10 *Sejam* **A** *e* **B** *matrizes n×n simétricas e sejam os autovalores* $\lambda_i(\mathbf{A})$, $\lambda_i(\mathbf{B})$ *e* $\lambda_i(\mathbf{A}+\mathbf{B})$, *em ordem crescente. então para cada k=1,...,n , teremos:*

$$\lambda_k(\mathbf{A}) + \lambda_1(\mathbf{B}) \leq \lambda_k(\mathbf{A}+\mathbf{B}) \leq \lambda_k(\mathbf{A}) + \lambda_n(\mathbf{B})$$

Prova:

Pelo teorema -2.8, temos que:

$$\lambda_1(\mathbf{B}) \leq \frac{\mathbf{x^T B x}}{\mathbf{x^T x}} \leq \lambda_n(\mathbf{B})$$

para qualquer $\mathbf{x} \in \Re^n$ não nulo. E pelo teorema -2.9 , para qualquer k=1,..,n temos que:

$$\lambda_k(\mathbf{A}+\mathbf{B}) = min_{\mathbf{w_1},...,\mathbf{w_{n-k}} \in \Re^n} max_{\mathbf{x}\neq\mathbf{0},\ \mathbf{x}\perp\mathbf{w_1},...,\mathbf{w_{n-k}}} \frac{\mathbf{x^T}(\mathbf{A}+\mathbf{B})\mathbf{x}}{\mathbf{x^T x}}$$

$$= min_{\mathbf{w_1},...,\mathbf{w_{n-k}} \in \Re^n} max_{\mathbf{x}\neq\mathbf{0},\ \mathbf{x}\perp\mathbf{w_1},...,\mathbf{w_{n-k}}} \left[\frac{\mathbf{x^T A x}}{\mathbf{x^T x}} + \frac{\mathbf{x^T B x}}{\mathbf{x^T x}}\right]$$

$$\geq min_{\mathbf{w_1},...,\mathbf{w_{n-k}} \in \Re^n} max_{\mathbf{x}\neq\mathbf{0},\ \mathbf{x}\perp\mathbf{w_1},...,\mathbf{w_{n-k}}} \left[\frac{\mathbf{x^T A x}}{\mathbf{x^T x}} + \lambda_1(\mathbf{B})\right]$$

$$= \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{B})$$

e

$$\lambda_k(\mathbf{A} + \mathbf{B}) =$$

$$= max_{\mathbf{w_1},\ldots,\mathbf{w_{k-1}} \in \Re^n} min_{\mathbf{x} \neq \mathbf{0},\ \mathbf{x} \in \Re^n,\ \mathbf{x} \perp \mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \frac{\mathbf{x}^T(\mathbf{A} + \mathbf{B})\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

$$= max_{\mathbf{w_1},\ldots,\mathbf{w_{k-1}} \in \Re^n} min_{\mathbf{x} \neq \mathbf{0},\ \mathbf{x} \perp \mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \left[\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} + \frac{\mathbf{x}^T\mathbf{B}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}\right]$$

$$\leq max_{\mathbf{w_1},\ldots,\mathbf{w_{k-1}} \in \Re^n} min_{\mathbf{x} \neq \mathbf{0},\ \mathbf{x} \perp \mathbf{w_1},\ldots,\mathbf{w_{k-1}}} \left[\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} + \lambda_n(\mathbf{B})\right]$$

$$= \lambda_k(\mathbf{A}) + \lambda_n(\mathbf{B})$$

e portanto $\lambda_k(\mathbf{A}) + \lambda_1(\mathbf{B}) \leq \lambda_k(\mathbf{A} + \mathbf{B}) \leq \lambda_k(\mathbf{A}) + \lambda_n(\mathbf{B})$. □

TEOREMA 2.11 *Seja* **A** *uma matriz com m linhas e n colunas, e se* $\| \bullet \|$ *é uma dada norma em* $\mathbf{M_{m,n}}$ *(apêndice A) . Então para todo* $\epsilon > 0$*, existe uma matriz* $\mathbf{A}_\epsilon$ *de tamanho* $m \times n$*, com valores singulares distintos , tal que* $\| \mathbf{A} - \mathbf{A}_\epsilon \| < \epsilon$.

Prova:
Suponhamos que m ≥ n e seja $\mathbf{A} = \mathbf{PDQ}^T$ a decomposição em valores singulares de **A** e $\mathbf{D}_\delta$ uma matriz nula exceto na diagonal $\sigma_1 + \delta$, $\sigma_2 + 2\delta$,..., $\sigma_n + n\delta$. Então se todos os valores singulares de **A** são iguais , $\mathbf{D}_\delta$ terá entradas distintas na diagonal para todo $\delta > 0$; e caso contrário , se escolhermos $\delta > 0$ talque nδ é menor do que a menor diferença entre os dois valores singulares distintos sucessivos , então $\mathbf{D}_\delta$ terá entradas distintas na diagonal. Nos dois casos $\mathbf{D}_\delta \rightarrow \mathbf{D}$ quando $\delta \rightarrow 0$.

Se definirmos $\mathbf{A}_\delta = \mathbf{PD}_\delta\mathbf{Q}^T$, então:

$$\| \mathbf{A} - \mathbf{A}_\delta \| = \| \mathbf{PDQ}^T - \mathbf{PD}_\delta\mathbf{Q}^T \| = \| \mathbf{D} - \mathbf{D}_\delta \| \rightarrow 0, \ quando\, \delta \rightarrow 0.$$

(Se m ≤ n , a prova é análogo). □

TEOREMA 2.12 *Seja* $\mathbf{A}$ *uma matriz* $m \times n$, σ_1 , ... , σ_q , *números reais não negativos, e seja* q *o* $min(m,n)$, *e defina* $\mathbf{A}^{"}_{(m+n)\times(m+n)}$ *como:*

$$\mathbf{A}^{"} = \begin{bmatrix} 0 & \mathbf{A} \\ \mathbf{A}^{T} & 0 \end{bmatrix}$$

Então os valores singulares de $\mathbf{A}$ *são* σ_1 , ... , σ_q *se e somente se os* $m+n$ *auto-valores de* $\mathbf{A}^{"}$ *são* σ_1 , ... , σ_q , $-\sigma_1$, ... , $-\sigma_q$, *e* $\mid m-n \mid$ *zeros.*

Prova:
Suponhamos que m $\geq$ n e seja $\mathbf{A} = \mathbf{PDQ}^{\mathbf{T}}$ a decomposição em valores singulares de $\mathbf{A}$. Escrevendo-se:

$$\mathbf{D} = \begin{bmatrix} \mathbf{S} \\ 0 \end{bmatrix}$$

onde $\mathbf{S}$ é uma matriz diagonal com elementos na diagonal σ_1 , ... , σ_n, e:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P_1} & \mathbf{P_2} \end{bmatrix} ,$$

onde $\mathbf{P}_1$ é de tamanho m×n e $\mathbf{P}_2$ é de tamanho m×(m-n). Seja $\mathbf{P}' = \frac{\mathbf{P_1}}{\sqrt{2}}$ e $\mathbf{Q}' = \frac{\mathbf{Q}}{\sqrt{2}}$, então a matriz:

$$\mathbf{U} = \begin{bmatrix} \mathbf{P}' & -\mathbf{P}' & \mathbf{P_2} \\ \mathbf{Q}' & \mathbf{Q}' & 0 \end{bmatrix}$$

é ortogonal pois:

$$\mathbf{U}^{\mathbf{T}}\mathbf{U} = \begin{bmatrix} \mathbf{P}'^{\mathbf{T}} & \mathbf{Q}'^{\mathbf{T}} \\ -\mathbf{P}'^{\mathbf{T}} & \mathbf{Q}'^{\mathbf{T}} \\ \mathbf{P}_2^{\mathbf{T}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{P}' & -\mathbf{P}' & \mathbf{P_2} \\ \mathbf{Q}' & \mathbf{Q}' & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{I_{n\times n}} & \mathbf{0} & \mathbf{P'^{T}P_2} \\ \mathbf{0} & \mathbf{I_{n\times n}} & \mathbf{-P'^{T}P_2} \\ \mathbf{P_2^{T}P'} & \mathbf{P_2^{T}P'} & \mathbf{I_{(m-n)\times(m-n)}} \end{bmatrix}$$

$$= \mathbf{I_{(m+n)\times(m+n)}}$$

e

$$\mathbf{U}\begin{bmatrix} \mathbf{S} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{-S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{U^{T}} =$$

$$= \begin{bmatrix} \mathbf{P'SP'} - \mathbf{P'SP'} & \mathbf{2P'SQ'} \\ \mathbf{Q'SP'} + \mathbf{Q'SP'} & \mathbf{Q'SQ'} - \mathbf{Q'SQ'} \end{bmatrix} =$$

$$= \begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A^{T}} & \mathbf{0} \end{bmatrix} = \mathbf{A''}$$

(Obs: se m<n a prova é análogo). □

TEOREMA 2.13 *Seja* **A** , **B** *matrizes com m linhas e n colunas , seja* **E** = **B** - **A** , *e seja* $q = min(m,n)$. *Se* $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_q$ *são os valores singulares de* **A** *e* $\tau_1 \geq \tau_2 \geq , \ldots , \tau_q$ *são os valores singulares de* **B**, *então:*

a)- $| \sigma_i - \tau_i | \leq \| \mathbf{E} \|_2$ *(Apêndice A, Definição A.3) para todo* $i=1,\ldots,q$, *e*

b)- $[\sum_{i=1}^{q} (\sigma_i - \tau_i)^2]^{\frac{1}{2}} \leq \| \mathbf{E} \|_2$

A prova deste teorema se encontra no apêndice C.

TEOREMA 2.14 *Seja* $\mathbf{A}$ *uma matriz com m linhas e n colunas e seja* $\mathbf{A}'$ *a matriz obtida ao retirar qualquer uma das colunas de* $\mathbf{A}$. *Seja* σ_i *os valores singulares de* $\mathbf{A}$ *e seja* σ_i' *os valores singulares de* $\mathbf{A}'$, *ambas em ordem crescente. Então:*
a)-Se $m \geq n$:

$$\sigma_1 \geq \sigma_1' \geq \sigma_2 \geq \sigma_2' \geq \sigma_3 \geq \ldots \geq \sigma_{n-1}' \geq \sigma_n \geq 0$$

b)-Se $m < n$:

$$\sigma_1 \geq \sigma_1' \geq \ldots \geq \sigma_{m-1} \geq \sigma_{m-1}' \geq \sigma_m \geq \sigma_m' \geq 0$$

Seja agora, $\mathbf{A}''$ *a matriz obtida ao retirar uma linha de* $\mathbf{A}$ *e seja* σ'' *os valores singulares de* $\mathbf{A}''$ *em ordem crescente, então:*
c)-Se $n \geq m$:

$$\sigma_1 \geq \sigma_1'' \geq \sigma_2 \geq \sigma_2'' \geq \sigma_3 \geq \ldots \geq \sigma_{m-1}'' \geq \sigma_m \geq 0$$

d)-Se $n < m$:

$$\sigma_1 \geq \sigma_1'' \geq \ldots \geq \sigma_{n-1} \geq \sigma_{n-1}'' \geq \sigma_n \geq \sigma_n'' \geq 0$$

Prova:
O quadrado do valor singular de $\mathbf{A}$ são os auto-valores da matriz simétrica $\mathbf{A^T A}$, e o quadrado dos valores singulares de $\mathbf{A}'$ são os auto-valores de $\mathbf{A}'^{\mathbf{T}}\mathbf{A}'$.
Como:

$$\mathbf{A^T A} = \begin{pmatrix} \sum_{i=1}^n a_{i1}^2 & \sum_{i=1}^n a_{i1}a_{i2} & \ldots & \sum_{i=1}^n a_{i1}a_{in} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \sum_{i=1}^n a_{in}a_{i1} & \sum_{i=1}^n a_{in}a_{i2} & \ldots & \sum_{i=1}^n a_{in}^2 \end{pmatrix}$$

e $\mathbf{A}'^{\mathbf{*}}\mathbf{A}'$, com a j-ésima coluna deletado é dado por:

$$\mathbf{A'^{*}A'} =$$

$$\begin{pmatrix}
\sum_{i=1}^{n} a_{i1}^2 & \cdots & \sum_{i=1}^{n} a_{i1}a_{i(j-1)} & \sum_{i=1}^{n} a_{i1}a_{i(j+1)} & \cdots & \sum_{i=1}^{n} a_{i1}a_{in} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\sum_{i=1}^{n} a_{i(j-1)}a_{i1} & \cdots & \sum_{i=1}^{n} a_{i(j-1)} & \sum_{i=1}^{n} a_{i(j-1)}a_{i(j+1)} & \cdots & \sum_{i=1}^{n} a_{i(j-1)}a_{in} \\
\sum_{i=1}^{n} a_{i(j+1)}a_{i1} & \cdots & \sum_{i=1}^{n} a_{i(j+1)}a_{i(j-1)} & \sum_{i=1}^{n} a_{i(j+1)}^2 & \cdots & \sum_{i=1}^{n} a_{i(j+1)}a_{in} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\sum_{i=1}^{n} a_{in}a_{i1} & \cdots & \sum_{i=1}^{n} a_{in}a_{i(j-1)} & \sum_{i=1}^{n} a_{in}a_{i(j+1)} & \cdots & \sum_{i=1}^{n} a_{in}^2
\end{pmatrix}$$

$\mathbf{A'^{T}A'}$ é uma submatriz de $\mathbf{A^{T}A}$ e como $\mathbf{A^{T}A}$ e $\mathbf{A'^{T}A'}$ são matrizes quadradas $\mathbf{A'^{T}A'}$ é uma submatriz principal de $\mathbf{A^{T}A}$. Seja λ_i e λ_i' os auto-valores de $\mathbf{A^{T}A}$ e $\mathbf{A'^{T}A'}$ respectivamente . Seja $\mathbf{x} = [\ \mathbf{x'}^{\mathbf{T}}\ \epsilon\ ]^T$, $\mathbf{x'} \in \Re^{n-1}$, $\epsilon \in \Re$; $\mathbf{w_i} = [\ \mathbf{w_i'}^{\mathbf{T}}\gamma\ ]^T$, $\mathbf{w_i'} \in \Re^{n-1}$, $\gamma \in \Re$ e $\mathbf{e_j}$ a j-ésima base padrão. Então pelo teorema -2.9 , temos que:

$$\lambda_{k+1}(\ \mathbf{A^{T}A}\ ) = min_{\mathbf{w_1},\ldots,\mathbf{w_{n-k-1}} \in \Re^n}\ max_{\mathbf{x} \neq 0\ ,\ \mathbf{x} \in \Re^n\ ,\ \mathbf{x} \perp \mathbf{w_1},\ldots,\mathbf{w_{n-k-1}}}\ \frac{\mathbf{x^{T}A^{T}Ax}}{\mathbf{x^{T}x}}$$

$$\geq\ min_{\mathbf{w_1},\ldots,\mathbf{w_{n-k-1}} \in \Re^n}\ max_{\mathbf{x} \neq 0\ ,\ \mathbf{x} \perp \mathbf{w_1},\ldots,\mathbf{w_{n-k-1}}\ ,\ \mathbf{x} \perp \mathbf{e_j}}\ \frac{\mathbf{x^{T}A^{T}Ax}}{\mathbf{x^{T}x}}$$

$$= min_{\mathbf{w_1'},\ldots,\mathbf{w_{n-k-1}'} \in \Re^{n-1}}\ max_{\mathbf{x'} \neq \mathbf{0}\ ,\ \mathbf{x'} \in \Re^{n-1}\ ,\ \mathbf{x'} \perp \mathbf{w_1'},\ldots,\mathbf{w_{n-k-1}'}}\ \frac{\mathbf{x'^{T}A'^{T}A'x'}}{\mathbf{x'^{T}x'}} = \lambda_k(\ \mathbf{A'^{T}A'}$$

e

$$\lambda_k(\ \mathbf{A^{T}A}\ ) = max_{\mathbf{w_1},\ldots,\mathbf{w_{k-1}} \in \Re^n}\ min_{\mathbf{x} \neq \mathbf{0}\ ,\ \mathbf{x} \in \Re^n\ ,\ \mathbf{x} \perp \mathbf{w_1},\ldots,\mathbf{w_{k-1}}}\ \frac{\mathbf{x^{T}A^{T}Ax}}{\mathbf{x^{T}x}}$$

$$\leq\ max_{\mathbf{w_1},\ldots,\mathbf{w_{k-1}} \in \Re^n}\ min_{\mathbf{x} \neq \mathbf{0}\ ,\ \mathbf{x} \in \Re^n\ ,\ \mathbf{x} \perp \mathbf{w_1},\ldots,\mathbf{w_{k-1}}\ ,\ \mathbf{x} \perp \mathbf{e_j}}\ \frac{\mathbf{x^{T}A^{T}Ax}}{\mathbf{x^{T}x}}$$

$$= max_{\mathbf{w_1'},\ldots,\mathbf{w_{k-1}'} \in \Re^{n-1}}\ min_{\mathbf{x'} \neq \mathbf{0}\ ,\ \mathbf{x'} \in \Re^{n-1}\ ,\ \mathbf{x'} \perp \mathbf{w_1'},\ldots,\mathbf{w_{k-1}'}}\ \frac{\mathbf{x'^{T}A'^{T}A'x'}}{\mathbf{x'^{T}x'}} = \lambda_k(\ \mathbf{A'^{T}A'}\ )$$

ou seja:

$$\lambda_k(\ \mathbf{A}^{\mathbf{T}}\mathbf{A}\ ) \ \leq\ \lambda_k(\ \mathbf{A'}^{\mathbf{T}}\mathbf{A'}\ ) \ \leq\ \lambda_{k+1}(\ \mathbf{A}^{\mathbf{T}}\mathbf{A}\ )\ ,\forall\ 1\ \leq\ k\ \leq\ n-1$$

e portanto resulta-se nos itens a e b.

Agora, temos que o quadrado do valor singular de $\mathbf{A}$ são os auto-valores de $\mathbf{A}^{\mathbf{T}}\mathbf{A}$ que são os mesmos auto-valores de $\mathbf{A}\mathbf{A}^{\mathbf{T}}$. Então:

$$\mathbf{A}\mathbf{A}^{\mathbf{T}} \ =\ \begin{pmatrix} \sum_{i=1}^{n} a_{1i}^2 & \sum_{i=1}^{n} a_{1i}a_{2i} & \cdots & \sum_{i=1}^{n} a_{1i}a_{mi} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \sum_{i=1}^{n} a_{mi}a_{1i} & \sum_{i=1}^{n} a_{mi}a_{2i} & \cdots & \sum_{i=1}^{n} a_{mi}^2 \end{pmatrix}$$

e $\mathbf{A"}\mathbf{A"}^{\mathbf{T}}$, com a j-ésima linha retirada é dada por:

$$\mathbf{A"}\mathbf{A"}^{\mathbf{T}} =$$

$$\begin{pmatrix} \sum_{i=1}^{n} a_{1i}^2 & \cdots & \sum_{i=1}^{n} a_{1i}a_{(j-1)i} & \sum_{i=1}^{n} a_{1i}a_{(j+1)i} & \cdots & \sum_{i=1}^{n} a_{1i}a_{mi} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sum_{i=1}^{n} a_{(j-1)i}a_{1i} & \cdots & \sum_{i=1}^{n} a_{(j-1)i} & \sum_{i=1}^{n} a_{(j-1)i}a_{(j+1)i} & \cdots & \sum_{i=1}^{n} a_{(j-1)i}a_{mi} \\ \sum_{i=1}^{n} a_{(j+1)i}a_{1i} & \cdots & \sum_{i=1}^{n} a_{(j+1)i}a_{(j-1)i} & \sum_{i=1}^{n} a_{(j+1)i}^2 & \cdots & \sum_{i=1}^{n} a_{(j+1)i}a_{mi} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sum_{i=1}^{n} a_{mi}a_{1i} & \cdots & \sum_{i=1}^{n} a_{mi}a_{(j-1)i} & \sum_{i=1}^{n} a_{mi}a_{(j+1)i} & \cdots & \sum_{i=1}^{n} a_{mi}^2 \end{pmatrix}$$

Assim, $\mathbf{A"}\mathbf{A"}^{\mathbf{T}}$ é uma submatriz de $\mathbf{A}\mathbf{A}^{\mathbf{T}}$ e como $\mathbf{A}\mathbf{A}^{\mathbf{T}}$ e $\mathbf{A"}\mathbf{A"}^{\mathbf{T}}$ são matrizes quadradas $\mathbf{A"}\mathbf{A"}^{\mathbf{T}}$ é uma submatriz principal de $\mathbf{A}\mathbf{A}^{\mathbf{T}}$ e a prova segue análogo à prova dos itens a e b. □

2.3 Inversa de Moore Penrose

DEFINIÇÃO 2.3 *A Inversa de Moore Penrose: definiremos a inversa de Moore Penrose de* $\mathbf{A} \in \mathbf{M_{m,n}}$, *como sendo a matriz* $\mathbf{X}$ *que satisfaz as seguintes condições.*

$$\mathbf{AXA} \ =\ \mathbf{A}\ , \qquad (1)$$

$$\mathbf{XAX} = \mathbf{X}\,, \qquad (2)$$
$$(\,\mathbf{AX}\,)^T = \mathbf{AX}\,, \qquad (3)$$
$$(\,\mathbf{XA}\,)^T = \mathbf{XA}\,, \qquad (4)$$

e denotaremos por $\mathbf{A^+}$.

TEOREMA 2.15 *Seja* $\mathbf{A} \in \Re^{m\times n}$. *Existe uma única matriz* $\mathbf{X} \in \mathbf{M_{m,n}}$ *que satisfaz as condições (1) a (4) acima.*

Prova:

Seja $\mathbf{x_1}$ *e* $\mathbf{x_2}$ duas matrizes satisfazendo as condições (1) a (4) acima. Por (2) e (3), temos que:

$$\mathbf{x_i^T} = \mathbf{x_i^T A^T x_i^T} = \mathbf{A x_i x_i^T}\,, \quad i = 1,2$$
$$\Rightarrow \quad \mathbf{x_1^T} - \mathbf{x_2^T} = \mathbf{A}(\,\mathbf{x_1 x_1^T} - \mathbf{x_2 x_2^T}$$

e então

$$Im(\,\mathbf{x_1^T} - \mathbf{x_2^T}\,) \subset Im(\,\mathbf{A}\,).$$

Agora usando-se as equações (1) e (4), temos que:

$$\mathbf{A^T} = \mathbf{A^T x_i^T A^T} = \mathbf{x_i A A^T}\,, \quad i = 1,2$$
$$\Rightarrow \quad (\,\mathbf{x_1} - \mathbf{x_2}\,)\mathbf{AA^T} = \mathbf{0}\ ou\ \mathbf{AA^T}(\,\mathbf{x_1^T} - \mathbf{x_2^T}\,) = \mathbf{0}$$

e então

$$Im(\,\mathbf{x_1^T} - \mathbf{x_2^T}\,) \subset Ker(\,\mathbf{AA^T}\,) =^{(I)} Ker(\,\mathbf{A^T}\,)$$

((I) Lancaster, seção 5.3). e como Ker($\mathbf{A^T}$) é ortogonal à Im($\mathbf{A}$) (Lancaster, seção 5.1), temos que:

$$\mathbf{x_1^T} - \mathbf{x_2^T} = \mathbf{0}$$

e portanto $\mathbf{x_1^T} = \mathbf{x_2^T}$. □

TEOREMA 2.16 *Se* $\mathbf{A} \in \mathbf{M_{m,n}}$, *então:*

$$Im(\,\mathbf{A}\,) + Ker(\,\mathbf{A^+}\,) = \Re^m$$
$$Ker(\,\mathbf{A}\,) + Im(\,\mathbf{A^+}\,) = \Re^n$$

Prova: Temos que Im($\mathbf{AA^+}$) = Im($\mathbf{A}$), pois:

$$(I) - se\ \mathbf{y} = \mathbf{AA^+x} \Rightarrow \mathbf{y} \in Im(\mathbf{AA^+})$$
$$e\ \mathbf{y} = \mathbf{A}(\mathbf{A^+x}) \Rightarrow \mathbf{y} \in Im(\mathbf{A})$$

então

$$Im(\mathbf{AA^+}) \subset Im(\mathbf{A})$$

$$(II) - se\ \mathbf{y} = \mathbf{Ax} \Rightarrow \mathbf{y} \in Im(\mathbf{A})$$
$$e\ como\ \mathbf{A} = \mathbf{AA^+A},$$
$$\mathbf{y} = \mathbf{AA^+Ax} = \mathbf{AA^+}(\mathbf{Ax}) \Rightarrow \mathbf{y} \in Im(\mathbf{AA^+})$$

então

$$Im(\mathbf{A}) \subset Im(\mathbf{AA^+})$$
$$e\ portanto\ Im(\mathbf{AA^+}) = Im(\mathbf{A})$$

Agora, seja $\mathbf{H} = \mathbf{AA^+}$, então:

$$\mathbf{A^+H} = \mathbf{A^+AA^+} = \mathbf{A^+}$$

e

$$posto(\mathbf{A^+}) = posto(\mathbf{A^+H}) \leq^{(1)} min(posto(\mathbf{A^+}), posto(\mathbf{H})) \leq posto(\mathbf{H})$$

e

$$posto(\mathbf{H}) = posto(\mathbf{AA^+}) \leq min(posto(\mathbf{A}), posto(\mathbf{A^+})) \leq posto(\mathbf{A^+})$$

((1) Ortega seção 3.5)
e portanto:

$$posto(\mathbf{H}) = posto(\mathbf{AA^+}) = posto(\mathbf{A^+})$$

Assim,

$$dim(Ker\mathbf{A^+}) = m - posto(\mathbf{A^+}) = m - posto(\mathbf{AA^+}) = dim(Ker\mathbf{AA^+})\ (I)$$

(onde dim() é a dimensão e (I) é dado por Lancaster, seção 3.8)

Dado que $\mathbf{x} \neq \mathbf{0}$, temos que $\mathbf{A^+x = 0}$, $\mathbf{x} \in \mathrm{Ker}\mathbf{A^+}$, e como se $\mathbf{A^+x = 0} \Rightarrow \mathbf{AA^+x = 0}$, $\mathbf{x} \in Ker\mathbf{AA^+}$, e portanto Ker($\mathbf{A^+}$) $\subset$ ker($\mathbf{AA^+}$).

Suponhamos que $\mathbf{AA^+ = 0}$, tal que $\mathbf{x} \in Ker(\ \mathbf{AA^+}\ )$, então $\mathbf{A^+AA^+x = 0}$ e como $\mathbf{A^+AA^+ = A^+}$, temos que $\mathbf{A^+x = 0}$, tal que $\mathbf{x} \in Ker(\ \mathbf{A^+}\ )$. Então $Ker(\ \mathbf{AA^+}\ ) \subset Ker(\ \mathbf{A^+}\ )$.
E portanto $ker(\ \mathbf{A^+}\ ) = ker(\ \mathbf{AA^+}\ )$, e pelo teorema 2, seção 5.8 Lancaster, temos que:

$$Ker(\ \mathbf{AA^+}\ ) + Im(\ \mathbf{AA^+}\ ) = \Re^m,$$

e como Im($\mathbf{AA^+}$) = Im($\mathbf{A}$) e Ker($\mathbf{AA^+}$) = Ker($\mathbf{A^+}$), temos que:

$$Ker(\ \mathbf{A^+}\ ) + Im(\ \mathbf{A}\ ) = \Re^n$$

e desenvolvendo de forma análoga ao anterior, usando a definição de inversa de Inversa de Moore, chegaremos ao resultado de que , Ker($\mathbf{A^+A}$) = $Ker(\ \mathbf{A}\ )$ e $Im(\ \mathbf{A^+A}\ )$ = Im($\mathbf{A^+}$). E portanto:

$$Ker(\ \mathbf{A}\ ) + Im(\ \mathbf{A^+}\ ) = \Re^n \qquad \square$$

TEOREMA 2.17 *Para qualquer matriz* $\mathbf{A} \in \mathbf{M_{m,n}}$, *temos que:*

$$Ker(\ \mathbf{A^+}\ ) = Ker(\ \mathbf{A^T}\ )$$

$$Im(\ \mathbf{A^+}\ ) = Im(\ \mathbf{A^T}\ )$$

Prova: Pela definição 2.3 temos que $\mathbf{AA^+}$ e $\mathbf{A^+A}$ são matrizes simétricas. Agora, como:

$$(\ \mathbf{AA^+}\ )(\ \mathbf{AA^+}\ ) = \mathbf{A}(\ \mathbf{A^+AA^+}\ ) = \mathbf{AA^+}$$

e

$$(\ \mathbf{A^+A}\ )(\ \mathbf{A^+A}\ ) = (\ \mathbf{A^+AA^+}\ )\mathbf{A} = \mathbf{A^+A}$$

$\mathbf{A^+A}$ *e* $\mathbf{AA^+}$ são matrizes idempotentes e por Lancaster seção 5.1, se uma matriz $\mathbf{B} \in \mathbf{M_{m,n}}$ é idempotente e simétrica , $Im(\ \mathbf{B}\ ) \oplus Ker(\ \mathbf{B}\ ) = \Re^n$. Então como $\mathbf{AA^+}$ e $\mathbf{A^+A}$ são matrizes simétricas e idempotentes, temos:

$$Im(\ \mathbf{AA^+}\ ) \oplus Ker(\ \mathbf{AA^+}\ ) = \Re^m$$

$$Im(\ \mathbf{A^+A}\ ) \oplus Ker(\ \mathbf{A^+A} = \Re^n$$

e como na prova do teorema 2.16 vimos que a $Im(\ \mathbf{AA^+}\ )$ = Im(**A**), $Ker(\ \mathbf{AA^+}\ ) = Ker(\ \mathbf{A^+}\ )$, $Im(\ \mathbf{A^+A}\ ) = Im(\ \mathbf{A^+}\ )$ e $Ker(\ \mathbf{A^+A}\ )$ = Ker(**A**), temos:

$$Im(\ \mathbf{A}\ ) \oplus Ker(\ \mathbf{A^+}\ ) = \Re^m$$

$$Im(\ \mathbf{A^+}\ ) \oplus Ker(\ \mathbf{A} = \Re^n$$

e como por Lancaster seção 5.1 temos que:

$$Im(\ \mathbf{A}\ ) \oplus Ker(\ \mathbf{A^T}\ ) = \Re^m$$

$$Im(\ \mathbf{A}\ ) \oplus Ker(\ \mathbf{A^T} = \Re^n$$

então:

$$Ker(\ \mathbf{A}+\ ) = Ker(\ \mathbf{A^T}\ ),$$

e

$$Im(\ \mathbf{A^+}\ ) = Im(\ \mathbf{A^T}\ ).$$

TEOREMA 2.18 *Se* $\mathbf{A} \in \mathbf{M_{m,n}}$, *então:*
1-$\mathbf{AA^+}$ *é um projetor ortogonal em* $Im(\ \mathbf{A}\ )$ *na direção de* $Ker(\ \mathbf{A^T}\ )$.
2-$\mathbf{A^+A}$ *é um projetor ortogonal em* $Im(\ \mathbf{A^T}\ )$ *na direção de* $Ker(\ \mathbf{A}\ )$.
3-$(\ \mathbf{I} - \mathbf{AA^+}\ )$ *é um projetor ortogonal em* $Ker(\ \mathbf{A^T}\ )$ *na direção de* $Im(\ \mathbf{A}\ )$.
4-$(\ \mathbf{I} - \mathbf{A^+A}\ )$ *é um projetor ortogonal em* $Ker(\ \mathbf{A}\ )$ *na direção de* $Im(\ \mathbf{A^T}\ )$.

Prova:
1- Como $\mathbf{AA^+}$ é simétrica e idempotente, pelo teorema 2.5 $\mathbf{AA^+}$ é um projetor ortogonal. Agora seja $\mathbf{x}$ um vetor $m \times 1$ $\forall$, então se $\mathbf{AA^+x = y}$, $\mathbf{y} \in$ $Im(\ \mathbf{AA^+}\ )$ e como $Im(\ \mathbf{AA^+}\ ) \oplus Ker(\ \mathbf{AA^+}\ ) = \Re^m$, $\mathbf{AA^+}$ é um projetor ortogonal em $Im(\ \mathbf{AA^+}\ )$ na direção de $Ker(\ \mathbf{AA^+}\ )$, e como pelo teorema 2.16 $Ker(\ \mathbf{AA^+}\ ) = Ker(\ \mathbf{A^+}\ )$ e $Im(\ \mathbf{AA^+}\ ) = Im(\ \mathbf{A}\ )$, e pelo teorema 2.17 $Ker(\ \mathbf{A^+}\ ) = Ker(\ \mathbf{A^T}\ )$, obtém-se o item 1.
2- Análogo ao item 1.

3-Temos que:
a-($\mathbf{I}$ - $\mathbf{AA}^+$)^T =($\mathbf{I}$ - ($\mathbf{AA}^{+\ T}$)) = ($\mathbf{I}$ - $\mathbf{AA}^+$), pois $\mathbf{AA}^+$ é simétrico.
b-($\mathbf{I}$ - $\mathbf{AA}^+$)($\mathbf{I}$ - $\mathbf{AA}^+$) = $\mathbf{I}$ - $\mathbf{AA}^+$ - $\mathbf{AA}^+$ + ($\mathbf{AA}^+$)($\mathbf{AA}^+$) = ($\mathbf{I}$ - $\mathbf{AA}^+$), pois $\mathbf{AA}^+$ é idempotente.

Portanto por a e b tem-se que ($\mathbf{I}$ - $\mathbf{AA}^+$) é um projetor ortogonal. Agora como $\mathbf{AA}^+$ é um projetor ortogonal em $Im(\ \mathbf{A}\ )$ na direção de $Ker(\ \mathbf{A}^\mathbf{T}\ )$, seja $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, onde $\mathbf{x}_1 \in Im(\ \mathbf{A}\ )$ e $\mathbf{x_2} \in Ker(\ \mathbf{A}^\mathbf{T}\ )$. Então:

$$\mathbf{AA}^+\mathbf{x} = \mathbf{x}_1$$

e

$$(\ \mathbf{I} - \mathbf{AA}^+\ )\mathbf{x} = \mathbf{x} - \mathbf{AA}^+\mathbf{x} =$$

$$= \mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_1 = \mathbf{x}_2$$

e portanto ($\mathbf{I}$ - $\mathbf{AA}^+$) é um projetor ortogonal em $Ker(\ \mathbf{A}^\mathbf{T}\ )$ na direção de $Im(\ \mathbf{A}\ )$
4-Análogo ao 3.

TEOREMA 2.19 *Seja* $\mathbf{X} \in \mathbf{M_{m,n}}$ *e* $\mathbf{X} = \mathbf{PDQ^T}$ *a decomposição em valores singulares de* $\mathbf{X}$. *Supondo que* $\mathbf{X}$ *tem posto* r *e se escrevermos esta decomposição da seguinte forma:*

$$\mathbf{X} = \mathbf{PDQ^T}$$

$$= \begin{bmatrix} \mathbf{P_1} & \mathbf{P_2} \end{bmatrix} \begin{bmatrix} \mathbf{D_1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{Q_1^T} \\ \mathbf{Q_2^T} \end{bmatrix}$$

onde $\mathbf{P_1} \in \mathbf{M_{m,r}}$, $\mathbf{P_2} \in \mathbf{M_{m,m-r}}$, $\mathbf{D_1} \in \mathbf{M_{r,r}}$, $\mathbf{Q_1^T} \in \mathbf{M_{r,n}}$ *e* $\mathbf{Q_2^T} \in \mathbf{M_{n-r,n}}$; *temos que:*

$$\mathbf{X^+} = \mathbf{Q_1 D_1^{-1} P_1^T}$$

é a inversa de Moore Penrose de $\mathbf{X}$.

Prova:Como $\mathbf{X}$ tem posto r:

$$\mathbf{X} = \mathbf{PDQ^T}$$

$$= \begin{bmatrix} \mathbf{P_1} & \mathbf{P_2} \end{bmatrix} \begin{bmatrix} \mathbf{D_1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{Q_1^T} \\ \mathbf{Q_2^T} \end{bmatrix}$$

$$= \mathbf{P_1 D_1 Q_1^T}$$

Agora pela definição 2.3, $\mathbf{X^+}$ é a inversa de Moore Penrose de $\mathbf{X}$ se ela satisfizer as condições (1) a (4). Suponha então que $\mathbf{X^+ = Q_1 D_1^{-1} P_1^T}$ é a inversa de Moore Penrose de $\mathbf{X}$ e iremos verificar se $\mathbf{X^+}$ satisfaz estas codições.
(1) $\mathbf{X^+X = X}$

Temos que:

$$\begin{aligned}\mathbf{XX^+X} &= \mathbf{P_1 D_1 Q_1^T Q_1 D_1^{-1} P_1^T P_1 D_1 Q_1^T} \\ &= \mathbf{P_1 D_1 D_1^{-1} D_1 Q_1^T} \\ &= \mathbf{P_1 D_1 Q_1^T = X}\end{aligned}$$

podemos provar as condições restantes (2) a (4) de forma análoga ao que foi feito em (1).

3 CAPÍTULO III

3.1 Problema dos Quadrados Mínimos

O problema do ajuste por quadrados mínimos pode ser expresso da seguinte maneira.

DEFINIÇÃO 3.1 *Dado uma matriz real* $\mathbf{A}_{(m \times n)}$ *onde* $(m \geq n)$ *, e um vetor real* $\mathbf{b}$ *de tamanho m, encontrar um vetor* $\mathbf{x}^{o}$ *de tamanho n, tal que:*

$$\| \mathbf{A}\mathbf{x}^{o} - \mathbf{b} \|_2 = min_{\mathbf{x} \in \Re^n} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2,$$

Analizaremos agora alguns aspectos de quadrados mínimos.
Define-se um conjunto $\mathbf{X}$ como segue:

$$\mathbf{X} = \{ \mathbf{x} : \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2 \leq \| \mathbf{A}\mathbf{z} - \mathbf{b} \|_2 \ , \ \forall \mathbf{z} \in \Re^n \}$$

1)-Se $\mathbf{x} \in \mathbf{X}$ então $\mathbf{A}^{\mathbf{T}}(\mathbf{A}\mathbf{x} - \mathbf{b}) = 0$, ou seja o resíduo $\mathbf{r} = \mathbf{A}\mathbf{x} - \mathbf{b}$ é ortogonal às colunas de $\mathbf{A}$ ($\mathbf{r} \in \mathrm{Im}(\mathbf{A})^{\perp}$).
Prova: Seja $\mathbf{x} \in \mathbf{X}$ e seja $\mathbf{w} = \mathbf{A}^{\mathbf{T}}(\mathbf{A}\mathbf{x} - \mathbf{b})$. Para todo $\alpha \in \Re$ teremos:

$$\| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2^2 \leq \| \mathbf{A}(\mathbf{x} + \alpha\mathbf{w}) - \mathbf{b} \|_2^2$$

Temos que:

$$\| \mathbf{A}(\mathbf{x} + \alpha\mathbf{w}) - \mathbf{b} \|_2^2 =$$

$$= ((\mathbf{A}\mathbf{x} + \alpha\mathbf{A}\mathbf{w})^T - \mathbf{b}^{\mathbf{T}})((\mathbf{A}\mathbf{x} + \alpha\mathbf{A}\mathbf{w}) - \mathbf{b}) =$$

$$= (\mathbf{A}\mathbf{x} + \alpha\mathbf{A}\mathbf{w})^T (\mathbf{A}\mathbf{x} + \alpha\mathbf{A}\mathbf{w}) - (\mathbf{A}\mathbf{x} + \alpha\mathbf{A}\mathbf{w})^T \mathbf{b} -$$

$$\mathbf{b}^{\mathbf{T}} (\mathbf{A}\mathbf{x} + \alpha\mathbf{A}\mathbf{w}) + \mathbf{b}^{\mathbf{T}}\mathbf{b} =$$

$$= \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2^2 + \alpha^2 \| \mathbf{A}\mathbf{w} \|_2^2 + 2\alpha(\mathbf{A}\mathbf{x})^T (\mathbf{A}\mathbf{w}) - 2\alpha\mathbf{b}^{\mathbf{T}}(\mathbf{A}\mathbf{w}) =$$

$$\| \mathbf{Ax} - \mathbf{b} \|_2^2 + \alpha^2 \| \mathbf{Aw} \|_2^2 + 2\alpha \| \mathbf{w} \|_2^2 =$$

$$= \| \mathbf{Ax} - \mathbf{b} \|_2^2 + \alpha(\alpha \| \mathbf{Aw} \|_2^2 + 2 \| \mathbf{w} \|_2^2)$$

e portanto se $\mathbf{w} \neq \mathbf{0}$, há uma contradição para α negativo pequeno. □

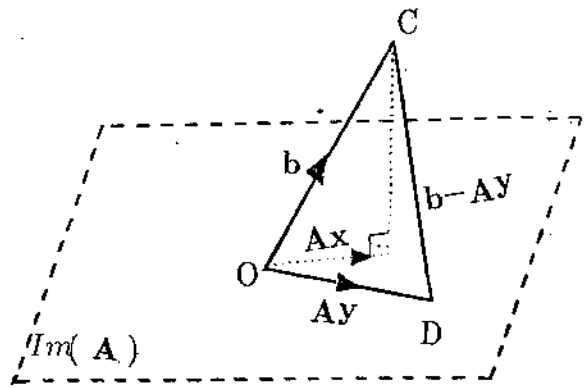


Fig 2.1
Aqui, $\mathbf{x} \in \mathbf{X}$ e $\mathbf{y} \in \mathbf{X}$. O método dos quadrados mínimos consiste em encontrar um D, tal que DC é mínimo.

2)- $\mathbf{X}$ é convexo e tem um único elemento de norma mínima $\mathbf{x^o}$:

$$\| \mathbf{x^o} \|_2 < \| \mathbf{x} \|_2 \qquad \forall \mathbf{x} \in \mathbf{X} - \mathbf{x^o}$$

Provaremos primeiro que $\mathbf{X}$ tem um único elemento de norma mínima $\mathbf{x^o}$ e enunciaremos um teorema a ser usado nesta demonstração.

TEOREMA 3.1 *Se* $\mathbf{A}$ *é uma matriz* $m \times n$ *e* $\mathrm{b} \in Im(\mathbf{A})$, *então a equação linear* $\mathbf{Ax} = \mathrm{b}$ *tem uma única solução* $\mathbf{x^o}$ *em* $Ker(\mathbf{A})^{\perp}$, *e qualquer solução pode ser escrita como* $\mathbf{x^o} + \mathbf{y}$, *onde* $\mathbf{y} \in Ker\mathbf{A}$.

Prova do teorema:
Seja $\mathbf{A} : \mathbf{R} \rightarrow \mathbf{S}$ e a equação linear :

$$\mathbf{Ax} = \mathrm{b} \qquad (I)$$

para um dado $\mathrm{b} \in \mathbf{S}$. É claro que esta equação pode ter uma solução se e somente se, $\mathrm{b} \in \mathrm{Im}(\mathbf{A})$. Então se $\mathrm{b} \in \mathrm{Im}(\mathbf{A})$, suponhamos que $\mathbf{x^o}$ seja uma solução. Então $\mathbf{x^o} + \mathbf{y}$ é também uma solução para qualquer $\mathbf{y} \in Ker(\mathbf{A})$, já que:

$$\mathbf{A}(\mathbf{x^o} + \mathbf{y}) = \mathbf{Ax^o} = \mathrm{b}$$

Em geral, qualquer uma das infinitas soluções de (I), podem ter a mesma regra de $\mathbf{x^o}$. No entanto, se $\mathbf{x^o}$ é restrita a $\mathrm{Ker}(\mathbf{A})^{\perp}$ então ele é único. Para isto, seja $\mathbf{x_1^o}$ e $\mathbf{x_2^o}$ duas soluções quaisquer de (I), e seja,

$$\mathbf{x_i^o} = \mathbf{u_i} + \mathbf{v_i} \quad , \qquad i = 1, 2$$

a decomposição de $\mathbf{x_i^o}$ nas partes $\mathbf{u_i}$ em $\mathrm{Ker}(\mathbf{A})$ e $\mathbf{v_i}$ em $\mathrm{Ker}(\mathbf{A})^{\perp}$. Então:

$$\mathbf{Ax_i^o} = \mathbf{Av_i} = \mathrm{b}$$

tal que,

$$\mathbf{A}(\mathbf{v_1} - \mathbf{v_2}) = 0$$

então $\mathbf{v_1}$ - $\mathbf{v_2}$ está no $\mathrm{Ker}(\mathbf{A})$ e no $\mathrm{Ker}(\mathbf{A})^{\perp}$, e portanto ele deve ser 0. Então $\mathbf{v_1} = \mathbf{v_2}$, e mostramos que o componente em $\mathrm{Ker}(\mathbf{A})^{\perp}$ de qualquer solução é único. □

Seja então $\mathbf{x^o} \in Ker(\mathbf{A})^{\perp}$. Então $\mathbf{x^o}$ é único. Seja $\mathbf{x} = \mathbf{x^o} + \mathbf{y}$, qualquer outra solução. Então:

$$\| \mathbf{x^o} + \mathbf{y} \|_2 = \| \mathbf{x^o} \|_2 + \| \mathbf{y} \|_2$$

e

$$\| \mathbf{x^o} \|_2 \leq \| \mathbf{x^o} + \mathbf{y} \|_2$$

e portanto:

$$\| \mathbf{x^o} \|_2 < \| \mathbf{x} \|_2, \ \forall \mathbf{x} \in \mathbf{X} - \mathbf{x^o} . \qquad \square$$

Provaremos agora que $\mathbf{X}$ é convexo.
Dado $\mathbf{x_1}$ *e* $\mathbf{x_2} \in \mathbf{X}$, queremos mostrar que:

$$\mathbf{x} = \alpha \mathbf{x_1} + (1 - \alpha) \mathbf{x_2} \in \mathbf{X};$$

$$\forall \alpha \in [0, 1]$$

Seja $\mathbf{x^o}$ o único elemento de norma mínima de $\mathbf{X}$. Então $\mathbf{x_i} = \mathbf{x^o} + \mathbf{y_i}$, $i = 1, 2$, onde $\mathbf{y_i} \in$ Ker($\mathbf{A}$).Então

$$\| \mathbf{Ax} - \mathbf{b} \|_2 =$$

$$= \| \mathbf{A}(\alpha \mathbf{x_1} + (1 - \alpha) \mathbf{x_2}) - \mathbf{b} \|_2 =$$

$$= \| \alpha \mathbf{A}(\mathbf{x^o} + \mathbf{y_1}) + (1 - \alpha) \mathbf{A}(\mathbf{x^o} + \mathbf{y_2} - \mathbf{b} \|_2 =$$

$$= \| \mathbf{Ax^o} - \mathbf{b} \|_2$$

e portanto $\| \mathbf{Ax} - \mathbf{b} \|_2 \in \mathbf{X}$. $\square$

3.2 Solução

3.2.1 Solução por equações normais

Seja o modelo de regressão:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

onde

$\mathbf{y}$ é o vetor $n \times 1$ de variáveis resposta

$\mathbf{X}$ é a matriz $n \times p$ de variáveis preditoras

β é o vetor $p \times 1$ de parâmetros do modelo

ϵ é o vetor erro n×1, onde $E(\ \epsilon_i\ ) = \mathbf{0}$ e $Var(\ \epsilon_i\ ) = \sigma^2$, i=1,...,n

Queremos encontrar β tal que minimize $\epsilon = \mathbf{y} - \mathbf{X}\ \beta$. Uma forma de fazermos isto é através de quadrados mínimos o que consiste em minimizar:

$$\| \mathbf{y} - \mathbf{X}\beta \| = \epsilon^T\epsilon = (\ \mathbf{y} - \mathbf{X}\beta\ )^T\ (\ \mathbf{y} - \mathbf{X}\beta\ ) =$$

$$= \mathbf{y}^\mathbf{T}\mathbf{y} - 2\beta^T\mathbf{X}^\mathbf{T}\mathbf{y} + \beta^T\mathbf{X}^\mathbf{T}\mathbf{X}\beta$$

Se derivarmos esta equação em relação a β, obteremos:

$$\epsilon^T\epsilon = -2\mathbf{X}^\mathbf{T}\mathbf{y} + 2\mathbf{X}^\mathbf{T}\mathbf{X}\beta$$

assim, igualando a zero obteremos:

$$\mathbf{X}^\mathbf{T}\mathbf{X}\beta = \mathbf{X}^\mathbf{T}\mathbf{y}$$

o qual é chamado de equações normais.

Se $\mathbf{X}$ tiver posto p, $\mathbf{X}^\mathbf{T}\mathbf{X}$ é definida positiva e portanto não singular; e esta equação terá uma única solução:

$$\beta^o = (\ \mathbf{X}^\mathbf{T}\mathbf{X}\ )^{-1}\ \mathbf{X}^\mathbf{T}\mathbf{y}.$$

Já que $\mathbf{X}\beta^o = \mathbf{X}(\ \mathbf{X}^\mathbf{T}\mathbf{X}\ )^{-1}\ \mathbf{X}^\mathbf{T}\mathbf{y} = \mathbf{P}\mathbf{y}$, $\mathbf{P}$ é a transformação linear representando a projeção ortogonal do espaço Euclidiano n-dimensional

En , em $Im(\ \mathbf{X}\ )$. Como havíamos visto, similarmente (**I** - **P**) representa a projeção ortogonal de **En** no complemento ortogonal , $Im(\ \mathbf{X}\ )^{\perp}$. Então podemos escrever **y** como $\mathbf{y} = \mathbf{Py} + (\ \mathbf{I} - \mathbf{P}\ )\mathbf{y}$ onde $\mathbf{Py} \in$ Im(**X**) e (**I** - **P**)y $\in$ Im(**X**)$^{\perp}$.

Agora se **X** não tiver posto p, $\mathbf{X^TX}$ também não terá, e a equação normal não poderá ser resolvido por : $\beta^o = (\ \mathbf{X^TX}\ )^{-1}\ \mathbf{X^Ty}$ já que $\mathbf{X^TX}$ não tem inversa e a equação normal não terá uma única solução. Ele terá infinitamente muitas soluções.

3.2.2 Solução por Decomposição de Valores Singulares

Seja $\mathbf{A} \in\ \mathbf{M_{m,n}}$, $\mathbf{b} \in\ \Re^n$ e considere a equação linear $\mathbf{Ax} = \mathbf{b}$. No caso em que o posto de **A** não é completo, ele terá infinitamente muitas soluções e queremos encontrar uma solução $\mathbf{x^o}$, tal que:

$$\|\ \mathbf{Ax^o} - \mathbf{b}\ \|_2 = min_{\mathbf{x}\in\Re^n}\ \|\ \mathbf{Ax} - \mathbf{b}\ \|_2$$

TEOREMA 3.2 *Se* $\mathbf{A} = \mathbf{PDQ^T}$ *, é a decomposição em valores singulares de* **A** *, então a solução por quadrados mínimos da equação linear* $\mathbf{Ax} = \mathbf{b}$ *, é dado por:*

$$\mathbf{x^o} = \sum_{i=1}^{r} \frac{(\ \mathbf{p_i^Tb}\ )}{\sigma_i}\ \mathbf{q_i}$$

onde r=posto(**A** *)*

Prova: Temos que:

$$\|\ \mathbf{Ax} - \mathbf{b}\ \|_2^2 = \|\ \mathbf{PDQ^Tx} - \mathbf{b}\ \|_2^2$$

denotando-se $\mathbf{y} = \mathbf{Q^Tx}$,

$$\|\ \mathbf{PDy} - \mathbf{b}\ \|_2^2 = \|\ \mathbf{Dy} - \mathbf{P^Tb}\ \|_2^2 =$$

$$= (\ \mathbf{Dy} - \mathbf{P^Tb}\ )^T\ (\ \mathbf{Dy} - \mathbf{P^Tb}\ ) =$$

$$= \sum_{i=1}^{r} (\sigma_i y_i - p_i^T b)^2 + \sum_{i=r+1}^{n} (p_i^T b)^2$$

onde $\mathbf{p_i}$ e a i-ésima coluna da matriz $\mathbf{P}$ e $\mathbf{y_i}$ é o i-ésimo elemento do vetor $\mathbf{y}$.

É claro que para minimizarmos esta soma de quadrados com o vetor $\mathbf{y^o}$ de norma mínima, devemos estabelecer:

$$y_i^o = \begin{cases} p_i^T b \ / \ \sigma_i & \text{para i=1,...,r} \\ 0 & \text{para 1=r+1,...,n} \end{cases}$$

ou seja,

$$x^o = Qy^o = \sum_{i=1}^{r} y_i^o q_i = \sum_{i=1}^{r} \frac{p_i^T b}{\sigma_i} q_i,$$

onde $\mathbf{q_i}$ é a i-ésima coluna de $\mathbf{Q}$.

3.2.3 Solução por Inversa de Moore Penrose

Seja a equação linear:

$$\mathbf{Ax = b}$$

onde $\mathbf{A} \in \mathbf{M_{m,n}}$, $\mathbf{b} \in \Re^m$ e desejamos resolver a equação para $\mathbf{x} \in \Re^n$. Sabemos que existe uma solução para esta equação se, e somente se, $\mathbf{b} \in Im(\mathbf{A})$. Seja $\mathbf{x^o = A^+ b}$, então $\mathbf{x^o}$ é uma das soluções da equação acima pois, se $\mathbf{Ax = b}$, tal que $\mathbf{b} \in Im(\mathbf{A})$, então como $\mathbf{A = AA^+A}$, temos que:

$$\mathbf{Ax = b}$$

$$\Rightarrow \quad \mathbf{AA^+ (Ax) = b}$$

$$\Rightarrow \quad \mathbf{AA^+ b = b}$$

$$\Rightarrow \quad \mathbf{Ax^o = b}$$

e será a única solução se, e somente se, $\mathbf{A}$ é não singular.

Considere agora o caso em que $\mathbf{b} \notin Im(\mathbf{A})$. Agora não existe solução para a equação $\mathbf{Ax} = \mathbf{b}$, mas existe uma solução de quadrados mínimos de $\mathbf{Ax} = \mathbf{b}$, que será o vetor $\mathbf{x^o}$ para o qual $\| \mathbf{Ax} - \mathbf{b} \|$ é minimizada.

$$\| \mathbf{Ax^o} - \mathbf{b} \|_2 = min_{\mathbf{x} \in \Re^n} \| \mathbf{Ax} - \mathbf{b} \|_2 \qquad (1)$$

Então para qualquer $\mathbf{A} \in \mathbf{M_{m,n}}$ e qualquer $\mathbf{b} \in \Re^m$, a melhor equação aproximada de $\mathbf{Ax} = \mathbf{b}$ é um vetor $\mathbf{x^o} \in \Re^n$ de norma mínima euclidiana, satisfazendo (1).

TEOREMA 3.3 *O vetor* $\mathbf{x^o} = \mathbf{A^+b}$ *é a melhor solução aproximada de* $\mathbf{Ax} = \mathbf{b}$.

Prova:Para qualquer $\mathbf{x} \in \Re^n$, podemos escrever:

$$\mathbf{Ax} - \mathbf{b} = \mathbf{A}(\mathbf{x} - \mathbf{A^+b}) + (\mathbf{I} - \mathbf{AA^+})(-\mathbf{b})$$

E pelo teorema-2.18 temos que $\mathbf{I} - \mathbf{AA^+}$ é um projetor ortogonal em $(Im(\mathbf{A}))^{\perp}$, então:

$$\mathbf{A}(\mathbf{x} - \mathbf{A^+b}) \in Im(\mathbf{A})$$

e

$$(\mathbf{I} - \mathbf{AA^+})(-\mathbf{b}) \in (Im(\mathbf{A}))^{\perp}$$

e portanto são vetores mutuamente ortogonais e a extensão do Teorema de Pitágoras implica que:

$$\| \mathbf{Ax} - \mathbf{b} \|_2^2 = \| \mathbf{A}(\mathbf{x} - \mathbf{A^+b}) \|_2^2 + \| (\mathbf{I} - \mathbf{AA^+})(-\mathbf{b}) \|_2^2 =$$

$$= \| \mathbf{A}(\mathbf{x} - \mathbf{x^o}) \|_2^2 + \| \mathbf{Ax^o} - \mathbf{b} \|_2^2 \qquad (2)$$

e como $\| \mathbf{A}(\mathbf{x} - \mathbf{x^o}) \|_2^2 \geq \mathbf{0}$, se tomarmos $\mathbf{x} = \mathbf{x^o}$,

$$\| \mathbf{Ax^o} - \mathbf{b} \|_2^2 = min_{\mathbf{x} \in \Re^n} \| \mathbf{Ax} - \mathbf{b} \|_2$$

Agora, mostraremos que $\| \mathbf{x_1} \|_2 \geq \| \mathbf{x^0} \|_2$ para qualquer vetor $\mathbf{x_1} \in \Re^n$ que minimiza $\| \mathbf{Ax} - \mathbf{b} \|_2$.

Se substituirmos $\mathbf{x_1}$ na equação (2) no lugar de $\mathbf{x}$, temos que $\| \mathbf{Ax_1} - \mathbf{b} \|_2^2 = \| \mathbf{Ax^o} - \mathbf{b} \|_2^2$ e $\mathbf{A}(\mathbf{x_1} - \mathbf{x^o}) = \mathbf{0}$, e portanto $\mathbf{x_1} - \mathbf{x^o} \in Ker(\mathbf{A})$; e $\mathbf{x^o} = \mathbf{A^+ b} \in Im(\mathbf{A^+})$ e como $Im(\mathbf{A^+}) = Im(\mathbf{A^T})$ (teorema 2.17), $\mathbf{x^o} \subset Im(\mathbf{A^T})$ e como $Ker(\mathbf{A}) \perp Im(\mathbf{A^T})$, temos que $\mathbf{x^o} \perp (\mathbf{x_1} - \mathbf{x^o})$. E portanto:

$$\mathbf{x_1} = \mathbf{x^o} + \mathbf{x_1} - \mathbf{x^o}$$

$$\Rightarrow \quad \| \mathbf{x_1} \|_2^2 = \| \mathbf{x^o} \|_2^2 + \| \mathbf{x_1} - \mathbf{x^o} \|_2^2 \geq \| \mathbf{x^o} \|_2^2 . \quad \square$$

4 CAPÍTULO IV

4.1 Número de Condição

O número de condição é uma medida de sensibilidade do vetor de solução $\mathbf{x}$ do sistema linear $\mathbf{Ax} = \mathbf{b}$ para pequenas mudanças nos elementos de $\mathbf{b}$ e $\mathbf{A}$. Seguindo o procedimento de Forsythe e Moler (1967) dividiremos esta seção em duas partes, sendo que na primeira parte iremos supor que o vetor $\mathbf{b}$ está sujeito à incerteza e na segunda parte iremos supor que a matriz $\mathbf{A}$ está sujeita à incerteza e assim veremos o efeito desta incerteza (*em* $\mathbf{A}$ *e* $\mathbf{b}$) no vetor de solução $\mathbf{x}$.

1)- Suponhamos então que o vetor $\mathbf{b}$ esteja sujeito a uma incerteza e seja $\mathbf{b} + \delta\mathbf{b}$ um vetor "próximo" de $\mathbf{b}$ e que $\mathbf{A}(\mathbf{x} + \delta\,\mathbf{x}) = \mathbf{b} + \delta\,\mathbf{b}$. Como $\mathbf{Ax} = \mathbf{b}$, temos que $\delta\,\mathbf{x} = \mathbf{A}^{-1}\delta\mathbf{b}$ e portanto:

$$\| \delta\mathbf{x} \| \leq \| \mathbf{A}^{-1} \| \| \delta\mathbf{b} \|$$

e

$$\| \mathbf{b} \| \leq \| \mathbf{A} \| \| \mathbf{x} \|$$

e então,

$$\| \delta\mathbf{x} \| \| \mathbf{b} \| \leq \| \mathbf{A} \| \| \mathbf{A}^{-1} \| \| \mathbf{x} \| \| \delta\mathbf{b} \|$$

$$\frac{\| \delta\mathbf{x} \|}{\| \mathbf{x} \|} \leq \| \mathbf{A} \| \| \mathbf{A}^{-1} \| \frac{\| \delta\mathbf{b} \|}{\| \mathbf{b} \|}$$

Define-se como o número de condição de uma matriz não singular $\mathbf{A}$ o número $\| \mathbf{A} \| \| \mathbf{A}^{-1} \| = \mathrm{k}(\mathbf{A})$, então:

$$\frac{\| \delta\mathbf{x} \|}{\| \mathbf{x} \|} \leq K(\mathbf{A}) \frac{\| \delta\mathbf{b} \|}{\| \mathbf{b} \|} \quad (3.3.1.1)$$

onde $\frac{\| \delta\mathbf{b} \|}{\| \mathbf{b} \|}$ pode ser interpretado como a medida de incerteza relativa no vetor $\mathbf{b}$ e $\frac{\| \delta\mathbf{x} \|}{\| \mathbf{x} \|}$ pode ser interpretado como a medida de incerteza relativa no vetor $\mathbf{x}$ devido à incerteza em $\mathbf{b}$. A equação (3.3.1.1) nos

mostra que o número de condição k(**A**) nos dá um limite para a medida de incerteza relativa no vetor de solução **x** ($\| \delta \mathbf{x} \| / \| \mathbf{x} \|$) dada a medida de incerteza em **b**.

2)- Suponhamos que a matriz **A** esteja sujeito à incerteza, então:

$$\mathbf{x} + \delta\mathbf{x} = (\mathbf{A} + \delta\mathbf{A})^{-1} \mathbf{b}$$

e

$$\delta\mathbf{x} = (\mathbf{A} + \delta\mathbf{A})^{-1} \mathbf{b} - \mathbf{A}^{-1} \mathbf{b} = [(\mathbf{A} + \delta\mathbf{A})^{-1} - \mathbf{A}^{-1}] \mathbf{b} =$$

$$= [\mathbf{A}^{-1}\mathbf{A} (\mathbf{A} + \delta\mathbf{A})^{-1} - \mathbf{A}^{-1} (\mathbf{A} + \delta\mathbf{A})(\mathbf{A} + \delta\mathbf{A})^{-1}] \mathbf{b} =$$

$$= [\mathbf{A}^{-1} (\mathbf{A} - (\mathbf{A} + \delta\mathbf{A}))(\mathbf{A} + \delta\mathbf{A})^{-1}] \mathbf{b} =$$

$$= - \mathbf{A}^{-1} \delta\mathbf{A} (\mathbf{A} + \delta\mathbf{A})^{-1} \mathbf{b} =$$

$$= - \mathbf{A}^{-1} \delta\mathbf{A} (\mathbf{x} + \delta\mathbf{x})$$

e portanto:

$$\| \delta\mathbf{x} \| \leq \| \mathbf{A}^{-1} \| \| \mathbf{A} \| \frac{\| \delta\mathbf{A} \|}{\| \mathbf{A} \|} \| \mathbf{x} + \delta\mathbf{x} \|$$

$$\frac{\| \delta\mathbf{x} \|}{\| \mathbf{x} + \delta\mathbf{x} \|} \leq k(\mathbf{A}) \frac{\| \delta\mathbf{A} \|}{\| \mathbf{A} \|}$$

e a incerteza de **x** em relação à $\mathbf{x} + \delta\mathbf{x}$ é limitada pela incerteza relativa de **A** multiplicada pelo número de condição de **A**.

Assim o número de condição depende da norma usada, e no caso da norma definida por: $|||\ \mathbf{A}\ ||| = max_{\|\ \mathbf{x}\ \| = 1}\ \|\ \mathbf{Ax}\ \|$ (Apêndice A , Definição $A.3$).

$$|||\ \mathbf{A}\ ||| = |||\ \mathbf{PDQ^T}\ ||| = |||\ \mathbf{D}\ ||| = \sqrt{\lambda_{max}} = \sigma_{max}$$

e

$$|||\ \mathbf{A^{-1}}\ ||| = |||\ \mathbf{QD^{-1}P^T}\ ||| = |||\ \mathbf{D^{-1}}\ ||| = \sqrt{\lambda_{min}^{-1}} = \sigma_{min}^{-1}$$

e portanto o número de condição k(**A**) será dado por:

$$k(\ \mathbf{A}\ ) = \frac{\sigma_{max}}{\sigma_{min}} \geq 1 \quad (3.3.1.2)$$

e portanto se a matriz $\mathbf{X^TX}$ é singular , temos que $\lambda_{min} = 0$ e k(**X**) $\rightarrow \infty$ e se as variáveis da matriz preditora $\mathbf{X}$=[$\mathbf{x_1}$...$\mathbf{x_p}$] vistas como vetores forem ortonormais, ou seja, $\mathbf{X^TX} = \mathbf{I}$, então todos os auto-valores de $\mathbf{X^TX}$ são iguais a um e k(**X**) = 1. Assim, a variação do número de condição é de $[\ 1, \infty\ )$, sendo que quanto maior o número de condição k(x), a matriz **X** é mais mal condicionada.

4.2 Multicolinearidade

Considere o modelo de regressão (2.2.1) . Dizemos que existe uma multicolinearidade exata quando o posto de **X** é menor do que p, ou seja, as colunas de **X** denotada por $\mathbf{x_1}$, . . . , $\mathbf{x_p}$ estão em um subespaço de dimensão menor do que p, ou ainda, existem constantes não nulas a_i , i=1,..,p , tal que :

$$\sum_{i=1}^{p} a_i\ \mathbf{x_i} = \mathbf{0}$$

Agora quando há constantes não nulas a_i tal que:

$$\sum_{i=1}^{p} a_i\ \mathbf{x_i} \approx \mathbf{0}$$

dizemos que **X** está "próximo de multicolinearidade" , onde $\approx$ denota aproximadamente igual.

Notamos que a colinearidade é relacionado às características especificamente da matriz do modelo $\mathbf{X}$ e não de aspectos estatísticos do modelo de regressão $\mathbf{y} = \mathbf{X}\,\beta + \epsilon$. Se escrevermos a matriz de variáveis preditora $\mathbf{X}$ em termos de decomposição em valores singulares , teremos:

$$\mathbf{y} = \mathbf{X}\,\beta + \epsilon = \mathbf{PDQ^T}\beta + \epsilon =$$

$$= \mathbf{X^*}\beta^* + \epsilon$$

onde

$$\mathbf{X^*} = \mathbf{PD} \;\; e \;\; \beta^* = \mathbf{Q^T}\beta$$

$$\mathbf{X^TX} = \mathbf{QDP^TPDQ^T} = \mathbf{QD^2Q^T} \quad (3.1.2)$$

Quando a matriz produto (3.1.2) é singular, a sua inversa não existe , e é o caso de uma multicolinearidade exata. Neste caso existe pelo menos um auto-valor igual a zero, e isto é fácil de ser visto se colocarmos a matriz produto na forma de decomposição em valores singulares, pois:

$$(\,\mathbf{X^TX}\,)^{-1} = (\,\mathbf{QD^2Q^T}\,)^{-1}$$

$$= \mathbf{Q^TD^{-2}Q}$$

quando $\mathbf{D^{-2}}$ existe. Sabemos que como $\mathbf{Q} \in \mathbf{M_p}$ é uma matriz ortogonal (Apêndice B) a sua inversa sempre existe e é dado por $\mathbf{Q^T}$. Como $\mathbf{D^{-2}}$ é uma matriz diagonal, cujos elementos da diagonal são as inversas dos auto-valores de $\mathbf{X^TX}$, $\mathbf{X^TX}$ é não singular se e somente se $\lambda_i \neq 0$, $\forall\ \lambda_i$ i=1,...,p. Suponha que o posto de $\mathbf{X}$ = p - k e k > 1. Então existe k auto-valores nulos e k autovetores $\mathbf{v_1}$, . . . , $\mathbf{v_k}$ associados a estes k auto-valores tal que $\mathbf{Xv_i} = \mathbf{0}$, pois $\mathbf{X^TXv_i} = \lambda_i\mathbf{v_i}$, ou seja, $\mathbf{v_i^TX^TXv_i} = \lambda_i\mathbf{v_i^Tv_i}$ e como $\mathbf{v_i}$ é ortonormal, $\mathbf{v_i^TX^TXv_i} = \lambda_i$ e como estes k auto-valores são iguais a zero, para cada um destes k auto-vetores teremos a relação : $\mathbf{v_i^TX^TXv_i} = 0 = (\,\mathbf{Xv_i}\,)^T(\,\mathbf{Xv_i}\,)$. Portanto $\mathbf{Xv_i} = \mathbf{0}$, ou seja, todas as linhas de $\mathbf{X}$ são ortogonais a $\mathbf{v_i}$ i=1,...,k . Agora, quando a matriz $\mathbf{X^TX}$ está "próximo de multicolinearidade", $\mathbf{X^TX}$ é "quase singular" e existem auto-valores pequenos em relação ao maior auto-valor causando mal condicionamento da matriz

(Ver 3.3.1.2) .

Suponhamos agora que a matriz $\mathbf{X} = [\, \mathbf{x_1}...\mathbf{x_p} \,]$ possui colunas ortogonais, ou seja, $\mathbf{x_i^T} * \mathbf{x_j} = 0$ para i$\neq$j , assim a matriz não tem problemas de multicolinearidade. Mas mesmo assim esta matriz pode ser mal condicionada, pois dado por exemplo, que o maior elemento da matriz $\mathbf{X^TX}$ é igual a 10^4 e o menor elemento é igual a 10^{-4} , então como $\mathbf{X^TX}$ é uma matriz diagonal e os próprios elementos da diagonal de $\mathbf{X^TX}$ são os auto-valores de $\mathbf{X^TX}$. Assim o número de condição de $\mathbf{X}$ seria: $K(\mathbf{X}) = \frac{\sqrt{10^4}}{\sqrt{10^{-4}}} = 10^4$ e a matriz é mal condicionada.

4.3 Ortogonalidade

Dizemos que a matriz $\mathbf{X}$ tem regressores ortogonais se $\mathbf{X^TX} = \mathbf{I}$. Quando isto não ocorre, a matriz dos dados podem ser transformados para o espaço de variáveis ortogonais. Aqui vamos colocar a matriz preditora $\mathbf{X}$ na forma de decomposição em valores singulares e transformaremos em uma matriz $\mathbf{Z} = [\, \mathbf{z_1}...\mathbf{z_n} \,]$, tal que $\mathbf{z_i^T} * \mathbf{z_j} = 0$ para i $\neq$ j, ou seja, as colunas da matriz $\mathbf{Z}$, vistas como vetores são ortogonais e mostraremos que apesar disto a variância do parâmetro estimado α pode ser grande.

Seja $\mathbf{X} = \mathbf{PDQ^T}$ a decomposição em valores singulares de $\mathbf{X}$. Então ,

$$\mathbf{X^TX} = \mathbf{QD^2Q^T}$$

e portanto,

$$\mathbf{Q^TX^TXQ} = \mathbf{D^2} = diag\,[\, \lambda_1 \,, ..., \, \lambda_p \,] \quad (3.2.1)$$

Consideraremos o modelo de regressão (2.2.1):

$$\mathbf{y} = \mathbf{X}\beta + \epsilon \,,$$

e seja:

$$\mathbf{Z} = \mathbf{XQ} \;\; e \;\; \alpha = \mathbf{Q^T}\beta$$

então:

$\mathbf{X} = \mathbf{ZQ^T}$ *e* $\beta = \mathbf{Q}\alpha$ *pois* $\mathbf{Q}$ *e uma matriz ortogonal.*

então como:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon\,,$$

temos que:

$$\mathbf{y} = \mathbf{ZQ^TQ}\alpha + \epsilon$$

e

$$\mathbf{y} = \mathbf{Z}\alpha + \epsilon$$

onde $\mathbf{Z} = \{\ \mathbf{z_1}\ ,...,\ \mathbf{z_p}\ \}$ e $\mathbf{z_1}$, ..., $\mathbf{z_p}$ vistos como vetores são ortogonais O sistema de equações normais é dado por:

$$(\ \mathbf{Z^TZ}\ )\ \hat{\alpha} = \mathbf{Z^Ty}$$

e por (3.2.1) , temos que:

$$\hat{\alpha} = \mathbf{D^{-2}Z^Ty}$$

e a matriz de variância-covariância de $\hat{\alpha}$ é dado por:

$$\mathbf{V}(\ \hat{\alpha}\ ) = \mathbf{V}(\ \mathbf{D^{-2}Z^Ty}\ ) = \mathbf{D^{-2}Z^TV}(\ \mathbf{y}\ )\ \mathbf{ZD^{-2}}$$

$$= \sigma^2\mathbf{D^{-2}Z^TZD^{-2}} \qquad (3.2.2)$$

$$= \sigma^2\mathbf{D^{-2}D^2D^{-2}}$$

$$= \sigma^2\mathbf{D^{-2}}$$

ou seja, $\mathbf{V}(\ \hat{\alpha_i}\ ) = \frac{\sigma^2}{\lambda_i}$, onde σ^2 é a variância do erro e λ_i é o i-ésimo auto-valor de $\mathbf{X^TX}$. Assim, podemos ver que apesar da matriz $\mathbf{Z}$ ter colunas ortogonais, dependendo do tamanho do auto-valor de $\mathbf{X^TX}$ a variância de $\hat{\alpha}$ pode ser grande. Portanto a multicolinearidade explica apenas uma parte do mal condicionamento , pois aqui os vetores $\mathbf{z_i}$ são ortogonais mas a $\mathbf{V}(\ \hat{\alpha}\ )$ pode ser grande.

4.4 Decomposição da Variância do Coeficiente de Regressão

Nesta seção iremos usar a decomposição em valores singulares para podermos decompor a variância de cada beta estimado em termos de elementos da matriz de auto-vetores ortonormais de $\mathbf{X^T X}$($\mathbf{Q}$) e dos auto-valores de $\mathbf{X^T X}$ para uma melhor visão e compreensão sobre a magnitude desta variância.

Consideraremos o modelo de regressão (2.2.1):

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

e seja $\mathbf{b}$ o estimador de quadrados mínimos de β , então $\mathbf{V}(\mathbf{b}) = \sigma^2 (\mathbf{X^T X})^{-1}$. Usando a decomposição em valores singulares, temos que:

$$\mathbf{V}(\mathbf{b}) = \sigma^2(\mathbf{X^T X})^{-1} = \sigma^2(\mathbf{QDP^T PDQ^T})^{-1} =$$

$$= \sigma^2 \mathbf{QD^{-2}Q^T}$$

e portanto para o k-ésimo componente de b , teremos:

$$\mathbf{V}(\mathbf{b_k}) = \sigma^2 \sum_j \frac{q_{kj}^2}{\lambda_j} \quad ; \quad (3.3.1)$$

ou seja $\mathbf{V}(\mathbf{b_k})$ se decompõe em somas de componentes , onde cada componente é associado com um dos p auto-valores de $\mathbf{X^T X}$. Na equação (3.3.1) mesmo que a matriz $\mathbf{X^T X}$ tenha um autovalor pequeno em relação ao maior auto-valor, não necessariamente teremos $\mathbf{V}(\mathbf{b_k})$ "grande", pois se q_{kj}^2 por sua vez for menor ou estiver "próximo" do λ_j esta variância não será afetada.

Define-se (Belsley, 1980) k,j -ésima proporção da decomposição da variância (π_{kj}) como sendo a proporção da variância do k-ésimo coeficiente de regressão associado com o j-ésimo componente da sua decomposição em (3.3.1). Ou seja:

$$\pi_{kj} = \frac{\phi_{kj}}{\phi_k} \quad , \quad k, j = 1, ..., p$$

$$onde \ \ \phi_{kj} = \frac{q_{kj}^2}{\lambda_j} \ \ e \ \ \phi_k = \sum_{j=1}^{p} \phi_{kj} \ \ k = 1, ..., p$$

Para associarmos a concentração da variância de qualquer coeficiente de regressão em algum dos seus componentes como sinal de que a multicolinearidade está causando problemas, é necessário que pelo menos duas das variâncias do coeficiente de regressão tenha alta proporção na decomposição de variaĥcia associada com um dos λ_i

Para isto considere uma matriz de variáveis preditora , tal que: $\mathbf{X^T X = I}$. Então os seus auto-valores são todos iguais a um e $q_{kj} = 0$ para k $\neq$ j e $q_{kj} = 1$ para k=j. Assim,

$auto - valor \ associado$	$prop\mathbf{V(b_1)}$	$prop\mathbf{V(b_2)}$	...	$prop\mathbf{V(b_p)}$
λ_1	1	0	...	0
λ_2	0	1	...	0
.	.	.	...	.
λ_p	0	0	...	1

onde por prop $\mathbf{V(b_i)}$ entende-se : proporção da variância de $\mathbf{b_i}$. Mas é claro que esta alta concentração da proporção da decomposição de variância em um dos auto-valores associados não indica problema de multicolinearidade.

4.5 Uma Possível Solução para o Problema de Multicolinearidade

Pelo teorema de Gauss-Markov temos que $\mathbf{b} = (\ \mathbf{X^T X}\ )^{-1}\mathbf{X^T y}$ é um estimador não viciado de β com variância mínima (entre as funções lineares de $\mathbf{y}$ que sejam estimadores não viciados), mas mesmo assim vimos nas seções 3.4 e 3.5 que a variância do $\mathbf{b}$ pode ser "grande". Suponhamos por exemplo que a matriz de variáveis preditora $\mathbf{X} \in \mathbf{M_{n,p}}$ tenha colunas ortogonais e que o p-ésimo auto-valor seja aproximadamente zero, então pela seção 3.4, temos que:

$$\mathbf{V(\ b\ )} = \sigma^2 \mathbf{D^{-2}}$$

e

$$\mathbf{V(\ b_p\ )\ =\ \frac{\sigma^2}{\lambda_p}}$$

e como $\lambda_p \approx 0$, a variância do $\mathbf{b_P}$ é grande apesar da matriz de variáveis preditora ser ortogonal e o estimador possuir variância mínima entre todos os estimadores não viciados, ou seja, o problema existente aqui é um problema referente à matriz de variáveis preditora **X**. Neste caso se pudermos aumentar o tamanho do λ_p , a $\mathbf{V(\ b_p\ )}$ irá diminuir. No caso da matriz de variáveis preditora não ser ortogonal, a variância de $\mathbf{b_k}$ foi dado na seção 3.5 da seguinte forma:

$$\mathbf{V(\ b_k\ )\ =\ \sigma^2 \sum_j \frac{q_{k,j}^2}{\lambda_j}} \quad para\ k = 1, ..., p$$

ou seja, como havíamos comentado na seção 3.4 mesmo que haja algum λ_j próximo de zero, se o valor do $q_{k,j}^2$ referente à λ_j for aproximadamente igual à λ_j ou menor, a variância do $\mathbf{b_k}$ não será afetada. Caso isto não ocorra, se tivermos um λ_j próximo de zero , a $\mathbf{V(\ b_k\ )}$ será "grande". Assim iremos estudar uma maneira de aumentarmos a magnitude destes auto-valores.

Pelo teorema 2.14 se $\mathbf{X} \in \mathbf{M_{n,p}}$ e σ_i i=1,... ,p , forem os valores singulares de **X**, então se $\mathbf{X^*}$ for a matriz **X** com uma linha inclusa tal que $\mathbf{X^* \in M_{n+1,p}}$ e se σ_i^* i=1,...,p , forem os valores singulares de $\mathbf{X^*}$, temos que:

$$\sigma_1^* \geq \sigma_1 \geq \sigma_2^* \geq \sigma_2 \geq \ ... \ \geq \sigma_P^* \geq \sigma_P \geq 0$$

ou seja, se incluirmos uma nova linha na matriz **X**, cada auto-valor de $\mathbf{X^*} = \begin{bmatrix} \mathbf{X} \\ ..... \\ \mathbf{x_{n+1}^T} \end{bmatrix}$ vai ser maior ou igual ao correspondente auto-valor de **X**. Então uma maneira de resolvermos este problema poderia ser incluindo uma nova linha na matriz de variável preditora **X**. Mas como o número de condição de **X** é dado por $\mathrm{k}(\ \mathbf{X}\ ) = \frac{\sigma_1}{\sigma_P}$, se aumentarmos o valor do σ_p, mas o valor do σ_1 aumentar também, o número de condição pode não melhorar, ou seja, queremos aumentar o menor auto-valor mas queremos que

o maior auto-valor continue o mesmo. Aqui se tivermos mais de um auto-valor "pequeno" em relação ao maior auto-valor, por exemplo k auto-valores "pequenos" em relação ao maior auto-valor, então iremos querer aumentar o valor destes k auto-valores de tal maneira que não aumente o valor do maior auto-valor. Sabemos que se incluirmos uma nova linha na matriz de variáveis preditora $\mathbf{X}$, cada um dos auto valores de $\mathbf{X}^* = \begin{bmatrix} \mathbf{X} \\ ..... \\ \mathbf{x}_{n+1}^T \end{bmatrix}$ vai ser maior ou igual aos respectivos auto-valores da matriz $\mathbf{X}$ antes da inclusão desta linha, ou seja, o problema agora é encontrarmos uma nova linha de tal maneira que ela aumente os k auto-valores "pequenos" em relação ao maior auto-valor, mas que ao mesmo tempo ela não altere o valor do maior auto-valor.

Como foi visto na seção 3.1, quando existe multicolinearidade, $\mathbf{X}\mathbf{v}_i = \mathbf{0}$ para os auto-vetores $\mathbf{v}_i$ correspondentes aos auto-valores λ_i nulos, e assim as linhas da matriz $\mathbf{X}$ são ortogonais a $\mathbf{v}_i$, não sendo possível o cálculo de todos os coeficientes de regressão a partir das observações. Se a matriz estiver "próximo de multicolinearidade" a matriz $\mathbf{X}^T\mathbf{X}$ terá pelo menos um auto-valor "pequeno" em relação ao maior auto-valor.

Suponhamos então que $\mathbf{X}^T\mathbf{X}$ tem k auto-valores nulos, (k > 1), ou seja, o posto de $\mathbf{X}$ = p − k e portanto existem k auto-vetores $\mathbf{v_1}$,..., $\mathbf{v_k}$ associados a estes k auto-valores tal que $\mathbf{X}\mathbf{v}_i = \mathbf{0}$. Ou seja se nós adicionarmos uma nova linha na matriz $\mathbf{X}$ de tal maneira que esta linha não seja ortogonal a $\mathbf{v}_i$, teremos $\mathbf{X}\mathbf{v}_i \neq \mathbf{0}$, e uma maneira natural de escolhermos esta linha seria escolhé-la exatamente na direção de $\mathbf{v}_i$. Assim poderíamos resolver o problema de multicolinearidade adicionando uma nova linha de observações em cada uma das direções $\mathbf{v_1}$,...,$\mathbf{v_k}$. Esta escolha não é única e não poderemos resolver este problema se adicionarmos à matriz de observações um número de linhas inferior a k.

Agora, suponhamos que a matriz esteja "próximo de multicolinearidade" , ou seja , $\mathbf{X}^T\mathbf{X}$ tem pelo menos um auto-valor "pequeno" em relação ao maior auto-valor. Sabemos que se incluirmos uma nova linha na matriz $\mathbf{X}$, os auto-valores correspondentes a esta nova matriz vai ser maior ou igual aos respectivos auto-valores da matriz $\mathbf{X}$. Assim se incluirmos uma nova linha qualquer pode não ocorrer melhora alguma na matriz de va-

riáveis preditora $\mathbf{X}$. Agora se $\mathbf{X}$ tem k auto-valores "pequenos" em relação ao maior auto-valor, a estimação na direção dos auto-vetores correspondentes a estes k auto-valores não serão precisos. Assim, se colocarmos uma nova linha em cada uma das direções dos auto-vetores correspondentes a estes k auto-valores "pequenos" , poderemos melhorar as estimativas nestas direções.

Suponhamos que a matriz $\mathbf{X}$ tem um auto-valor "pequeno" λ_p na direção do auto-vetor $\mathbf{q_p}$ e como queremos aumentar o valor do auto-valor λ_p, na direção $\mathbf{q_p}$ iremos adicionar primeiro uma nova linha $\mathbf{x_{n+1}} = \mathbf{q_p}$. Assim, teremos:

$$\mathbf{y}^* = \mathbf{X}^*\beta + \epsilon^*$$

$$onde, \mathbf{y}^* = \begin{pmatrix} \mathbf{y} \\ .... \\ y_{n+1} \end{pmatrix} , \mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ .... \\ \mathbf{x_{n+1}^T} \end{pmatrix} e \ \epsilon^* = \begin{pmatrix} \epsilon \\ .... \\ \epsilon_{n+1} \end{pmatrix}$$

e vamos calcular o auto-valor referente a $\mathbf{q_p}$, então:

$$\mathbf{X^{*T}X^*} = \mathbf{X^TX} + \mathbf{x_{n+1}x_{n+1}^T} = \mathbf{X^TX} + \mathbf{q_pq_p^T}$$

e

$$\mathbf{X^{*T}X^*q_p} = \mathbf{X^TXq_p} + \mathbf{q_pq_p^Tq_p} =$$

$$= \lambda_p\mathbf{q_p} + \mathbf{q_p} =$$

$$= (\lambda_p + 1)\mathbf{q_p}$$

ou seja o valor do auto-valor de $\mathbf{q_p}$ se tornou $\lambda_P + 1$, assim pudemos aumentar o valor do λ_p em 1 . Neste caso $\mathbf{x_{n+1}}$ tem norma Euclidina 1 , pois $\mathbf{x_{n+1}^Tx_{n+1}} = \mathbf{q_p^Tq_p} = 1$, ou seja, se adicionarmos uma nova linha na direção de $\mathbf{q_p}$ com norma Euclidiana l , o valor do auto-valor λ_p deveria aumentar de l^2 . Para isto, seja então:

$$\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ .... \\ \mathbf{x_{n+1}^T} \end{pmatrix} \qquad onde \ \mathbf{x_{n+1}} = l\mathbf{q_p}.$$

Assim,

$$\mathbf{X^{*T}X^{*}} = \mathbf{X^{T}X} + \mathbf{x_{n+1}x_{n+1}^{T}} = \mathbf{X^{T}X} + l^2\mathbf{q_p q_p^{T}}$$

e

$$\mathbf{X^{*T}X^{*}q_p} = \mathbf{X^{T}Xq_p} + \mathbf{q_p q_p^{T}} =$$

$$= \lambda_p \mathbf{q_p} + l^2 \mathbf{q_p} =$$

$$= (\lambda_p + l^2) \mathbf{q_p}$$

ou seja:

$$\mathbf{X^{T}Xq_p} = \lambda_p \mathbf{q_p}$$

e

$$\mathbf{X^{*T}X^{*}q_p} = (\lambda_p + l^2) \mathbf{q_p}$$

e portanto o valor do p-ésimo auto-valor que era "pequeno" em relação ao maior auto-valor , λ_1, aumentou em l^2. Vamos verificar agora o que aconteceu com os demais auto-valores da matriz $\mathbf{X^{*T}X^{*}}$.

Seja então λ_1 o maior auto-valor e $\mathbf{q_1}$ o auto-vetor correspondente a λ_1 e como $\mathbf{q_1}$ é ortonormal a $\mathbf{q_p}$, teremos:

$$\mathbf{X^{*T}X^{*}q_1} = \mathbf{X^{T}Xq_1} + \mathbf{x_{n+1}x_{n+1}^{T}q_1} =$$

$$= \mathbf{X^{T}Xq_1} + l^2\mathbf{q_p q_p^{T}q_1}$$

$$= \mathbf{X^{T}Xq_1}$$

$$= \lambda_1 \mathbf{q_1}$$

ou seja:

$$\mathbf{X^T X q_1} \;=\; \lambda_1\, \mathbf{q_1}$$

e

$$\mathbf{X^{*T} X^* q_1} \;=\; \lambda_1\, \mathbf{q_1}$$

e assim, o fato de adicionarmos uma nova linha $\mathbf{x_{n+1}}$ na direção do p-ésimo auto-vetor com norma Euclidiana l, ficando assim a variável preditora $\mathbf{X^*} = \begin{bmatrix} \mathbf{X} \\ .... \\ \mathbf{x_{n+1}} \end{bmatrix}$, não alterou o valor do maior auto-valor (λ_1), que era exatamente o que precisávamos para aumentar o número de condição de $\mathbf{X}$.

Se proceguirmos da mesma maneira, teremos que todos os auto-valores de $\mathbf{X^T X}$ referentes a auto-vetores ortogonais a $\mathbf{q_p}$ continuam inalterados, ou seja:

$$\mathbf{X^T X q_i} \;=\; \lambda_i\, \mathbf{q_i} \qquad para\ i = 1, ..., p-1$$

e

$$\mathbf{X^{*T} X^* q_i} \;=\; \lambda_i\, \mathbf{q_i} \quad para\ i = 1, ..., p-1$$

Então o fato de acrescentarmos uma nova linha na direção do auto vetor do menor auto-valor, só modificará valor do menor auto-valor e os valores de todos os outros auto-valores de $\mathbf{X^{*T} X^*}$ continuarão inalterados.

No caso de termos mais de um auto-valor "pequeno" em relação ao maior auto-valor , deveremos acrescentar então tantas linhas quantas o número de auto-valores que forem pequenos em relação ao maior auto-valor. Assim, o número de condição k($\mathbf{X}$) irá aumentar a cada linha inclusa deste modo, podendo melhorar o mal condicionamento da matriz $\mathbf{X}$ e diminuir o valor da variância do $\mathbf{b}$, fazendo com que os estimadores se tornem mais estáveis.

4.6 Método de Estimação Recursivo para a Solução Apresentada na Seção 3.5

Muitas vezes na prática ocorrem problemas que envolvem um volume muito grande de dados , e como trabalhamos com um espaço de memória limitada em micro computadores, seria bastante útil desenvolvermos um método recursivo para a solução que foi encontrada na seção 3.5 .

Seja o modelo de regressão:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

onde

$\mathbf{y}$ é o vetor $n \times 1$ de variáveis resposta

$\mathbf{X}$ é a matriz $n \times p$ de variáveis preditoras

β é o vetor $p \times 1$ de parâmetros do modelo

ϵ é o vetor erro n×1, onde $E(\ \epsilon_i\ ) = \mathbf{0}$ e a $Var(\ \epsilon_i\ ) = \sigma^2$, para i=1,...,n .

Então , se acrescentarmos uma nova linha na matriz $\mathbf{X}$, $\mathbf{x_{n+1}^T}$, tal que $\mathbf{x_{n+1}^T} = l\mathbf{q_p^T}$ (onde $\mathbf{q_p}$ é o p-ésimo auto-vetor de $\mathbf{X^TX}$ associado ao p-ésimo auto-valor de $\mathbf{X^TX}$), teremos:

$$\mathbf{b_{n+1}} = \mathbf{b_n} + \mathbf{q_p} h_p \ ;$$

onde

$\mathbf{b_n}$ é o vetor de parâmetros estimadomoss no modelo de regressão com n dados

$\mathbf{b_{n+1}}$ é o vetor de parâmetros estimados no modelo de regressão com n+1 dados , onde a linha n+1 é da forma $l\mathbf{q_p}$

$$h_p = \frac{l}{\lambda_P + l^2}\ (\ y_{n+1} - \frac{l}{\sigma_p}\mathbf{p_p^T y}\ )$$

onde λ_p é o p-ésimo auto-valor de $\mathbf{X^T X}$, σ_p é o p-ésimo valor singular de $\mathbf{X}$, y_{n+1} é o n+1-ésimo valor do vetor $\mathbf{y_{n+1}} = \begin{bmatrix} \mathbf{y} \\ y_{n+1} \end{bmatrix}$ foiova:

Temos que:

$$\mathbf{b_n} = (\mathbf{X^T X})^{-1} \mathbf{X^T y}$$

usando-se a decomposição em valores singulares,

$$\mathbf{b_n} = [(\mathbf{PDQ^T})^T \mathbf{PDQ^T}]^{-1} (\mathbf{PDQ^T})^T \mathbf{y} =$$

$$= (\mathbf{QDP^T PDQ^T})^{-1} \mathbf{QDP^T y} =$$

como $\mathbf{P}$ é uma matriz ortogonal,

$$= (\mathbf{QD^2 Q^T})^{-1} \mathbf{QDP^T y} =$$

$$= \mathbf{QD^{-2} Q^T QDP^T y} =$$

como $\mathbf{Q}$ é uma matriz ortogonal,

$$= \mathbf{QD^{-1} P^T y}$$

$$= \mathbf{QD'^{-1} P'^T y}$$

onde

$\mathbf{Q^T}$ é a matriz dos auto-vetores ortogonais de $\mathbf{X^T X}$, ou seja , $\mathbf{Q_{pxp}} = \mathbf{q_1}, ..., \mathbf{q_p}$ onde $\mathbf{q_i}$ é um vetor $p \times 1$

$\mathbf{D'_{pxp}}$ é a matriz diagonal com valores singulares de $\mathbf{X}$

$\mathbf{P'_{nxp}} = \{\mathbf{p_1}, ..., \mathbf{p_p}\}$ onde, $\mathbf{p_i}$ é um vetor $n \times 1$ e $\mathbf{p_i} = \frac{1}{\sigma_i} \mathbf{X q_i}$

Seja $\mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ ..... \\ \mathbf{x_{n+1}} \end{bmatrix}$ onde $\mathbf{x_{n+1}^T} = \mathbf{q_p^T}$, $\mathbf{y_{n+1}} = \begin{bmatrix} \mathbf{y} \\ ..... \\ y_{n+1} \end{bmatrix}$

então : $\mathbf{b_{n+1}} = (\ \mathbf{Z^T Z}\ )^{-1}\ \mathbf{Z^T} y_{n+1}$

Vimos que os auto-vetores de $\mathbf{Z^T Z}$ continuam sendo os mesmos auto- vetores de $\mathbf{X^T X}$ e os auto-valores de $\mathbf{Z^T Z}$ são os mesmos de $\mathbf{X^T X}$, exceto o p-ésimo auto-valor que é dado por $\lambda_p + l^2$. Então,

$$\mathbf{b_{n+1}} = \mathbf{Q D_{n+1}^{-1} P_{n+1}^T} y_{n+1}$$

onde

$$\mathbf{D_{n+1}} = \begin{bmatrix} \sigma_1 & 0 & .... & 0 & 0 \\ 0 & \sigma_2 & .... & 0 & 0 \\ . & . & .... & . & . \\ . & . & .... & . & . \\ . & . & .... & . & . \\ 0 & 0 & .... & \sigma_{p-1} & 0 \\ 0 & 0 & .... & 0 & \sqrt{\lambda_p + l^2} \end{bmatrix}$$

$$\mathbf{y_{n+1}} = \begin{bmatrix} \mathbf{y} \\ ..... \\ y_{n+1} \end{bmatrix}$$

$$\mathbf{P_{n+1}} = \begin{bmatrix} \mathbf{p_1} & ... & \mathbf{P_{p-1}} & a_p\ \mathbf{p_p} \\ ... & ... & ... & ..... \\ 0 & ... & 0 & \frac{a_p l}{\sigma_p} \end{bmatrix}$$

pois, como os auto-ovetores de $\mathbf{X^T X}$ continuam inalterados e $\mathbf{p_i} = \frac{1}{\sigma_i}\ \mathbf{X q_i}$, temos que:

$$\mathbf{p_{i_{mosn+1}}} = \frac{1}{\sigma_i} \begin{bmatrix} x_{11} & x_{12} ... x_{1p} & \\ ... & ..... & ... \\ x_{n1} & x_{n2} ... x_{np} & \\ x_{n+1\ 1} & x_{n+1\ 2} ... x_{n+1\ p} & \end{bmatrix} \mathbf{q_i} \quad , \quad para\ i = 1, .., p-1$$

$$\mathbf{p}_{\mathbf{i}_{n+1}} = \begin{bmatrix} \mathbf{p_i} \\ \cdots\cdots \\ 0 \end{bmatrix}$$

e

$$\mathbf{p}_{\mathbf{p}_{n+1}\mathbf{k}} = k\frac{1}{\sqrt{\lambda_p + l^2}}\mathbf{X}\mathbf{q}_\mathbf{p} = a_p \begin{bmatrix} \mathbf{p_p} \\ \cdots\cdots \\ \frac{l}{\sigma_p} \end{bmatrix} \quad onde \quad a_p = \frac{\sigma_p}{\sqrt{\lambda_p + l^2}}$$

Agoramos , temos que:

$$\mathbf{D}^{-1}_{n+1}\mathbf{P}^{t}_{n+1} = \begin{bmatrix} \frac{1}{\sigma_1}\mathbf{p}_1^\mathbf{T} & 0 \\ \frac{1}{\sigma_2}\mathbf{P}_\mathbf{2}^\mathbf{T} & 0 \\ \cdots\cdots\cdots & \cdot \\ \frac{1}{\sigma_{P-1}}\mathbf{p}_{\mathbf{p}-1}^\mathbf{T} & 0 \\ \frac{\lambda_p}{\lambda_p + l^2}\frac{1}{\sigma_p}\mathbf{P}_\mathbf{p}^\mathbf{T} & \frac{l}{\lambda_p + l^2} \end{bmatrix}$$

$$\mathbf{D}^{-1}_{n+1}\mathbf{P}^{T}_{n+1} = \begin{bmatrix} \frac{1}{\sigma_1}\mathbf{p}_1^\mathbf{T} & 0 \\ \cdots\cdots & \cdot \\ \frac{1}{\sigma_{p-1}}\mathbf{p}_{\mathbf{p}-1}^\mathbf{T} & 0 \\ \frac{1}{\sigma_p}\mathbf{P}_\mathbf{p}^\mathbf{T} & \frac{l}{\lambda_p + l^2} \end{bmatrix} \begin{bmatrix} 1 & 0\ldots0 & 0 \\ 0 & 1\ldots0 & 0 \\ \cdot & \cdots\cdots & \cdot \\ 0 & 0\ldots1 & 0 \\ \frac{-l}{\sigma_p}\mathbf{p}_\mathbf{p}^\mathbf{T} & & 1 \end{bmatrix}$$

e

$$\begin{bmatrix} \mathbf{I}_{(p-1)\mathbf{x}n} & 0 \\ \frac{-l}{\sigma_P}\mathbf{p}_\mathbf{P}^\mathbf{T} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \frac{-l}{\sigma_p}\mathbf{P}_\mathbf{p}^\mathbf{T}\mathbf{y} + y_{n+1} \end{bmatrix}$$

e portanto:

$$\mathbf{b}_{n+1} = \mathbf{Q}\mathbf{D}^{-1}_{n+1}\mathbf{P}^{T}_{n+1}\mathbf{y}_{n+1}$$

$$= \mathbf{Q} \begin{bmatrix} \frac{1}{\sigma_1}\,\mathbf{p_1^T} & 0 \\ \cdots\cdots & . \\ \cdots\cdots & . \\ \cdots\cdots & . \\ \frac{1}{\sigma_{p-1}}\,\mathbf{p_p^T}_{-1} & 0 \\ \frac{1}{\sigma_p}\,\mathbf{p_p}^T & \frac{l}{\lambda_p + l^2} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \frac{-l}{\sigma_p}\,\mathbf{p_p^T y} + y_{n+1} \end{bmatrix}$$

$$= \mathbf{Q} \begin{bmatrix} \frac{1}{\sigma_1}\,\mathbf{p_1^T y} \\ . \\ . \\ . \\ \frac{1}{\sigma_{p-1}}\,\mathbf{p_p^T}_{-1}\mathbf{y} \\ \frac{1}{\sigma_p}\,\mathbf{p_p^T y} - \frac{l^2}{\lambda_p + l^2}\,\frac{1}{\sigma_p}\,\mathbf{p_p^T y} + \frac{l}{\lambda_p + l^2}\,y_{n+1} \end{bmatrix}$$

$$= \mathbf{Q} \begin{bmatrix} \frac{1}{\sigma_1}\,\mathbf{p_1^T y} \\ . \\ . \\ . \\ \frac{1}{\sigma_{p-1}}\,\mathbf{p_p^T}_{-1}\mathbf{y} \\ \frac{1}{\sigma_p}\,\mathbf{p_p^T y} \end{bmatrix} + \mathbf{Q} \begin{bmatrix} 0 \\ . \\ . \\ . \\ 0 \\ \frac{l}{\lambda_p + l^2}\left(y_{n+1} - \frac{l}{\sigma_p}\,\mathbf{p_p^T y} \right) \end{bmatrix}$$

$$= \mathbf{Q}\mathbf{D}^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{y} + \mathbf{q_p}\,h_p$$

$$= \mathbf{b_n} + \mathbf{q_p}\,h_p$$

onde , $h_p = \frac{l}{\lambda_p + l^2}\left(y_{n+1} - \frac{l}{\sigma_p}\,\mathbf{p_p^T y} \right)$

Se ao invés de adicionarmos à matriz $\mathbf{X}$ a linha $\mathbf{x_{n+1}} = l\,\mathbf{q_p}$ e adicionarmos uma linha na direção do auto-valor referente ao i-ésimo auto-vetor , teremos analogamente que:

$$\mathbf{b_{n+1}} = \mathbf{b_n} + \mathbf{q_i}\, h_i,$$

onde , $h_i = \frac{l}{\lambda_i + l^2}\,(\, y_{n+1} - \frac{l}{\sigma_i}\,\mathbf{p_i^T y}$

Mas, como os auto-valores estão ordenados em ordem decrescente e queremos naturalmente aumentar o valor do menor auto-valor , foi apresentado a forma recursiva ao adicionarmos a nova linha $\mathbf{x_{n+1}} = l\,\mathbf{q_p}$.

Se tivermos mais de um auto-valor pequeno e quisermos acrescentar um número de linhas maior do que um, ou seja, por exemplo as linhas $\mathbf{x_{n+1}^T} = l\,\mathbf{q_p^T}$ e $\mathbf{x_{n+2}^t} = l'\mathbf{q_{p-1}^T}$, teremos analogamente:

$$\mathbf{b_{n+1}} = \mathbf{b_n} + \mathbf{q_p}\, h_p$$

e

$$\mathbf{b_{n+2}} = \mathbf{b_{n+1}} + \mathbf{q_{p-1}}\, h_{p-1}$$

onde, $h_{p-1} = \frac{l'}{\lambda_p + l'^2}\,(\, y_{n+2} - \frac{l'}{\sigma_{p-1}}\,\mathbf{p_{p-1}^T y}\,)$ Assim, a cada linha que incluirmos na direção de algum auto-vetor, não será necessário um novo cálculo do β estimado usando todas as matrizes que foram empregados na primeira estimativa de β, e sim utilizarmos apenas os valores da estimativa anterior do β , o vetor $\mathbf{y}$, e sendo que foi acrescentado uma nova linha na direção do i-ésimo auto-vetor com norma l, os vetores $\mathbf{q_i}$, $\mathbf{p_i}$, o auto- valor λ_i e o valor l.

4.7 Exemplos de Aplicação do Método

Para uma melhor visão da solução apresentada na seção 3.5 iremos colocar dois exemplos de matrizes mal condicionadas e aplicaremos o método proposto

1)- Seja $\mathbf{H}$ uma matriz de Hilbert 4×4, então:

$$\mathbf{H} = \begin{bmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 & 1/5 \\ 1/3 & 1/4 & 1/5 & 1/6 \\ 1/4 & 1/5 & 1/6 & 1/7 \end{bmatrix}$$

e utilizaremos esta como sendo a matriz de variáveis preditoras $\mathbf{X}$.

Seja então o modelo de regressão (2.2.1):

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

suponhamos que os erros ϵ_i sejam independente e identicamente distribuídos com distribuição normal com média zero e variância 1. Então $\mathbf{y} \sim N(\mathbf{X}\beta, \mathbf{I})$ e para simulação de $\mathbf{y}$ suponhamos que o verdadeiro valor de β é dado por:

$$\beta = [\, 1 \; 1 \; 1 \; 1 \,]^T;$$

Para a obtenção do vetor $\mathbf{y}$ geramos cinco valores amostrais de uma distribuição normal padrão com a utilização do método de Box-Miller, e depois acrescentamos os valores da média , $\mathbf{X}\beta$.

Assim,

$$\mathbf{y} \sim N(\, \mathbf{X}\beta \, , \, \mathbf{I} \,)$$

e os valores de $\mathbf{X}\beta$ foram:

$$\mathbf{X}\beta = \begin{bmatrix} 2.083 \\ 1.283 \\ 0.95 \\ 0.73 \end{bmatrix}$$

e

$$\mathbf{y} = \begin{bmatrix} 8.615 \\ 1.444 \\ 0.188 \\ 1.527 \end{bmatrix}$$

Com a utilização do IML do SAS calculamos os auto-valores associados à matriz $\mathbf{X^TX}$, o qual foi:

$$\lambda = \begin{bmatrix} 2.2452943 \\ 0.0263739 \\ 0.0002713 \\ 0.0000106 \end{bmatrix}$$

Notamos que o λ_4 é pequeno em relação ao λ_1 , o que indica que a matriz deve ser mal condicionada. Calculamos então o número de condição que foi :

$$K(\mathbf{X}) = 460$$

e o valor do β estimado :

$$\mathrm{b} = \begin{bmatrix} 34.179 \\ -102.527 \\ 121.753 \\ -59.541 \end{bmatrix}$$

Notamos que o número de condição é grande indicando o mal condicionamento da matriz. A fim de termos uma melhor visão geramos uma matriz de perturbação $\delta\mathbf{X}$, onde os elementos desta matriz de perturbação tem distribuição uniforme no intervalo de 0 a 10^{-3}, e foi dado por:

$$\delta\mathbf{X} = \begin{bmatrix} 2.32830643E-10 & 3.34111507E-04 & 6.05867477E-04 & 9.21498751E-04 \\ 3.23983840E-05 & 1.87925994E-04 & 3.85236460E-04 & 6.32863026E-04 \\ 8.83923377E-04 & 8.87059601E-04 & 6.82666432E-05 & 3.12527175E-04 \\ 7.05339014E-04 & 8.10669967E-04 & 9.45096835E-04 & 1.56293623E-04 \end{bmatrix}$$

Calculamos a matriz de variáveis preditora perturbada $\mathbf{XP} = \mathbf{X} + \delta\mathbf{X}$, e o beta estimado com a matriz de variáveis preditora $\mathbf{XP}$. Os valores desta($\mathbf{bp}$) foram:

$$\mathrm{bp} = \begin{bmatrix} 36.08 \\ -112.204 \\ 126.433 \\ -53.989 \end{bmatrix}$$

Assim, enquanto a matriz de perturbação $\delta\mathbf{X}$ possuía elementos que variavam da ordem de 0 a 10^{-3}, a maior diferença entre os betas estimados, **b** e **bp**, foi de 9.68. Acrescentamos então uma nova linha na direção do auto-vetor do quarto auto-valor com norma $l = 1.50$, ficando assim:

$$\mathbf{x_{n+1}} = L\ \mathbf{q_4} = 1.50 * \begin{bmatrix} 0.146 \\ -0.768 \\ 0.607 \\ 0.143 \end{bmatrix} = \begin{bmatrix} 0.22 \\ -1.151 \\ 0.909 \\ 0.214 \end{bmatrix}$$

Escolhemos a norma l de tal maneira que o menor auto-valor (λ_4) ficasse igual ao maior auto-valor(λ_1). É claro que a norma do vetor acrescentado l não precisava ser de tal maneira que fosse o mesmo do maior auto-valor, mas como sabemos que quanto maior o auto-valor, menor será a variância do **b** naquela direção , foi utilizado este valor. Neste caso o número de condição foi:

$$K(\ \mathbf{X}\ ) = 91$$

o que indica uma "grande" diminuição no valor do mesmo. O valor do beta estimado com esta nova matriz foi:

$$\mathbf{b} = \begin{bmatrix} 12.212 \\ 12.572 \\ 30.832 \\ -80.928 \end{bmatrix}$$

e após somarmos uma matriz de perturbação à matriz de variáveis preditora de forma análoga ao que foi feito anteriormente o valor do beta estimado foi:

$$\mathbf{bp} = \begin{bmatrix} 12.460 \\ 12.123 \\ 29.872 \\ -79.519 \end{bmatrix}$$

e a maior diferença entre o **b** e **bp** neste caso foi de 1.41, ou seja, esta diferença é aproximadamente sete vezes menor do que a diferença anterior. Como o número de condição foi de k(**X**) = 91, acrescentamos uma nova linha na direção do auto-vetor do penúltimo auto-valor(λ_3) com norma $l = 1.50$, ficando assim $\lambda_1 = \lambda_3$. A nova linha $\mathbf{x_{n+2}^T}$ foi dada por:

$$\mathbf{x_{n+2}} = L\,\mathbf{q_3} = 1.50 * \begin{bmatrix} 0.061 \\ -0.219 \\ -0.490 \\ 0.842 \end{bmatrix} = \begin{bmatrix} 0.092 \\ -0.328 \\ -0.734 \\ 1.261 \end{bmatrix}$$

e neste caso o valor do número de condição foi:

$$K(\,\mathbf{X}\,) = 9$$

o que indica que a matriz não deve ser mal condicionado. Foi calculado o valor do beta estimado(**b**) e do beta estimado com a matriz de variáveis preditora perturbada (**bp**), que foi:

$$\mathbf{b} = \begin{bmatrix} 17.431 \\ -6.109 \\ -10.99 \\ -9.026 \end{bmatrix}$$

e

$$\mathbf{bp} = \begin{bmatrix} 17.460 \\ -6.119 \\ -11.001 \\ -9.034 \end{bmatrix}$$

Assim, a maior diferença entre o **b** e **bp** foi de 0.028, enquanto que antes da inclusão das linhas foi de 9.68.

2)-Como um segundo exemplo , foram utilizados dados reais retirado do artigo "An appraisal of least squares programs for the electronic computer from the point of view of the user" de James W.Longley e Silver Spring.

A matriz de variáveis preditora **X** é dado por:

$$\mathbf{X} = \{\, \mathbf{x_1}\ \mathbf{x_2}\ \mathbf{x_3}\ \mathbf{x_4}\ \mathbf{x_5}\ \mathbf{x_6}\ \};$$

onde

$\mathbf{x_1}$ = deflator de preço implícito, GNP.
$\mathbf{x_2}$ = Produto nacional bruto.

$\mathbf{x_3}$ = Desemprego.
$\mathbf{x_4}$ = Tamanho da força armada.
$\mathbf{x_5}$ = População não institucional com idades iguais ou acima de 14.
$\mathbf{x_6}$ = Tempo.

e a variável dependente :
$\mathbf{y}$ = Emprego total retirado.

e foi usado o valor de b:

$$\mathbf{b} = \begin{bmatrix} -3482258 \\ 15.062 \\ -0.036 \\ -2.02 \\ -1.033 \\ -0.051 \\ 1829 \end{bmatrix}$$

retirado da tabela 10 do artigo, como sendo o verdadeiro valor do β para o cálculo de $\mathbf{y} \sim$ N($\mathbf{X}\beta$, $\mathbf{I}$) com a utilização do método de Box-Miller. Com a utilização do IML do SAS calculamos os valores dos auto-valores e auto-vetores de $\mathbf{X^T X}$:

$$\lambda = \begin{bmatrix} 2.76E12 \\ 7.04E9 \\ 11608546 \\ 2504095 \\ 1734.17 \\ 13.38 \\ 1.19E-7 \end{bmatrix}$$

e

$$\mathbf{Q} =$$

$$\begin{bmatrix} 2.3418E-6 & 0.00001 & -9.71E-6 & -2.80E-6 & 0.00051 & -0.000035 & 0.99999 \\ 0.000243 & 0.000616 & -0.00049 & 0.00144 & 0.104358 & 0.994538 & -0.000019 \\ 0.960337 & -0.27879 & -0.00223 & -0.00464 & 0.00135 & -0.00019 & 3.054E-8 \\ 0.007777 & 0.010278 & 0.785382 & 0.618637 & 0.01704 & -0.0023 & 4.5594E-7 \\ 0.006267 & 0.013421 & -0.61873 & 0.78542 & -0.00807 & -0.0006 & 1.3124E-7 \\ 0.278625 & 0.959975 & -5.37E-6 & -0.0188 & -0.0213 & 0.001607 & -1.03E-7 \\ 0.004579 & 0.020892 & -0.01843 & -0.0047 & 0.99413 & -0.10433 & -0.000511 \end{bmatrix}$$

Notamos que pelo menos três dos auto-valores , λ_5, λ_6, λ_7, são "extremamente pequenos" em relação ao λ_1, sendo que o λ_3 e λ_4 estão "bastante distantes" de λ_1 também.

Calculamos o valor de **b** , obtendo-se:

$$\mathbf{b} = \begin{bmatrix} -3485363 \\ 14.383 \\ -0.036 \\ -2.02 \\ -1.034 \\ -0.05 \\ 1830 \end{bmatrix}$$

e geramos a matriz de perturbação $\delta\mathbf{X}$ de maneira análoga ao exemplo 1 e acrescentams à matriz **X** ficando assim, $\mathbf{XP} = \mathbf{X} + \delta\mathbf{X}$ e calculamos o valor de beta estimado com a matriz perturbada **XP** , o qual foi :

$$\mathbf{bp} = \begin{bmatrix} -212414 \\ -42.313 \\ 0.061 \\ -0.527 \\ -0.567 \\ -o.357 \\ 155 \end{bmatrix}$$

Notamos que enquanto a matriz de perturbação é composta de elementos com valores que variam da ordem de 10^{-3} a 10^{-5} a maior diferença entre **b** e **bp** foi da ordem de 3×10^{6} e no valor do $\mathbf{b_2}$ e $\mathbf{b_3}$ houve uma mudança

no sinal. O número de condição foi: k($\mathbf{X}$) = 4,8 $\times$ 10^8 indicando assim que existe problema de multicolinearidade. Acrescentamos então uma nova linha na direção do auto-vetor correspondente a λ_7 e escolhemos como sendo o valor de l o valor l=1,67 $\times$ 10^5, ficando assim o valor de λ_1 igual ao valor do λ_7 . Assim a nova linha foi dada por:

$$\mathbf{x_{n+1}} = L\ \mathbf{q_7} = \begin{bmatrix} 1663609 \\ -31.092 \\ 0.05 \\ 0.758 \\ 0.218 \\ -0.171 \\ -850 \end{bmatrix}$$

e o valor do beta estimado com esta nova matriz foi:

$$\mathbf{b} = \begin{bmatrix} -34821876 \\ 600 \\ -0.992 \\ -16.309 \\ -5.147 \\ 3.177 \\ 17858 \end{bmatrix}$$

e o valor do beta estimado com a matriz de variáveis preditora perturbada, foi:

$$\mathbf{bp} = \begin{bmatrix} -34821876 \\ 1334.699 \\ -1.426 \\ -15.855 \\ 0.992 \\ 8.24 \\ 17600 \end{bmatrix}$$

Neste caso a maior diferença entre o beta estimado com a matriz de variáveis preditora não perturbada e o beta estimado com a matriz de variáveis preditora perturbada foi de 734.44 e houve uma mudança no valor do $\mathbf{b_5}$. O número de condição foi de :

$$k(\mathbf{X}) = 4,5 \times\ 10^3$$

indicando assim um forte indício de multicolinearidade ainda existente na matriz **X**.

Foi então acrescentada mais uma linha na direção do auto-vetor correspondente à λ_6 com norma l =1,67 $\times$ 10^5, fazendo com que este auto-valor fique com o mesmo valor do λ_1. Esta nova linha foi dada por:

$$\mathbf{x_{n+2}} = L\ \mathbf{q_6} = \begin{bmatrix} -57.848 \\ 1654523 \\ -33.418 \\ -3827 \\ -10123 \\ 2674 \\ -173567 \end{bmatrix}$$

Foi calculado o valor do beta estimado com esta nova matriz que foi:

$$\mathbf{b} = \begin{bmatrix} -34821876 \\ 600 \\ -0.993 \\ -16.311 \\ -5.148 \\ 3.178 \\ 17858 \end{bmatrix}$$

e o valor do beta estimado com a matriz de variáveis preditora perturbada foi:

$$\mathbf{bp} = \begin{bmatrix} -34821878 \\ 582.063 \\ -1.276 \\ -14.115 \\ 1.453 \\ 7.023 \\ 17679 \end{bmatrix}$$

Neste caso a maior diferença entre o beta estimado com a matriz de variáveis preditora não perturbada e o beta estimado com a matriz

de variáveis preditora perturbada foi de 178.5 e continuou havendo uma mudança de sinal no $\mathbf{b_5}$. O número de condição neste caso foi de:

$$k(\mathbf{X}) = 39949$$

indicando ainda a existência de multicolinearidade.

Acrescentamos então a terceira linha na direção do auto-vetor correspondente ao auto-valor λ_5 com norma $l = 1{,}67 \times 10^5$, fazendo com que este também fique com o mesmo valor do λ_1. Assim, a nova linha foi dada por:

$$\mathbf{x_{n+3}} = L\,\mathbf{q_5} = \begin{bmatrix} 849 \\ 173611 \\ 2246 \\ 28355 \\ -13436 \\ -35568 \\ 1653843 \end{bmatrix}$$

Neste caso o valor do beta estimado foi:

$$\mathbf{b} = \begin{bmatrix} -34821876 \\ 600 \\ -0.992 \\ -16.31 \\ -5.147 \\ 3.176 \\ 17858 \end{bmatrix}$$

e o valor do beta estimado com a matriz de variáveis preditora perturbada foi:

$$\mathbf{bp} = \begin{bmatrix} -34821876 \\ 600 \\ -1.033 \\ -11.053 \\ 0.003 \\ 3.184 \\ 17858 \end{bmatrix}$$

Neste caso a maior diferença entre **b** e **bp** foi de 5.26 e continua havendo uma mudança de sinal no $\mathbf{b_5}$. O número de condição foi de:

$$k(\mathbf{X}) = 1051$$

indicando ainda a existência de multicolinearidade. Acrescentamos então de um modo análogo a quarta linha $\mathbf{x}_{n+4}^{T}$:

$$\mathbf{x_{n+4}} = L\,\mathbf{q_4} = \begin{bmatrix} -4.66 \\ 2396 \\ -7726 \\ 1029170 \\ 1306632 \\ -31359 \\ -7945 \end{bmatrix}$$

e neste caso os valores do beta estimado com a matriz de variáveis preditora não perturbada e o do beta estimado com a matriz de variáveis preditora perturbada foram:

$$\mathbf{b} = \begin{bmatrix} -34821876 \\ 600 \\ -0.993 \\ -16.31 \\ -5,146 \\ 3.176 \\ 17858 \end{bmatrix}$$

e

$$\mathbf{bp} = \begin{bmatrix} -34821876 \\ 600 \\ -0.99 \\ -15.562 \\ -5.722 \\ 3.321 \\ 17858 \end{bmatrix}$$

assim a maior diferença diferença entre eles foi de 0.75 e não houve mais, mudança no sinal.O número de condição foi de:

$$k(\mathbf{X}) = 488$$

O que mostra que ainda há problema de multicolinearidade . Acrescentamos então a quinta linha de maneira análoga, obtendo-se:

$$\mathbf{x_{n+5}} = L\ \mathbf{q_5} = \begin{bmatrix} -16.163 \\ -825 \\ -3713 \\ 1306566 \\ -1029326 \\ -8.945 \\ -30672 \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} -34821876 \\ 600 \\ -0.99 \\ -16.309 \\ -5.146 \\ 3.176 \\ 17858 \end{bmatrix}$$

e

$$\mathbf{bp} = \begin{bmatrix} -34821876 \\ 600 \\ -0.99 \\ -16.297 \\ -5.143 \\ 3.321 \\ 17858 \end{bmatrix}$$

Neste caso a maior diferença entre **b** e **bp** foi de 0.14. O número de condição neste caso foi:

$$k(\mathrm{x}) = 19.83$$

Se incluirmos mais uma linha de forma análoga , teremos:

$$\mathbf{x_{n+6}} = L\,\mathbf{q_6} = \begin{bmatrix} 17.94 \\ 1024 \\ -463209 \\ 17077 \\ 22299 \\ 1594992 \\ 34712 \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} -34821876 \\ 600 \\ -0.99 \\ -16.31 \\ -5.146 \\ 3.176 \\ 17858 \end{bmatrix}$$

e

$$\mathbf{bp} = \begin{bmatrix} -34821876 \\ 600 \\ -0.96 \\ -16.298 \\ -5.145 \\ 3.203 \\ 17858 \end{bmatrix}$$

e a maior diferença entre **b** e **bp** é de 0.0297 e o número de condição é 1.
Na prática pode não ser possível o acréscimo de linhas na direção e com a norma desejada, como por exemplo este segundo exemplo que é constituído de dados econômicos, e que foi utilizada por ser um conjunto de dados bastante citado na literatura. No entanto existem muitas situações onde isto é possível , por exemplo em experimentos de laboratório ou em máquinas onde a pessoa pode controlar os dados de entrada. Mas ainda assim pode ser que não seja possível acrescentar a linha com a norma desejada por ser um valor "muito grande" por exemplo, tendo que diminuir o valor desta. Neste caso é aconselhável colocar esta linha com o maior valor da norma dentro das possibilidades.

A Definições

Aqui apresentaremos algumas definições sobre elementos básicos da teoria de matrizes que serão necessários ao desenvolvimento desta dissertação.

DEFINIÇÃO A.1 (Rudin, 1976, pg. 3) *Um corpo é um conjunto F com duas operações denominadas adição e multiplicação, que satisfazem os seguintes axiomas:*

(A)- Axiomas de adição

(A1)- se $x, y \in F$, então a soma também pertence à F, isto é, $x+y \in F$.
(A2)- adição é comutativa: $x + y = y + x$; $\forall x, y \in F$.
(A3)- adição é associativa: $(x + y) + z = x + (y + z)$; $\forall x, y, z \in F$.
(A4)- F contém um elemento 0 tal que $0 + x = x$, qualquer que seja $x \in F$.
(A5)- para todo $x \in F$ existe um elemento $-x \in F$, tal que

$$x + (-x) = 0$$

e

(M)- Axiomas de multiplicação

(M1)- se $x, y \in F$, então o produto também pertence à F, isto é, $xy \in F$.
(M2)- multiplicação é comutativa: $xy = yx$; $\forall x, y \in F$
(M3)- multiplicação é associativa: $(xy)z = x(yz)$; $\forall x, y \in F$.
(M4)- F contém um elemento $1 \neq 0$ tal que $1x = x$, qualquer que seja $x \in F$.
(M5)- se $x \in F$ e $x \neq 0$, então existe um elemento $1/x$, tal que

$$x \cdot (1/x) = 1$$

e

(D)- A lei distributiva

$$x(y+z) \;=\; xy + xz$$

qualquer que seja $x, y, z \in F$.

Nesta dissertação, sempre que fizermos referência a um corpo estaremos considerando os números reais ($\Re$) ou os números complexos (C), com suas operações usuais de adição e multiplicação.

O objeto fundamental de estudo desta dissertação é a matriz, que pode ser definido de duas formas: tábua retangular de escalares ou como transformações lineares entre dois espaços vetorias, especificado suas respectivas bases, onde escalares são elementos de um determinado corpo F. O conjunto de todas as matrizes $m \times n$ sobre F será denotado por $M_{m,n}(F)$ e $M_{n,n}(F)$ será abreviado por $M_n(F)$ e denotaremos as matrizes por letras maiusculas em negrito, por exemplo, $\mathbf{A}, \mathbf{B}$, *etc* os vetores serão denotados por letras minúsculas em negrito, por exemplo, $\mathbf{a}, \mathbf{y}, \mathbf{x}$, *etc*.

A seguir definiremos norma de vetores e matrizes, apresentando alguns tipos de normas que serão utilizados.

DEFINIÇÃO A.2 (Horn and Jonhson, 1990, pg. 259) *Seja V um espaço vetorial. Uma função $\| \cdot \|: V \to \Re$ é uma norma de vetor se para todo $\mathbf{x}, \mathbf{y} \in F$:*

1- $\| \mathbf{x} \| \geq 0 \;\; e \;\; \| \mathbf{x} \| = 0 \iff \mathbf{x} = \mathbf{0}$.
2- $\| c\mathbf{x} \| = | c | \| \mathbf{x} \|$, *para todo escalar c (real ou complexo).*
3- $\| \mathbf{x} + \mathbf{y} \| \leq \| \mathbf{x} \| + \| \mathbf{y} \|$

Existem vários tipos de norma de vetores, entre estas, citamos a norma euclidiana (ou, norma l_2)sobre $\Re^n$, que é definida por:

$$\| \mathbf{x} \|_2 = \sqrt{(\mathbf{x}^{\mathbf{T}}\mathbf{x})} \;=\; \left(x_1^2 + \cdots + x_n^2\right)^{\frac{1}{2}}$$

DEFINIÇÃO A.3 (Horn and Jonhson, 1990, pg. 290) *Norma de matrizes*

Seja $\mathbf{M_n}$ *um espaço vetorial de dimensão* n^2, *denotaremos a função* $|||\cdot|||: M_n \to \Re$, *como norma de matriz se para todo* $\mathbf{A}, \mathbf{B} \in \mathbf{M_n}$, *as*

seguinte condições forem satisfeitas:

1- $|||\ A\ |||\geq 0$ *e* $|||\ \mathbf{A}\ |||= 0$ *se, e só se,* $\mathbf{A} = \mathbf{0}$.
2- $|||\ c\mathbf{A}\ |||=|\ c\ |\ |||\ \mathbf{cA}\ |||$ *para todo escalar c (real ou complexo).*
3- $|||\ \mathbf{A}+\mathbf{B}\ |||\leq|||\ \mathbf{A}\ |||\ +\ |||\ \mathbf{B}\ |||$
4- $|||\ \mathbf{AB}\ |||\leq|||\ \mathbf{A}\ |||\ \mathbf{B}\ |||$

A seguir apresentaremos algumas normas sobre o $\mathbf{M_n}$. A norma Euclidiana ou norma l_2 definida para $\mathbf{A} \in \mathbf{M_n}$ (Horn and Johnson, 1990, pg. 291) é dada por:

$$\| \mathbf{A} \|_2 = \left(\sum_{i,j=1}^{n} a_{ij}^2 \right)^{\frac{1}{2}}$$

Um outro tipo de norma de matriz a ser utilizado aqui é definido da seguinte maneira (Horn and Johnson pg 292): seja $\|\ .\ \|$ uma norma de vetor em $\Re^n$. Define-se $|||\ .\ |||$ em $\mathbf{M_n}$ por:

$$|||\ \mathbf{A}\ ||| = max_{\| \mathbf{x} \| = 1} \| \mathbf{Ax} \|$$

B Matrizes Ortogonais

Neste apêndice, vamos difinir matrizes ortogonais e apresentar alguns resultados que serão utilizados nesta dissertação.

DEFINIÇÃO B.1 *Uma matriz* $\mathbf{U} \in \mathbf{M_n}$ *é dita ser ortogonal se* $\mathbf{U^T U = I}$.

Aqui temos como objetivo estabelecer algumas condições que são equivalentes à definição acima, a fim de que nã haja dúvidas sobre as diversas definições de matrizes ortogonais encontradas na literatura, isto é, queremos mostrar a equivalência entre as diversas definições encontradas na literatura.

TEOREMA B.1 *Se* $\mathbf{A} \in \mathbf{M_n}$ *e* $\mathbf{BA = I}$. *Então,* $\mathbf{A}$ *é não singular,* $\mathbf{B}$ *é única e*

$$\mathbf{AB} = I$$

Prova: Se $\mathbf{A} \in \mathbf{M_n}$ e $\mathbf{BA} = \mathbf{I}$, então se $\mathbf{Ax} = \mathbf{0}$, temos que

$$\mathbf{x} = \mathbf{Ix} = \mathbf{BAx}$$

Portanto, $\mathbf{Ax} = \mathbf{0} \iff \mathbf{x} = \mathbf{0}$, o que quer dizer que $\mathbf{A}$ é não singular. Agora, a não singularidade de $\mathbf{A}$ implica que as equações:

$$\mathbf{Ax} = \mathbf{y} \qquad \mathbf{x}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}} = \mathbf{y}^{\mathbf{T}}$$

têm ambas uma única solução qualquer que seja $\mathbf{y} \in \Re$. Desta forma, sabemos que existe $\mathbf{B_L} \in \mathbf{M_n}$ e $\mathbf{B_R} \in \mathbf{M_n}$, que são as únicas soluções de:

$$\mathbf{B_L A} = \mathbf{I} \qquad \mathbf{AB_R} = \mathbf{I}$$

respectivamente. Agora, da expressão

$$\mathbf{B_L} = \mathbf{B_L A B_R} = \mathbf{B_R}$$

concluímos que $\mathbf{B}$ é unica e

$$\mathbf{AB} = I$$

□

Este resultado apresenta algumas interpretações interessantes, por exemplo, se as condições do teorema são válidas, então B é a inversa de A. Apresentaremos agora um resultado que nos dá algumas condições equivalentes à matriz ser ortogonal.

TEOREMA B.2 *Se* $\mathbf{U} \in \mathbf{M_n}$, *as seguintes relações são equivalentes:*

a- $\mathbf{U}$ *é ortogonal*
b- $\mathbf{U}$ *é não singular e* $\mathbf{U^T} = \mathbf{U^{-1}}$
c- $\mathbf{UU^T} = \mathbf{I}$
d- $\mathbf{U^T}$ *é ortogonal*
e- as colunas de $\mathbf{U}$ *formam um conjunto ortogonal*
f- as linhas de $\mathbf{U}$ *formam um conjunto ortogonal*
g- Se para $\mathbf{x} \in \Re^n$, $\mathbf{y} = \mathbf{Ux}$ *segue que* $\mathbf{y^T y} = \mathbf{x^T x}$

Prova: Inicialmente, vamos mostrar que as condições (a) à (d) são equivalentes, pois isto é conseqüência direta do teorema 1. Assim, se $\mathbf{U}$ é ortogonal

$$\mathbf{U^T U = I}$$

Portanto, segue do teorema 1 que $\mathbf{U}$ é não singular e

$$\mathbf{U U^T = I}$$

Desde que $\mathbf{U^{-1}}$ é única, concluímos que $\mathbf{U^T = U^{-1}}$. Agora, usando o fato de que $(\mathbf{U^T})^T = \mathbf{U}$, podemos mostrar que (a) à (d) são equivalentes.

Denotando por $\mathbf{u^{(i)}}$ a i-ésima coluna de $\mathbf{U}$, segue que se $\mathbf{U}$ é ortogonal

$$\mathbf{u^{(i)\ T}\ u^{(j)}} \begin{cases} 0 \quad ; \quad se \ \ i \neq j \\ \\ 1 \quad ; \quad se \ \ i = j \end{cases}$$

Portanto, $\mathbf{U^T U} = I$ é uma forma diferente de dizer que as colunas de $\mathbf{U}$ são ortogonais. Desta forma similar, segue que (a) à (f) são equivalentes.

Se (a) é válido e $\mathbf{y = Ux}$, então

$$\mathbf{y^T y \ = \ x^T U^T U x \ = \ x^T I x \ = \ x^T x}$$

isto é, (a) implica (g). A volta, isto é, (g) implica (a) têm uma demonstração um pouco mais complicada e inicialmente faremos para o caso bidimensional. Portanto, para n=2, assumindo (g) e difinindo

$$\mathbf{x} \ = \ \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix}$$

obtemos que

$$1 \ = \ \mathbf{x^T x} \ = \ \mathbf{y^T y} \ = \ \mathbf{x^T U^T U x} \qquad (1)$$

Desta forma, se

$$\mathbf{U^T U} = \begin{bmatrix} a_{11} & a_{12} \\ \\ a_{21} & a_{22} \end{bmatrix}$$

da equação (1) podemos concluir que

$$a_{11} = 1$$

Da mesma forma, se tomarmos

$$\mathbf{x} = \begin{bmatrix} 0 \\ \\ 1 \end{bmatrix}$$

obtemos

$$a_{22} = 1$$

Denotando os vetores colunas de $\mathbf{U}$ por $\mathbf{u_1}$ e $\mathbf{u_2}$, repectivamente. Segue que

$$\mathbf{U^T U} = \begin{bmatrix} 1 & a \\ \\ a & 1 \end{bmatrix}$$

onde a é o produto interno de $\mathbf{u_1}$ com $\mathbf{u_2}$, isto é,

$$a = u_1^T u_2$$

Assim, tomando

$$\mathbf{x} = \begin{bmatrix} 1 \\ \\ 1 \end{bmatrix}$$

obtemos que

$$2 = \mathbf{x^T x} = \mathbf{y^T y} = \mathbf{x^T U^T U x} = 2 + 2a$$

portanto, $a = 0$. Isto significa que se $\mathbf{x^T U^T U x} = \mathbf{x^T x}$ qualquer que seja $x \in \Re^n$, então

$$\mathbf{U^T U} = \mathbf{I}$$

isto é, $\mathbf{U}$ é ortogonal. Agora, para $n > 2$, considere $\mathbf{A} = \mathbf{U^T U}$. Assim, seja $\mathbf{x} \in \Re^n$ tal que todos os componentes são iguais à zero com excessão da i-ésima e j-ésima coordenada, com $i \leq j$. Então,

$$\mathbf{x^T A x} = [x_i \;\; x_j]\mathbf{A}(\{\mathbf{i,j}\}) \begin{bmatrix} x_i \\ \\ x_j \end{bmatrix}$$

onde $\mathbf{A}(\{\mathbf{i,j}\})$ representa a submatriz de $\mathbf{A}$ que é formada tomando-se a interseção da i-ésima e j-ésima linha com a iésima e j-ésima coluna. Assim, como já mostramos

$$\mathbf{A}(\{\mathbf{i,j}\}) = \mathbf{I} \in \mathbf{M_2}$$

Desde que i e j são arbitrários, concluímos que qualquer submatriz principal de $\mathbf{A}_{2\times 2}$ é igual à identidade. A única matriz $\mathbf{A}$ satisfazendo isto é a identidade. Portanto

$$\mathbf{U^T U} = \mathbf{I}$$

para $\mathbf{U} \in \mathbf{M_n}$. □

Maiores informações sobre matrizes ortogonais podem ser encontrados, entre outros, em textos tais como Horn Johnson (1990).

COROLÁRIO B.1 *Seja* $\mathbf{U} \in \mathbf{M_n}$ *uma matriz ortogonal,* $\mathbf{x}$ *um vetor* $n \times 1$ *e* $\mathbf{A} \in \mathbf{M_n}$. *Então:*
a)- $\| \mathbf{Ux} \|_2 = \| \mathbf{x} \|_2$
b)- $\| \mathbf{UA} \|_2 = \| \mathbf{A} \|_2$

Prova: a)- Seja $\mathbf{Ux} = \mathbf{y}$, ou seja, $\mathbf{y}$ é um vetor $n \times 1$, então:
$\| \mathbf{y} \|_2 = \mathbf{y^T y} = \mathbf{x^T U^T U x} = \mathbf{x^T x} = \| \mathbf{x} \|_2$
e portanto a norma Euclidiana é unitariamente invariante.

b)-Seja $\mathbf{A} = [\, \mathbf{a}_1, \ldots, \mathbf{a}_n \,] \in \mathbf{M_n}$, onde $\mathbf{a}_i$ é um vetor $n \times 1$. Então:

$$\| \mathbf{A} \|_2^2 = \| \mathbf{a_1} \|_2^2 + ... + \| \mathbf{a_n} \|_2^2$$

e

$$\| \mathbf{UA} \|_2^2 = \| \mathbf{Ua_1} \|_2^2 + ... + \| \mathbf{Ua_n} \|_2^2 =$$

$$= \| \mathbf{a_1} \|_2^2 + ... + \| \mathbf{a_n} \|_2^2 = \| \mathbf{A} \|_2^2$$

como $\| \mathbf{B^T} \|_2 = \| \mathbf{B} \|_2$ para todo $\mathbf{B} \in \mathbf{M_n}$, temos que:

$$\| \mathbf{AU} \|_2 = \| \mathbf{U^T A^T} \|_2 = \| \mathbf{A^T} \|_2 = \| \mathbf{A} \|_2$$

então a norma l_2 em $\mathbf{M_n}$ é uma norma de matriz unitariamente invariante.

C Prova de alguns teoremas

Prova do teorema 2.9

Como $\mathbf{A}$ é simétrica, existe uma matriz ortogonal $\mathbf{U}$ tal que $\mathbf{A} = \mathbf{UDU^T}$ e $\mathbf{D} = \text{diag}\ (\ \lambda_1, \ ...\ ,\ \lambda_n\ )$. Se $\mathbf{x} \neq \mathbf{0}$, temos que:

$$\frac{\mathbf{x^T A x}}{\mathbf{x^T x}} = \frac{(\ \mathbf{U^T x}\ )^T \mathbf{D} (\ \mathbf{U^T x}\ )}{\mathbf{x^T x}} = \frac{(\ \mathbf{U^T x}\ )^T \mathbf{D} (\ \mathbf{U^T x}\ )}{(\ \mathbf{U^T x}\ )^T (\ \mathbf{U^T x}\ )}$$

e como $\{\ \mathbf{U^T x} : \mathbf{x} \in \Re^n \ \text{e}\ \mathbf{x} \neq \mathbf{0}\ \} = \{\ \mathbf{y} \in \Re^n : \mathbf{y} \neq \mathbf{0}\ \}$ para $\mathbf{y} = \mathbf{U^T x}$, então se $\mathbf{w_1}$, $\mathbf{w_2}$, ... , $\mathbf{w_{n-k}} \in \Re^n$ é dado, temos que:

$$sup_{\mathbf{x} \neq \mathbf{0}\ ,\ \mathbf{x} \perp \mathbf{w_1}, ..., \mathbf{w_{n-k}}} \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} = sup_{\mathbf{y} \neq \mathbf{0}\ ,\ \mathbf{y} \perp \mathbf{U^T w_1}, ..., \mathbf{U^T w_{n-k}}} \frac{\mathbf{y^T D y}}{\mathbf{y^T y}} =$$

$$= sup_{\mathbf{y^T y} = 1\ ,\ \mathbf{y} \perp \mathbf{U^T w_1}, ..., \mathbf{U^T w_{n-k}}} \sum_{i=1}^{n} \lambda_i \mid y_i \mid^2$$

$$\geq \; sup_{\mathbf{y^T y=1}\ ,\ \mathbf{y \perp U^T w_1,...,U^T w_{n-k}}\ ,\ y_1=...=y_{k-1}=0} \; \sum_{i=1}^{n} \lambda_i \mid y_i \mid^2$$

$$= sup_{\mathbf{y \perp U^T w_1,...,U^T w_{n-k}}\ ,\ y_k^2 + ... + y_n^2 = 1} \; \sum_{i=k}^{n} \lambda_i \mid y_i \mid^2 \geq \lambda_k \quad (1.9.3)$$

e

$$inf_{\mathbf{x \neq 0}\ ,\ \mathbf{x \perp w_1,...,w_{k-1}}} \; \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} \; = \; inf_{\mathbf{y \neq 0}\ ,\ \mathbf{y \perp U^T w_1,...,U^T w_{k-1}}} \; \frac{\mathbf{y^T D y}}{\mathbf{y^T y}} \; =$$

$$inf_{\mathbf{y^T y=1}\ ,\ \mathbf{y \perp U^T w_1,...,U^T w_{k-1}}} \; \sum_{i=1}^{n} \lambda_i \mid y_i \mid^2$$

$$\geq \; inf_{\mathbf{y^T y=1}\ ,\ \mathbf{y \perp U^T w_1,...,U^T w_{k-1}}\ ,\ y_1=...=y_{k-1}=0} \; \sum_{i=1}^{n} \lambda_i \mid y_i \mid^2$$

$$= \; inf_{\mathbf{y \perp U^T w_1,...,U^T w_{k-1}}\ ,\ y_k^2+...+y_n^2=1} \; \sum_{i=1}^{n} \lambda_i \mid y_i \mid^2 \geq \lambda_k \quad (1.9.4)$$

Agora, no teorema 2.8 mostramos que:

$$\lambda_1 \; = \; min_{\mathbf{x \neq 0}} \; \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} \; = \; min_{\mathbf{x^T x = 1}} \; \mathbf{x^T A x}$$

e

$$\lambda_n \; = \; max_{\mathbf{x \neq 0}} \; \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} \; = \; max_{\mathbf{x^T x = 1}} \; \mathbf{x^T A x}$$

Suponha que $\mathbf{A = U D U^T}$ com $\mathbf{U} = [\ \mathbf{u_1}, \ ...\ , \mathbf{u_n}\ ]$, onde as colunas de U são os auto-vetores ortonormais de **A**. Se considerarmos apenas os vetores $\mathbf{x} \in \Re^n$ que são ortogonais a $\mathbf{u_1}$, teremos:

$$\mathbf{x^T A x} \; = \; \sum_{i=1}^{n} \lambda_i \mid (\ \mathbf{u_i^T x}\ ) \mid^2 = \; \sum_{i=2}^{n} \lambda_i \mid \mathbf{u_i^T x} \mid^2$$

$$\geq \lambda_2 \sum_{i=2}^{n} \mid \mathbf{u_i^T x} \mid^2 = \lambda_2 \sum_{i=1}^{n} \mid (\mathbf{u_i^T x}) \mid^2 = \lambda_2 \mathbf{x^T x}$$

e esta desigualdade se torna igualdade se escolhermos $\mathbf{x = u_2}$, então:

$$min_{\mathbf{x} \neq \mathbf{0}\ ,\ \mathbf{x} \perp \mathbf{u_1}} \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} = min_{\mathbf{x^T x} = 1\ ,\ \mathbf{x} \perp \mathbf{u_1}} \mathbf{x^T A x} = \lambda_2$$

assim:

$$min_{\mathbf{x} \neq \mathbf{0}\ ,\ \mathbf{x} \perp \mathbf{u_1}, \ldots, \mathbf{u_{k-1}}} \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} =$$

$$= min_{\mathbf{x^T x} = 1\ ,\ \mathbf{x} \perp \mathbf{u_1}, \ldots, \mathbf{u_{k-1}}} \mathbf{x^T A x} = \lambda_k\ , \quad k = 2, \ldots, n \quad (1.9.5)$$

e

$$max_{\mathbf{x} \neq \mathbf{0}\ ,\ \mathbf{x} \perp \mathbf{u_1}, \ldots, \mathbf{u_{n-k+1}}} \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} =$$

$$= max_{\mathbf{x^T x} = 1\ ,\ \mathbf{x} \perp \mathbf{u_n}, \ldots, \mathbf{u_{n-k+1}}} \mathbf{x^T A x} = \lambda_{n-k} \quad k = 1, \ldots, n-1 \quad (1.9.6)$$

Tínhamos em (1.9.3) que:

$$sup_{\mathbf{x} \neq \mathbf{0}\ ,\ \mathbf{x} \perp \mathbf{w_1}, \ldots, \mathbf{w_{n-k}}} \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} \geq \lambda_k$$

para quaisquer n-k vetores $\mathbf{w_1}, \ldots, \mathbf{w_{n-k}}$, e por (1.9.6) esta desigualdade será igualdade para uma escolha de vetores $\mathbf{w_i}$; $\mathbf{w_i} = \mathbf{u_{n-i+1}}$. Então:

$$inf_{\mathbf{w_1}, \ldots, \mathbf{w_{n-k}}} \ sup_{\mathbf{x} \neq \mathbf{0}\ ,\ \mathbf{x} \perp \mathbf{w_1}, \ldots, \mathbf{w_{n-k}}} \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} = \lambda_k \quad (1.9.7)$$

Em (1.9.4), tínhamos que:

$$inf_{\mathbf{x} \neq \mathbf{0}\ ,\ \mathbf{x} \perp \mathbf{w_1}, \ldots, \mathbf{w_{k-1}}} \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} \geq \lambda_k$$

para quaisquer k-1 vetores $\mathbf{w_1},...,\mathbf{w_{k-1}}$, e por (1.9.5) esta desigualdade será igualdade para uma escolha de vetores $\mathbf{w_i}$; $\mathbf{w_i} = \mathbf{u_i}$. Então:

$$sup_{\mathbf{w_1},...,\mathbf{w_{k-1}}} \; inf_{\mathbf{x} \neq \mathbf{0} \, , \, \mathbf{x} \perp \mathbf{w_1},...,\mathbf{w_{k-1}}} \; \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} = \lambda_k \quad (1.9.8)$$

e podemos substituir o ínfimo e o supremo com mínimo e máximo respectivamente, pois o extremo é atingido. Assim,

$$min_{\mathbf{w_1},...,\mathbf{w_{n-k}}} \; max_{\mathbf{x} \neq \mathbf{0} \, , \, \mathbf{x} \perp \mathbf{w_1},...,\mathbf{w_{n-k}}} \; \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} = \lambda_k$$

e

$$max_{\mathbf{w_1},...,\mathbf{w_{k-1}}} \; min_{\mathbf{x} \neq \mathbf{0} \, , \, \mathbf{x} \perp \mathbf{w_1},...,\mathbf{w_{k-1}}} \; \frac{\mathbf{x^T A x}}{\mathbf{x^T x}} = \lambda_k$$

Prova do teorema 2.13.

a)- Com a utilização do teorema anterior , temos que, $\mathbf{A}^{"}$, $\mathbf{B}^{"}$ e $\mathbf{E}^{"}$ são matrizes (m+n)×(m+n) simétricas e seja $\lambda_i(\mathbf{E}^{"})$ o i-ésimo auto-valor ordenado de $\mathbf{E}^{"}$. Como $\mathbf{E}^{"} = \mathbf{B}^{"} - \mathbf{A}^{"}$, temos $\mathbf{B}^{"} = \mathbf{A}^{"} + \mathbf{E}^{"}$ e pelo teorema 2.10 temos:

$$\lambda_k(\mathbf{A}^{"}) + \lambda_1(\mathbf{E}^{"}) \leq \lambda_k(\mathbf{A}^{"} + \mathbf{E}^{"}) \leq \lambda_k(\mathbf{A}^{"}) + \lambda_n(\mathbf{E}^{"})$$

agora, como pelo teorema 2.12:

$$\lambda_1(\mathbf{E}^{"}) = -\sigma_n(\mathbf{E})$$

e

$$\lambda_n(\mathbf{E}^{"}) = \sigma_n(\mathbf{E})$$

temos:

$$\lambda_k(\mathbf{A}^{"}) - \sigma_n(\mathbf{E}) \leq \lambda_k(\mathbf{B}^{"}) \leq \lambda_k(\mathbf{A}^{"}) + \sigma_n(\mathbf{E})$$

$$-\sigma_n(\mathbf{E}) \leq \lambda_k(\mathbf{B}^{"}) - \lambda_k(\mathbf{A}^{"}) \leq \sigma_n(\mathbf{E})$$

$\Rightarrow$

$$| \lambda_k(\mathbf{B}^{"}) - \lambda_k(\mathbf{A}^{"}) | \leq \sigma_n(\mathbf{E})$$

e assim,

$$| \lambda_k(\mathbf{B}^{"}) - \lambda_k(\mathbf{A}^{"}) | \leq | \lambda_n(\mathbf{E}^{"}) | \quad (1.13.1)$$

Agora seja $\rho(\mathbf{E}^{"})$ o raio espectral da matriz $\mathbf{E}^{"}_{(\mathbf{m+n})\times(\mathbf{m+n})}$, ou seja:

$$\rho(\mathbf{E}^{"}) = max\{ | \lambda | : \lambda \, e \, um \, auto - valor \, de \, \mathbf{E}^{"} \}$$

Seja então , λ qualquer auto-valor de $\mathbf{E}^{"}$, então $| \lambda | \leq \rho(\mathbf{E}^{"})$ e existe pelo menos um auto-valor de λ para o qual $| \lambda | = \rho(\mathbf{E}^{"})$. Se $\mathbf{E}^{"}\mathbf{x} = \lambda\mathbf{x}$, $\mathbf{x} \neq \mathbf{0}$ e se $| \lambda | = \rho(\mathbf{E}^{"})$, considere a matriz $\mathbf{X}_{(\mathbf{n}\times\mathbf{n})}$ onde todas as suas n colunas são o vetor $\mathbf{x}$, então $\mathbf{E}^{"}\mathbf{X} = \lambda\mathbf{X}$ e:

$$| \lambda | \| \mathbf{X} \|_2 = \| \lambda\mathbf{X} \|_2 = \| \mathbf{E}^{"}\mathbf{X} \|_2 \leq \| \mathbf{E}^{"} \|_2 \| \mathbf{X} \|_2$$

e portanto $| \lambda | = \rho(\mathbf{E}^{"}) \leq \| \mathbf{E}^{"} \|_2$.

Assim, a desigualdade (1.13.1) é dado por:

$$| \lambda_k(\mathbf{B}^{"}) - \lambda_k(\mathbf{A}^{"}) | \leq | \lambda_n(\mathbf{E}^{"}) | \leq \| \mathbf{E}^{"} \|_2$$

e pelo teorema 2.12 , temos que:

$$| \tau_i - \sigma_i | \leq \| \mathbf{E} \|_2 \quad para \, todo \, i = 1, ..., q. \quad \square$$

b)-Com a utilização do teorema 2.12 , temos que $\mathbf{A}^{"}$, $\mathbf{B}^{"}$ e $\mathbf{E}^{"}$ são matrizes (m+n)×(m+n) simétricas . Seja λ_1 ,..., λ_{n+m} os auto- valores de $\mathbf{A}^{"}$ em alguma ordem dada, e seja λ'_1 ,..., λ'_{n+m} os auto-valores de $\mathbf{B}^{"}$ em alguma ordem. Seja as matrize $\mathbf{D} = \text{diag}(\lambda_1 , ..., \lambda_{m+n})$, $\mathbf{D}' = \text{diag}(\lambda'_1 , ..., \lambda'_{n+m})$, $\mathbf{V}_{(\mathbf{m+n})\times(\mathbf{m+n})}$ uma matriz ortogonal tal que $\mathbf{A}^{"} = \mathbf{VDV^T}$ e $\mathbf{W}_{(\mathbf{m+n})\times(\mathbf{m+n})}$ uma matriz ortogonal tal que $\mathbf{B}^{"} = \mathbf{WD'W^T}$. Então:

$$\| \mathbf{E}^{"} \|_2^2 = \| (\mathbf{A}^{"} + \mathbf{E}^{"}) - \mathbf{A}^{"} \|_2^2$$

$$= \| \ \mathbf{WD'W^T} - \mathbf{VDV^T} \ \|_2^2$$

$$=^{(I)} \| \ \mathbf{V^T WD'W^T V} - \mathbf{D} \ \|_2^2$$

$$= \| \ \mathbf{ZD'Z^T} - \mathbf{D} \ \|_2^2$$

$$= \ tr(\ \mathbf{ZD'Z^T} - \mathbf{D}\ )(\ \mathbf{ZD'Z^T} - \mathbf{D}\ )^T$$

$$= \ tr(\ \mathbf{D'D'^T} + \mathbf{DD^T}\ ) - tr(\ \mathbf{ZD'Z^T D^T} + \mathbf{DZD'^T Z^T}\ )$$

$$= \sum_{i=1}^{n} (\ |\ \lambda_i'\ |^2 + |\ \lambda_i\ |^2\ ) - 2tr(\ \mathbf{ZD'Z^T D^T}\ )$$

onde $\mathbf{Z} = \mathbf{V^T W}$.

((I) as matrizes ortogonais são invariantes em $\| \quad \|_2$, Apêndice B, Corolário $B.1$)

Então,

$$\| \ \mathbf{E}'' \ \|_2^2 \geq \sum_{i=1}^{n} (\ |\ \lambda_i'\ |^2 + |\ \lambda_i\ |^2\ ) -$$

$$2\ max\{\ tr(\ \mathbf{UD'U^T D^T}\ )\ : \mathbf{U}\ e\ ortogonal\}.$$

e tr($\mathbf{UD'U^T D^T}$) = $\sum_{i=1}^{n} |\ u_{i,j}\ |^2\ (\ \lambda_i\ \lambda_j'\ )$, e queremos encontrar o valor máximo desta expressão. Se nós denotarmos por $c_{i,j} = |\ u_{i,j}\ |^2$, sendo a matriz C = [$c_{i,j}$] , uma matriz (m+n)×(m+n) com elementos $c_{i,j}$, i,j=1,...,m+n , então C é uma matriz duplamente estocástica e então é um conjunto convexo compacto em $\mathbf{C_{(m+n)\times(m+n)}}$ com um domínio maior(Horn and Johnson, Cap 8, seção 7), e portanto:

$$max\{\ tr(\ \mathbf{UD'U^T D^T}\ )\ :\ \mathbf{U}\ e\ ortogonal\}$$

$$= max\{ \sum_{i,j=1}^{n} \mid u_{i,j} \mid^2 (\lambda_i \lambda'_j) : \mathbf{U} \text{ e } ortogonal\}$$

$$\leq max\{ \sum_{i,j=1}^{n} c_{i,j} (\lambda_i \lambda'_j) : \mathbf{C} \text{ e } duplamente\ estocastico\}$$

e como a função a ser maximizada é uma função linear em um conjunto convexo compacto, o máximo ocorre em pontos extremos do conjunto convexo, e estes pontos em matrizes duplamente estocásticas são as matrizes de permutação (Horn and Johnson, Cap 8, seção 7) e portanto existe uma matriz permutação $\mathbf{P_{(m+n)x(m+n)}}$ tal que:

$$max\{ \sum_{i,j=1}^{n} c_{i,j} (\lambda_i \lambda'_j) : \mathbf{C} \text{ e } duplamente\ estocastico \}$$

$$= tr(\mathbf{PD'P^TD^T})$$

e como uma matriz permutação é ortogonal:

$$max\{ tr(\mathbf{UD'U^TD^T}) : \mathbf{U} \text{ e } ortogonal \} = tr(PD'P^TD^T)$$

Se $\mathbf{P}\ \mathbf{e_i} = \mathbf{e}_{\sigma_i}$ para i=1,...,n , então:

$$tr(\mathbf{PD'P^TD^T}) = \sum_{i=1}^{n} (\lambda'_{\sigma_{(i)}} \lambda_i)$$

e

$$\| \mathbf{E}'' \|^2 \geq \sum_{i=1}^{n} [\mid \lambda'_{\sigma_{(i)}} \mid^2 + \mid \lambda_i^2 - 2 (\lambda'_{\sigma_{(i)}} \lambda_i)]$$

$$= \sum_{i=1}^{n} \mid \lambda'_{\sigma_{(i)}} - \lambda_i \mid^2$$

e portanto $[\sum_{i=1}^{n} \mid \lambda'_{\sigma_{(i)}} - \lambda_i \mid^2]^{\frac{1}{2}} \leq \| \mathbf{E}'' \|$ e pelo teorema 2.12,

$$[\sum_{i=1}^{n} \mid \tau_i - \sigma_i)^2]^{\frac{1}{2}} \leq \| \mathbf{E} \| \quad \square.$$

Referências

[1] A. Albert (1972) - *Regression and The Moore-Penrose Pseudoinverse*, Academic Press, N. Y.

[2] D. A. Belsley, E. Kuh e R. E. Welsch (1980) - *Regression Diagnostics: Identifying data and Sources of Collinearity*, John Wiley and Sons

[3] O. H. Bustos e A. C. Frery (1992) - it Simulação Estocástica: teoria e Algorítimos, Monografia de matemática, no. 49, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Instituto de Matemática Pura e Aplicada.

[4] J. Dagpunar (1988) - *Principles of Random Variate Generation*, Clavendon Press, Oxford

[5] N. R. Drapper e H. Smith (19) - *Applied Regression Analysis*, 2nd Edition, John Wiley and Sons, New York

[6] G. Forsythe e C. B. Moler (1967) - *Computer Solution of Linear Algebraic Systems*, Printice-Hall, INC

[7] R. A. Horn and C. R. Johnson (1990) - *Matrix Analysis*, Cambridge University Press

[8] P. Lancaster e M. Tismenetsky (1985) - *The theory of Matrices*, Academic Press, N. Y.

[9] E. D. Nering (1970) - *Linear Algebra and Matrix Theory*, 2nd Edition, John Wiley and Sons, INC, N. Y.

[10] J. M. Ortega (1987) - *Matrix Theory: a second course*, Plenum Press, New York and London

[11] C. R. Rao e S. K. Mitra (1971) - *Generalized Inverse of Matrices and its Applications*, John Wiley and Sons, N. Y.

[12] W. Rudin (1976) - *Principles of Mathematical Analysis*, third edtion, McGraw-Hill

[13] S. A. Searle (1971) - *Linear Models*, John Wiley and Sons, INC, New York

[14] G. Seber (1977) - *Linear Regression Analysis*, John Wiley and Sons, INC

[15] G. W. Stewart (1973) - *Introduction to Matrix Computations*, Academic Press

[16] H. D. Vinod e A. Ullah (1981) - *Recent Advances in REgression Methods*, Marcel Dekker, INC