SILVIO JACKS DOS ANJOS GARNÉS

AJUSTAMENTO PARAMÉTRICO POR MÍNIMOS QUADRADOS COM ANÁLISE NA ESTABILIDADE DA SOLUÇÃO

Dissertação apresentada ao Curso de Pós-Graduação em Ciências Geodésicas da Universidade Federal do Paraná, como requisito para a obtenção do grau de Mestre em Ciências.

ORIENTADOR: Dr. Raimundo J. B. de Sampaio

CO-ORIENTADOR: Dr. Quintino Dalmolin

CURITIBA
1996

SILVIO JACKS DOS ANJOS GARNÉS

AJUSTAMENTO PARAMÉTRICO POR MÍNIMOS QUADRADOS COM ANÁLISE NA ESTABILIDADE DA SOLUÇÃO

Dissertação apresentada ao curso de Pós-Graduação em Ciências Geodésicas da Universidade Federal do Paraná como requisito para a obtenção do grau de Mestre em Ciências.

ORIENTADOR: Dr. Raimundo J. B. de Sampaio

CO-ORIENTADOR: Dr. Quintino Dalmolin

CURITIBA

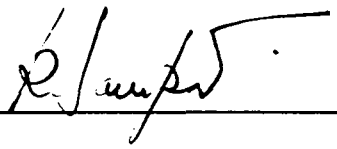
1996

AJUSTAMENTO PARAMÉTRICO POR MÍNIMOS QUADRADOS COM ANÁLISE NA ESTABILIDADE DA SOLUÇÃO.

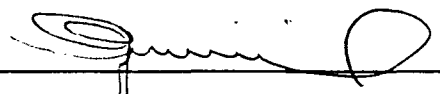
POR

SILVIO JACKS DOS ANJOS GARNÉS
ENGENHEIRO AGRIMENSOR

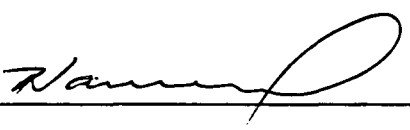
Dissertação aprovada como requisito parcial para a obtenção do grau de Mestre em Ciências no Curso de Pós-Graduação em Ciências Geodésicas da Universidade Federal do Paraná, pela comissão formada pelos seguintes professores:



Prof. Dr. RAIMUNDO J. B. DE SAMPAIO - ORIENTADOR



Prof. Dr. QUINTINO DALMOLIN - MEMBRO



Prof. MsC. ROMUALDO WANDRESEN - MEMBRO

Curitiba, 03 de Abril de 1996

DEDICATÓRIA

Dedico este trabalho a:

Etelvina dos Anjos Soares

Minha mãe, pelo amor e pelo apoio, que sempre me deu forças nos momentos mais dificeis de minha vida.

Alcides Lorca Garnés

Flávio dos Anjos Garnés

Gislany dos Anjos Garnés

Meu pai e meus irmãos, pelo companherismo e carinho que mesmo estando longe, estamos perto.

Maria Aparecida Vasconcelos dos Santos

Rafael Vasconcelos Garnés

Matheus Vasconcelos Garnés

Minha esposa e meus filhos pela compreensão, esperança, carinho e amor.

AGRADECIMENTOS

Desejo externar aqui meus sinceros agradecimentos:

ao Prof. Dr. Raimundo J. B. de Sampaio, pela ajuda na definição da meta, pela valiosa orientação e pelas correções durante a realização deste;

ao Co-orientador Prof. Dr. Quintino Dalmolin, pelo apoio, correções e idéias sugeridas;

ao Prof. Dr. Milton de Azevedo Campos, pelo apoio com relação ao tema no momento da escolha do mesmo;

ao amigo Alfonso R. T. Criollo por propiciar discuções a cerca do assunto;

ao Dr. Luiz Danilo D. Ferreira, pelas dicas indispensaveis a respeito do texto.

aos professores do departamento de matemática que ministraram o curso de especialização de matemática aplicada durante o ano de 1993, os quais muito contribuiram para o amadurecimento dos conhecimentos que foram utilizados aqui;

e a todos aqueles que de uma forma ou de outra contribuiram para a realização deste trabalho.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE SÍMBOLOS

$A_{(m,n)}$	Matriz A com m-linhas e n-colunas.
C(A)	Número de condição de A.
C^1	Diferenciavel continuamente.
C^2	Duas vezes diferenciavel continuamente.
$\frac{dA}{dt}$	Derivada de A com respeito a t.
H(x)	Hessiana.
inf	Ínfimo.
L(X,Y)	Espaço vetorial das transformações lineares de X em Y.
max	Máximo.
min(m,n)	Mínimo entre m e n.
Posto(A)	Posto da matriz A.
R	Conjunto dos números reais.
R^m	Espaço vetorial m-dimensional.
R^n	Espaço Vetorial n-dimensional.
$A \in R^{m \times n}$	Matriz A com elementos reais de m-linhas e n-colunas.
sup	Supremo.
Traço(A)	Soma dos elementos da diagonal de A.
$\|\cdot\|$	Norma de vetor ou matriz.
$\frac{\partial Y}{\partial x}$	Derivada parcial de Y com respeito a x.
$\nabla f(x)$	Gradiente da função f(x).
$\Re(A)$	Espaço coluna da matriz A.
$\aleph(A^T)$	Espaço nulo de A^T ou espaço nulo a esquerda de A.
$\aleph(A)$	Espaço nulo de A.
$\Re(A^T)$	Espaço linha de A.
$< x,y >$	Produto interno dos vetores x e y.

$\Rightarrow$	Iplica que.
$\Leftrightarrow$	Equivale a.
$\rightarrow$	Leva em.
$\forall$	Para todo.
$\neq$	Diferente.
$\approx$	Aproximadamente.
∞	Infinito.
$\cong$	Igual ou aproximadamente.
$\exists$	Exite.
$\therefore$	Donde.
$\{\ \}$	Conjunto vazio.
$\cup$	União.
$\cap$	Intersecção.
$\in$	Pertence.
$\notin$	Não pertence.

RESUMO

Este trabalho tem como objetivo estudar a solução do ajustamento paramétrico pelo método de mínimos quadrados e os problemas que podem ocorrer na busca dessa solução quanto a estabilidade.

Foram feitas as comparações de cinco métodos de solução para sistema de equações lineares redundantes, mostrando as vantangens e desvantagens de cada um através de testes.

Quando as equações residuais são não-lineares, são utilizados dois métodos para obter a solução, um com características de convergência local e outro com característica de convergência global, e no final é feita a aplicação em um exemplo prático da Geodésia para mostrar essas características.

Na conclusão são comentados os resultados e feita as recomendações sobre os métodos a serem utilizados em determinados tipo de problemas.

ABSTRACT

The purpose of this work is to apllie least squares methods for adjustment parameter problem that appear from Geodesy.

It was made comparison of five methods of solution for systems of redundant linear equations, point out the advantages and disadvantages from each one of them.

When the residual equation is nonlinear, they were applied two methods to get the solution, one with local characteristc of convergence and another with global convergence. In the end, it is made aplication to a practical exemple of Geodesy to show those characteristcs.

In the conclusion we comment the results and made same recommendations about the methods to be applied in different kind of problems.

1. INTRODUÇÃO

1.1 GENERALIDADES

O problema ou método de mínimos quadrados nasceu independentemente com dois grandes nomes da matemática; Gauss (1809) e Legendre (1806). A partir daí, a metodologia ganhou crescente importância e aplicações de maneira que hoje é a principal técnica de ajustamento de observações utilizado em Ciências como: Geodésia, Topografia, Astronomia e outras.

Também, tem concepções diferentes dependendo da Ciência que o estuda. Por exemplo: em Matemática, o problema interessa do ponto de vista da existência da solução e dos métodos que podem fornecer esta solução. Em Estatística, o problema decorre de observações e tem na sua concepção a distribuição de probabilidades.

As notacões utilizadas em torno do problema diferem tanto do contexto em que é tratado, como também dos autores que tratam dum mesmo contexto. Neste trabalho será utilizada a notação usual dos textos de programação matemática.

O problema de mínimos quadrados pode ser enunciado como:

$$\text{Minimizar } f(x) = \|V(x)\|_2^2 \qquad (1.1)$$

onde:

$V: R^n \rightarrow R^m$: é a função residual;

$\|\cdot\|_2^2$: é norma euclidiana ao quadrado.

Autores como DENNIS e SCHNABEL (1983) tratam o problema (1.1) de duas formas; primeiro: quando V é linear, chamam (1.1) de problema de mínimos quadrados linear; segundo: quando V é não-linear, chamam (1.1) de problema de mínimos quadrados não-linear. Neste trabalho será feita esta distinção para maior clareza no entendimento dos métodos de solução.

No caso do problema de mínimos quadrados linear, autores como LAWSON e HANSON (1974), o definem:

Dada uma matriz real $A_{(m \times n)}$ de posto(k)$\leq$min(m,n)[1] e um vetor real **b** m-dimensional, encontrar um vetor real $\mathbf{x}^*$ que minimiza o quadrado da norma euclidiana de **Ax-b**.

Não fazem distinção quanto ao tamanho relativo de m e n. Neste caso o problema pode ser formado por qualquer um dos casos ilustrado na fig. 01 (LAWSON e HANSON 1974).

Neste trabalho serão enfatizados os casos 2a e 2b, ou seja, solucionar um sistema de equações lineares inconsistente (impossível ou incompatível), com m>n, sem a consideração de que o sistema seja formado por observações.

Foi escolhido, trabalhar somente com a parte da solução do problema de mínimos quadrados, apenas fazendo mensão quando necessário a parte das observações, porque a grande maioria das literaturas na área de ajustamento, tratam a solução de um prisma muito elementar, quando podem ocorrer problemas graves nos resultados encontrados usando tal prisma (problema do mal-condicionamento como será visto adiante), além de que, por si só a parte da solução já é um assunto bastante extenso.

---

[1] posto, rank ou característica de uma matriz, são termos utilizados para dizer quantas linhas ou colunas são linearmente independentes nesta.

FIGURA 01 - CASOS PARA O PROBLEMA DE MÍNIMOS QUADRADOS LINEAR.

caso 1a

Ax = b

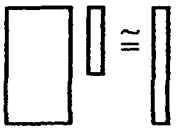
posto(A)=m=n

caso 1b

Ax ≅ b

posto(A)=k<m=n

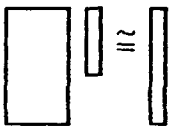
caso 2a



Ax ≅ b

posto(A)=n<m

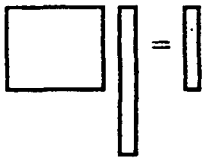
caso 2b



Ax ≅ b

posto(A)=k<n<m

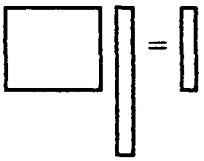
caso 3a



Ax = b

posto(A)=m<n

caso 3b



Ax ≅ b

posto(A)=k<m<n

Outra questão a ser levantada é com relação ao título do trabalho. Sabe-se que para a aplicação do critério de mínimos quadrados existe três técnicas muito comuns utilizadas em ajustamento de observações, as quais

são: a) Método das observações indiretas ou método paramétrico; b) Método das observações diretas condicionadas (LUGNANI 1983) ou método dos correlatos ; c) Método combinado.

O nome ajustamento paramétrico no título refere-se ao caso da técnica (a), isso porque a maioria dos problemas formulados para serem resolvidos pela técnica (b) podem ser formulados para serem resolvidos pela técnica (a). Mesmo utilizando as técnicas (b) e (c) para problemas lineares ou mesmo não-lineares, conforme tratou DALMOLIN (1976), o problema é resolvido através da solução de uma seqüência de sistema lineares o que acaba sendo o caso 1a mostrado na fig.01. Os métodos de solução do problema de mínimos quadrados linear para os casos 2a e 2b, são também aplicados a esse caso (para o caso 1b , dos métodos expostos aqui, somente o método utilizando a decomposição de valor singular pode ser usado).

1.2 ESTRUTURA DO TRABALHO

No capítulo 2 são colocados os conceitos básicos a serem utilizados no decorrer do trabalho.

O capítulo 3 fornece as solução para o problema de mínimos quadrados linear por diversos caminhos, inclusive quando as equações são ponderadas.

O capítulo 4 trata da análise do condicionamento do problema. Lá é mostrada a desvantagem em solucionar o problema através das equações normais e o que ocorre no problema de mínimos quadrados linear quando são perturbados a matriz de coeficientes das incógnitas e o vetor dos termos independentes.

O capítulo 5 traz os métodos numéricos para a solução do problema de mínimos quadrados linear, onde cada método vem com uma subrotina. Os

métodos considerados para obter a solução do problema foram: Gauss-Jordan com pivôtamento completo; subroutina Versol; decomposição de Cholesky; decomposição QR; e decomposição de Valor Singular.

O capítulo 6 mostra alguns testes realizados com os métodos acima quanto a estabilidade e a velocidade na solução, a fim de tirar algumas conclusões para serem aplicadas ao caso do problema de mínimos quadrados não-linear tratado no capítulo 7.

O capítulo 7 traz dois métodos de solução do problema de mínimos quadrados não-linear, são eles: o método de Gauss-Newton e o método de Levenberg-Marquardt. O método de Gauss-Newton porque é o mais básico, tem características de convergência local (os tipos convergência serão definidos adiante) e o método de Levenberg-Marquardt porque tem características de convergência global. Ainda no mesmo capítulo, é utilizado um exemplo prático e os resultados da aplicação de um programa computacional para mostrar tais características.

Por último é feita uma conclusão a respeito de todo o trabalho e algumas recomendações quanto aos métodos de solução a utilizar para resolver o problema de mínimos quadrados.

1.3 OBJETIVOS DESTA PESQUISA

O objetivo final deste trabalho é identificar os problemas que podem surgir na solução do problema de mínimos quadrados e fornecer métodos numéricos alternativos para a solução do mesmo, quanto a estabilidade, velocidade e convergência (no caso não-linear) para a solução.

2. ELEMENTOS DE ANÁLISE E SISTEMAS DE EQUAÇÕES LINEARES.

Neste capítulo serão apresentadas as ferramentas essenciais usadas no decorrer do trabalho. Serão fornecidos os conceitos de espaço vetorial, espaço normado, sistema de equações lineares, derivada de matrizes, conjunto e funções convexas, condições de otimalidade e convergência de sequência de números reais.

2.1 ESPAÇOS VETORIAIS

Definição: Espaço vetorial, é um conjunto V, não-vazio, sobre o qual estão definidos as operações de adição (+) : $V \times V \to V$ e multiplicação por escalar ($\cdot$) : $R \times V \to V$, tais que com estas operações se cumpre os seguintes axiomas:

a) Em relação à adição
   - i) $(u+v)+w = u+(v+w)$, $\forall \; u,v,w \in V$;
   - ii) $u+v = v+u$, $\forall \; u,v \in V$;
   - iii) $\exists \; 0 \in V$, $\forall \; u \in V$, $u+0=u$;
   - iv) $\forall \; u \in V$, $\exists \; (-u) \in V$, $u+(-u)=0$.

b) Em relação a multiplicação por escalar

   Para qualquer $u,v \in V$ e $\forall \; \alpha,\beta \in R$
   - vi) $(\alpha\beta)u = \alpha(\beta u)$;
   - vii) $(\alpha+\beta)u = \alpha u+\beta u$;
   - viii) $\alpha(u+v) = \alpha u+\alpha v$;
   - ix) $1u = u$.

Os elementos pertencentes ao espaço vetorial V são chamados de vetores, não importando sua natureza.

Quando, para o conjunto dos escalares da definição acima for tomado o conjunto C dos números complexos, o espaço vetorial V é chamado de espaço vetorial complexo.

2.2 NORMAS

Seja V um espaço vetorial real. Diz-se que $\varphi: V \rightarrow R$ é uma norma em V, se:

i) $\varphi(x) \geq 0$, $\forall\, x \in V$ e

$\varphi(x) = 0$ se, e somente se $x=0$;

ii) $\varphi(\lambda x) = |\lambda|.\varphi(x)$, $\forall\, \lambda \in R$, $x \in V$;

iii) $\varphi(x+y) \leq \varphi(x)+\varphi(y)$, $\forall\, x,y \in V$.

Define-se a distância entre dois pontos quaisquer do espaço como:

$d(x,y) = \varphi(x-y)$.

É claro que o valor vai depender de qual for a norma escolhida já que existem infinitas normas.

Aqui a norma de um vetor será denotada por $\|\cdot\|$. Algumas das normas mais comuns em R^n são:

i) $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ (norma-1) (2.1)

ii) $\|x\|_2 = (\sum_{i=1}^{n} |x_i|^2)^{1/2}$ (norma-2 ou euclidiana) (2.2)

iii) $\|x\|_\infty = \max_{1 \leq i \leq n} \{ |x_i| \}$ (norma máxima ou infinita) (2.3)

iv) $\|x\|_P = (\sum_{i=1}^{n} |xi|^P)^{1/P}$, $1 \leq P \leq \infty$ (norma P) (2.4)

A norma de uma matriz também será denotada por $\|\cdot\|$.

Seja $W = L(X,Y)$, e $A \in W$. A norma de A é definida como :

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|_y}{\|x\|_x} . \qquad (2.5)$$

Uma simples verificação mostra que a norma de A satisfaz às seguintes condições:

i) $||A|| \geq 0$, $\forall$ $A \in W$

ii) $||\lambda A|| = |\lambda| . ||A||$, $\forall$ $\lambda \in R$, $\forall$ $A \in W$

iii) $||A+B|| \leq ||A|| + ||B||$, $\forall$ A , $B \in W$.

2.2.1 Normas induzidas

Definição: Para qualquer matriz A e uma norma de vetor $||\cdot||$, a norma de matriz induzida ou subordinada $||A||$ é definida por:

$$\|A\| = \inf \{M \in R : \|Ax\| \leq M\|x\|\} . \qquad (2.6)$$

algumas vezes é mais conveniente expressar a definição acima como:

$$\|A\| = \max_{\|u\|=1} \|Au\| , \text{ onde } u = \frac{x}{\|x\|} \text{ ou} \qquad (2.7)$$

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} . \qquad (2.8)$$

As normas de matrizes subordinadas ou induzidas mais comuns são:

$$\|A\|_1 = \max_{1 \leq j \leq n} \{ \|a_j\|_1 \} \text{ (máx.absoluto da soma de col.)} \qquad (2.9)$$

$$\|A\|_2 = \sigma_{max}(A) \quad \text{(maior valor singular de A)} \qquad (2.10)$$

$$\|A\|_\infty = \max_{1 \leq i \leq n} \{ \|a_i\|_1 \} \text{ (máx. absoluto da soma de lin.)} \qquad (2.11)$$

A norma de um vetor e a norma de uma matriz subordinada são sempre compatíveis.

A compatibilidade entre a norma de uma matriz $||\cdot||'$ e a norma de uma vetor $||\cdot||$, é possivel se for satisfeita a seguinte condição:

$$||Ax|| \leq ||A||' . ||x||. \qquad (2.12)$$

Cabe destacar mais duas normas importantes que não são induzidas pela norma de vetores. São elas :

$$\|A\|_A = \max\{ |a_{ij}| \} \qquad \text{(norma absoluta)} \qquad (2.13)$$

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} \qquad \text{(norma Frobenius)} \qquad (2.14)$$

esta última satisfaz a relação

$$\|A\|_F^2 = \text{traço}(A^T A). \qquad (2.15)$$

A norma de Frobenius é compativel com a norma de vetor euclidiana, isto é, para qualquer A e **x**

$$\|Ax\|_2 \leq \|A\|_F \|x\|_2 . \qquad (2.16)$$

2.3 ESPAÇO VETORIAL EUCLIDIANO

Chama-se *espaço vetorial euclidiano*, a um espaço vetorial real normado de dimensão finita, no qual está definido a norma-2.

2.3.1 Produto interno

Produto interno em um espaço vetorial real V é uma função de $V \times V \rightarrow \mathbf{R}$ que a todo par de vetores $(\mathbf{u},\mathbf{v}) \in V \times V$ associa um número real, indicado por $\mathbf{u}.\mathbf{v}$ ou $\langle \mathbf{u},\mathbf{v} \rangle$, tal que os axiomas a seguir se verifiquem :

i) $\mathbf{u}.\mathbf{v} = \mathbf{v}.\mathbf{u}$;

ii) $\mathbf{u}.(\mathbf{v}+\mathbf{w}) = \mathbf{u}.\mathbf{v}+\mathbf{u}.\mathbf{w}$;

iii) $(\alpha u)v = \alpha(\mathbf{u}.\mathbf{v})$, $\forall\ \alpha \in \mathbf{R}$;

iv) $\mathbf{u}.\mathbf{u} \geq 0$ e $\mathbf{u}.\mathbf{u} = 0$ se, e só se $\mathbf{u}=0$.

2.3.2 Subespaço vetorial

Seja **V** um espaço vetorial , e $W \subset V$. **W** é subespaço vetorial de **V** se:

i) Para quaisquer dois vetores $x , y \in W$, sua soma $x+y \in W$;

ii) $\forall \lambda \in R, \ x \in W, \ \lambda.x \in W$.

2.4 SISTEMAS DE EQUAÇÕES LINEARES

Uma equação linear é uma equação da forma:

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b, \qquad (2.17)$$

onde:

$x_1, x_2, \ldots, x_n \in R$ são variáveis;

$a_1, a_2, \ldots, a_n \in R$ são coeficientes das variáveis; e

$b \in R$ é o termo independente.

Os valores das variáveis que tornam verdadeira a equação são chamadas ***raizes*** da equação linear.

Um sistema de equações lineares é um conjunto de equações lineares da forma:

$$\begin{aligned}
a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n &= b_1 \\
a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n &= b_2 \\
\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots & \\
\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots & \\
\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots & \\
a_{m1} x_1 + a_{m2} x_2 + \cdots + a_{mn} x_n &= b_m
\end{aligned} \qquad (2.18)$$

Os valores das variáveis que satisfazem as equações do sistema são chamadas ***raizes*** do sistema de equações lineares (solução do sistema).

Os sistemas podem ser classificados segundo suas soluções, como:

i) Sistema possível (consistente ou compatível) e determinado;

ii) Sistema possível e indeterminado;

iii) Sistema impossível (inconsistente ou incompatível).

Um sistema é possível se admitir raizes ou soluções.

O sistema possível é determinado quando admite solução única.

O sistema possível é indeterminado quando admite infinitas soluções.

O sistema é impossível quando não admite solução.

Dois sistemas são equivalentes quando admitem as mesmas raizes ou soluções.

O sistema de equações lineares (2.18) pode ser escrito na forma matricial como:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \qquad (2.19)$$

e representado por $\mathbf{Ax=b}$.

A classificação de (2.19) é usualmente feita analisando o posto de **A**, o posto de $\bar{\mathbf{A}}=[\mathbf{A},\mathbf{b}]$ (conhecida como matriz ampliada) e o número de variáveis.

A seguir será feito um breve resumo desta classificação.

Seja um sistema de equações lineares $\mathbf{Ax=b}$, com $\mathbf{A}_{(m,n)}$, $\mathbf{x}_{(n,1)}$ e $\mathbf{b}_{(m,1)}$ ou mais explicitamente como em (2.19). A matriz ampliada $\bar{\mathbf{A}}$ será:

$$\bar{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix} . \qquad (2.20)$$

Uma maneira de calcular o posto de uma matriz é reduzí-la à forma escada através de operações elementares. O posto da matriz será o número de linhas com elementos não todos nulos de uma matriz equivalente reduzida à forma escada.

Uma classificação geral não importando a dimensão do sistema poderá ser feito da seguinte forma:

i) se posto(A)=posto($\bar{A}$)=k, o sistema é possível;

ii) se posto(A) < posto($\bar{A}$), o sistema é impossível;

se o sistema for possível, poderá ser:

iii) determinado se k=n ;

iv) indeterminado se k<n.

2.4.1 Os subespaços fundamentais

As afirmações anteriores quanto a classificação dos sistemas de equações lineares, farão mais sentido depois desta secção, pois será dada uma interpretação geométrica para a solução dos sistemas.

Considere os seguintes subespaços obtidos das matrizes dos coeficientes:

1) *O espaço linha de* A; se for aplicado em A o processo de eliminação, reduzindo A à forma escada, o número de linhas com elementos não todos nulos será o *posto(*A)=k e essas linhas formarão uma base para o espaço linha de A.

2) *O espaço nulo de A*; O espaço nulo de A tem dimensão n-k, que é o número de variáveis livres do sistema Ax=0. Uma base para esse subespaço é obtido atribuindo valores a essas variáveis livres.

3) *O espaço coluna de A*; tem dimensão igual ao espaço linha de A,ou seja, posto(A)=k. Uma base para esse espaço pode ser obtida verificando-se

na matriz reduzida à forma escada quais das colunas tem pivôs, as correspondentes colunas em A formarão uma base para esse subespaço.

4) *O espaço nulo de* A^T; Este subespaço é constituido dos vetores y tal que $A^Ty=0$. A dimensão para esse subespaço é m-k. Uma base pode ser formada das últimas m-k linhas de $L^{-1}P$, onde : L é uma matriz de operações elementares e P uma matriz de permutação que quando aplicadas a matriz A, gera a matriz U, reduzida à forma escada (decomposição LU).

A simbologia para os subespaços acima bem como suas dimensões são resumidas abaixo:

i) $\Re(A^T)$ = espaço linha de A ; dimensão k.

ii) $\aleph(A)$ = espaço nulo de A ; dimensão n-k.

iii) $\Re(A)$ = espaço coluna de A; dimensão k.

iv) $\aleph(A^T)$ = espaço nulo de A^T; dimensão m-k.

Uma vez de posse dos conceitos sobre os quatro subespaços relacionados ao sistema Ax=b, pode-se checar a classificação dada anteriormente, utilizando agora os subespaços.

Foi dito que um sistema é possível se o *posto(A)* for igual ao *posto(Ā)*. Analisando esse fato, segue que posto(A) significa exatamente o número de linhas ou colunas linearmente independentes da matriz A. O mesmo acontecendo com a matriz Ā, pois, posto(Ā) equivale ao número de linhas ou colunas linearmente independente de Ā.

Da afirmação acima, conclui-se que uma linha (uma coluna) de Ā deve ser linearmente dependente para que o sistema seja possível. Em particular o vetor b já que o mesmo foi acrescido as colunas de A para formar Ā.

Daí, conclui-se sobre as afirmações (i) e (ii) no critério de classificação que:

a) Se o vetor **b** for combinação linear das colunas de **A**, então o sistema é possível;

b) Se o vetor **b** não for combinação linear das colunas de **A**, o sistema é impossível.

Do ponto de vista geométrico isso quer dizer que se o vetor $\mathbf{b} \in \Re(\mathbf{A})$, então o sistema será possível e será impossivel caso contrário (ver fig. 02).

Caso $\mathbf{b} \in \Re(\mathbf{A})$, então, as condições (iii) e (iv) devem ser verificadas, ou seja, se **b** é formado unicamente ou não.

O sistema possível será determinado se o espaço coluna tiver dimensão total, ou seja, todas as colunas de **A** forem linearmente independente (posto completo). Isto significa que só existe uma maneira de combinar as colunas de **A** (através de um vetor **x**) para formar **b**.

Se alguma coluna de **A** for linearmente dependente, isso implica numa deficiência de posto, e o espaço coluna será gerado por uma base de dimensão menor do que (n) . Neste caso as colunas de **A** formarão apenas um espaço de dimensão $k<n$. Assim sendo, **b** poderia ser escrito de infinitas maneiras.

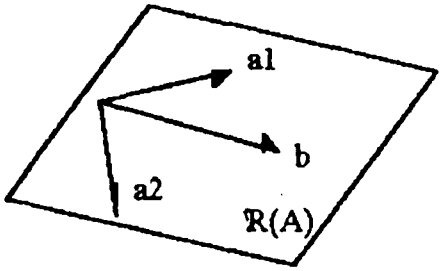
Observe que para estas interpretações foi utilizado somente o espaço coluna de **A**, de modo que é suficiente. Já os demais serão estudados na análise do problema de mínimos quadrados.

2.5 DERIVADA DE MATRIZES

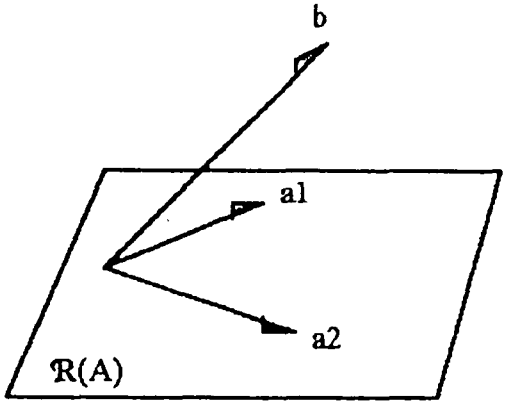
Seja $\mathbf{x} \in \mathbf{R}^n$ com elementos x_i função de $t \in R$ e pelo menos uma vez diferenciavel em t. Então, a derivada de **x** com respeito a t será:

$$\frac{d\mathbf{x}}{dt}=(\frac{dx_1}{dt}\ \frac{dx_2}{dt}\ \cdots\ \frac{dx_n}{dt})^T. \qquad (2.21)$$

FIGURA 02 - REPRESENTAÇÃO GEOMÉTRICA DA CLASSIFICAÇÃO DO SISTEMA LINEAR.



2.a Sistema possível



2.b Sistema impossível

De maneira análoga a derivada de uma matriz $A_{(m,n)}$ com respeito a t será:

$$\frac{dA}{dt}=\begin{bmatrix} \frac{da_{11}}{dt} & \frac{da_{12}}{dt} & \cdots & \frac{da_{1n}}{dt} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \frac{da_{m1}}{dt} & \frac{da_{m2}}{dt} & & \frac{da_{mn}}{dt} \end{bmatrix} \qquad (2.22)$$

exemplo:

$$\mathbf{x}=(e^t\ \mathrm{sen}(t)\quad \cos(t))^T \qquad \therefore \qquad \frac{d\mathbf{x}}{dt}=(e^t\cos(t)+e^t\,\mathrm{sen}(t)\quad -\mathrm{sen}(t))^T\ ;$$

$$\mathbf{A} = \begin{bmatrix} \cos(t) & \text{sen}(t) \\ 1 & e^{2t} \end{bmatrix} \quad \therefore \quad \frac{dA}{dt} = \begin{bmatrix} -\text{set}(t) & \cos(t) \\ 0 & 2e^{2t} \end{bmatrix} .$$

Sejam agora duas matrizes **A** e **B**. Suponha ser possível o produto entre elas. Então, a derivada do produto de **A** por **B** escreve-se:

$$\frac{d(\mathbf{AB})}{dt} = \frac{d(\mathbf{A})}{dt}\mathbf{B} + \mathbf{A}\frac{d(\mathbf{B})}{dt} \tag{2.23}$$

2.5.1 Derivada parcial de uma matriz

Seja $\mathbf{y} \in \mathbf{R}^m$ e $\mathbf{x} \in \mathbf{R}^n$ com y_i função de x_j então as derivadas parciais de y_i com respeito a x_j escreve-se:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \tag{2.24}$$

a (2.24) é conhecida como matriz jacobiano.

Exemplo :

$$\mathbf{y} = \begin{bmatrix} 3x_1 + 4x_2 + 5 \\ 10x_1^2 + 3x_2^3 \\ x_1^3 + 2x_2^2 \end{bmatrix} \qquad \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} 3 & 4 \\ 20x_1 & 9x_2^2 \\ 3x_1^2 & 4x_2 \end{bmatrix}$$

o diferencial de y pode ser escrito como:

$$dy = \frac{\partial y}{\partial x} dx \qquad (2.25)$$

2.5.2 Derivada de expressão linear

Seja $\mathbf{y} = \mathbf{Ax}$ (2.26)

onde:

$\mathbf{y} \in \mathbf{R}^m$;

$\mathbf{A} \in \mathbf{R}^{mxn}$, com elementos constantes;

$\mathbf{x} \in \mathbf{R}^n$, vetor das variáveis independentes.

Então:

$$dy = d(\mathbf{Ax}) = \frac{\partial(\mathbf{Ax})}{\partial \mathbf{x}} d\mathbf{x}\,, \qquad (2.27)$$

logo a derivada de y com respeito a x será:

$$\frac{dy}{d\mathbf{x}} = \mathbf{A}\,. \qquad (2.28)$$

2.5.3 Derivada da forma bilinear

Chama-se forma bilinear a expressão:

$$u = \mathbf{x}^T \mathbf{A} y \qquad (2.29)$$

onde:

$u \in \mathbf{R}$;

$\mathbf{A} \in \mathbf{R}^{mxn}$, com elementos constantes;

$y \in \mathbf{R}^n$;

$\mathbf{x} \in \mathbf{R}^m$, vetor das variáveis independentes.

O diferencial se expressa (LUGNANI 1983) por:

$$du = \mathbf{x}^T \mathbf{A} dy + d\mathbf{x}^T \mathbf{A} y \qquad (2.30)$$

2.5.4 Derivada da forma quadrática

A forma quadrática é defida pela expressão:

$$q=\mathbf{x}^T\mathbf{A}\mathbf{x} \qquad (2.31)$$

onde:

$q \in \mathbf{R}$;

$\mathbf{A} \in \mathbf{R}^{n\times n}$, simétrica;

$\mathbf{x} \in \mathbf{R}^n$.

O cálculo da derivada é (MIKHAIL e GRACIE 1981):

$$\frac{\partial q}{\partial \mathbf{x}}=2\mathbf{x}^T\mathbf{A} \qquad (2.32)$$

2.6 FUNÇÕES CONVEXAS

As propriedades que esta classe de funções possui são de grande importância na teoria de otimização. Uma vez que a função é identificada a-priori, certos resultados ficam assegurados. No caso do problema de mínimos quadrados linear, a função residual a ser minimizada pertence a classe das funções convexas. Este é o motivo desta secção.

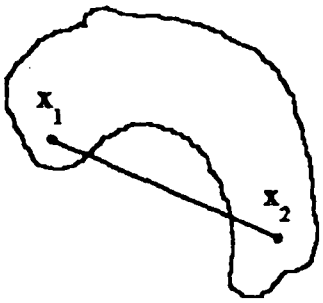
Para introduzir alguns conceitos sobre funções convexas, é útil primeiramente definir conjuntos convexos.

2.6.1 Conjuntos convexos

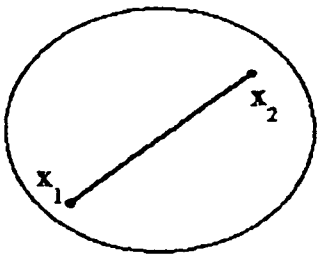
Definição: Um conjunto C no E^n(espaço euclidiano) é dito ser convexo se para qualquer $\mathbf{x}_1,\mathbf{x}_2 \in C$ e qualquer número real α , $0<\alpha<1$, o ponto $\alpha\mathbf{x}_1+(1-\alpha)\mathbf{x}_2 \in C$.

Esta definição pode ser interpretada geométricamente como: se dois pontos quaisquer pertencente ao conjunto forem unidos através de um segmento de reta, qualquer ponto tomado ao longo deste segmento pertence ao conjunto, ver fig.03.

FIGURA 03 - CONVEXIDADE DOS CONJUNTOS



3.a Conjunto não-convexo



3.b Conjunto convexo

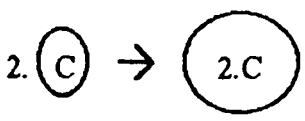
Existem algumas operações que preservam a convexidade dos conjuntos. São elas:

i) Se C é um conjunto convexo e β é um número real, o conjunto $\beta C = \{x: x=\beta c , c \in C\}$, é convexo.

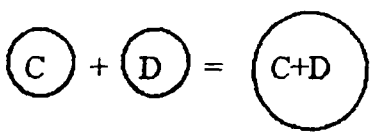
ii) Se C e D são conjuntos convexos, então o conjunto $C+D= \{x: x=c+d, c \in C, d \in D\}$, é convexo.

iii) A interseção de um número arbitrário de conjuntos convexos é um conjunto convexo.

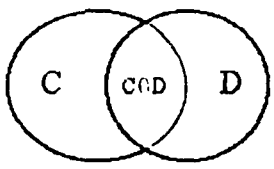
FIGURA 04 - PROPRIEDADES DOS CONJUNTOS CONVEXOS



4.a propriedade i



4.b propriedade ii



4.c propriedade iii

2.6.2 Funções convexas

Uma função f definida em um conjunto convexo Ω é dita ser convexa se, $\forall\ \mathbf{x}_1, \mathbf{x}_2 \in \Omega$ e $\forall\ \lambda,\ 0 \le \lambda \le 1$, ocorrer que:

$$f(\lambda \mathbf{x}_2 + (1-\lambda)\mathbf{x}_1) \le \lambda f(\mathbf{x}_2) + (1-\lambda) f(\mathbf{x}_1) . \qquad (2.33)$$

Se

$$f(\lambda \mathbf{x}_2 + (1-\lambda)\mathbf{x}_1) < \lambda f(\mathbf{x}_2) + (1-\lambda) f(\mathbf{x}_1) \qquad (2.34)$$

então, f é dita ser estritamente convexa.

Geometricamente, uma função é convexa se a linha que une os pontos $(\mathbf{x}_1, f(\mathbf{x}_1))$ e $(\mathbf{x}_2, f(\mathbf{x}_2))$, não ficar abaixo do gráfico da função.

Uma função f é concava se -(f) for convexa.

2.6.2.1 Combinações de funções convexas

Proposição 1. Seja f_1 e f_2 funções convexas em um conjunto convexo Ω. Então a função $f_1 + f_2$ é convexa em Ω.

Prova: seja $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ e $0 < \alpha < 1$, então:

$$f_1(\alpha \mathbf{x}_2 + (1-\alpha)\mathbf{x}_1) + f_2(\alpha \mathbf{x}_2 + (1-\alpha)\mathbf{x}_1) \le \alpha[f_1(\mathbf{x}_2) + f_2(\mathbf{x}_2)] + (1-\alpha)[f_1(\mathbf{x}_1) + f_2(\mathbf{x}_1)] =$$

$$= \alpha(f_1 + f_2)(\mathbf{x}_2) + (1-\alpha)(f_1 + f_2)(\mathbf{x}_1) .$$

Proposição 2. Seja f uma função convexa em um conjunto convexo Ω. Então, af é convexa para qualquer $a \ge 0$.

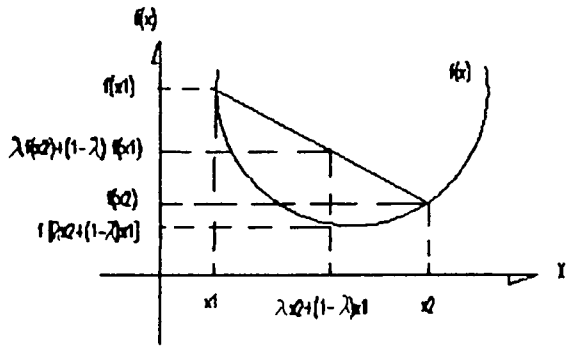
Prova: imediata.

A partir das duas proposições acima pode ser demonstrado que uma combinação positiva de funções convexas resulta numa função convexa.

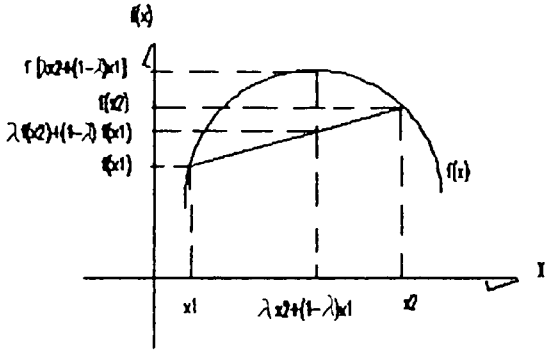
Proposição 3. Seja f uma função convexa em um conjunto convexo Ω. O conjunto $T_c = \{\mathbf{x} : \mathbf{x} \in \Omega,\ f(\mathbf{x}) \le c\}$ é convexo para qualquer número real c.

Prova: Seja $x_1, x_2 \in T_c$. Então $f(x_1) \leq c$, $f(x_2) \leq c$ e para $0<\alpha<1$, $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2) \leq c$. Assim $\alpha x_1 + (1-\alpha)x_2 \in T_c$.

FIGURA 05 - FUNÇÕES CONCAVAS E CONVEXAS PARA O CASO UNIDIMENSIONAL.



5.a Função convexa



5.b Função concava

Note que, como consequência é também convexo o conjunto de pontos satisfazendo a: $f_1(x) \leq c_1$, $f_2(x) \leq c_2$, . . . , $f_m(x) \leq c_m$, onde cada f_i, $1 \leq i \leq m$, é uma função convexa. Como caso particular disso, obtem-se que o conjunto solução de um sistema de equações lineares é um conjunto convexo.

2.6.2.2 Propriedades de funções convexas diferenciaveis

Uma definição equivalente a (2.33) pode ser formulada usando diferenciação de função. A proposição 4 a seguir é formulada para funções de classe C^1 * e a proposição 5 para funções de classe C^2 **.

---

* Uma função cuja derivadas parciais existem e são contínuas é dita ser de classe C^1.

** Uma função cuja derivadas parciais até a segunda ordem existem e são contínuas é dita ser de classe C^2.

Proposição 4. Seja $f \in C^1$. Então, f é convexa em um conjunto convexo Ω, se e somente se

$$f(x_2) \geq f(x_1) + \nabla f^T(x_1)(x_2 - x_1) \ , \ \forall \ x_1, x_2 \in \Omega . \qquad (2.35)$$

Prova: primeiro suponha f ser convexa. Então, a partir da definição (2.33) tem-se que :

$$f(\lambda x_2 + (1-\lambda)x_1) \leq \lambda f(x_2) + (1-\lambda) f(x_1) \quad \text{ou}$$

$$f(x_1 + \lambda(x_2 - x_1)) \leq f(x_1) + \lambda(f(x_2) - f(x_1)) \ , \text{ assim para } 0 < \lambda \leq 1$$

$$\frac{f(x_1 + \lambda(x_2 - x_1)) - f(x_1)}{\lambda} \leq f(x_2) - f(x_1)$$

levando ao limite com $\lambda \to 0$, vem que

$$f(x_2) - f(x_1) \geq \nabla f^T(x_1)(x_2 - x_1)$$

ou

$$f(x_2) \geq f(x_1) + \nabla f^T(x_1)(x_2 - x_1) \ .$$

Com isso fica demonstrada a parte " somente se ".

Assuma agora

$$f(x_2) \geq f(x_1) + \nabla f^T(x_1)(x_2 - x_1) \ , \ \forall \ x_1 \ , \ x_2 \in \Omega .$$

Tomando dois pontos fixos y_1 e $y_2 \in \Omega$, $0 \leq \lambda \leq 1$ e chamando $x_1 = \lambda(y_1) + (1-\lambda)(y_2)$, sendo que alternativamente $x_2 = y_1$ ou $x_2 = y_2$, tem-se que :

$$f(y_1) \geq f(x_1) + \nabla f^T(x_1)(y_1 - x_1) \qquad (2.36)$$

$$f(y_2) \geq f(x_1) + \nabla f^T(x_1)(y_2 - x_1) \qquad (2.37)$$

multiplicando (2.36) por λ , (2.37) por $(1-\lambda)$ e somando ambas resulta

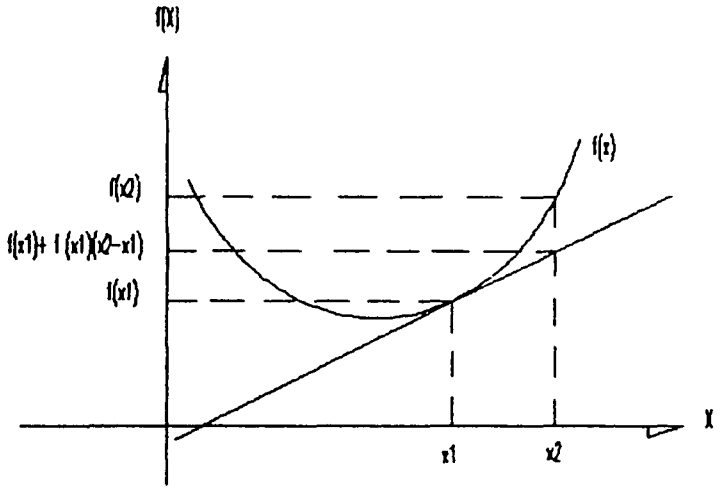
$$\lambda f(y_1) + (1-\lambda) f(y_2) \geq f(x_1) + \nabla f^T(x^1)[\lambda y_1 + (1-\lambda) y_2 - x_1]$$

substituindo x_1 , tem-se

$$\lambda f(y_1) + (1-\lambda) f(y_2) \geq f(\lambda y_1 + (1-\lambda) y_2) \ .$$

Com isso completa-se a demonstração da proposição 4.

FIGURA 6 - DEFINIÇÃO DE FUNÇÃO CONVEXA POR DIFERENCIAÇÃO



A primeira definição afirma que a interpolação linear entre dois pontos superestima a função, enquanto que a segunda afirma que a aproximação linear baseada na derivada local subestima a função.

Para funções duas vezes continuamente diferenciaveis, existe uma outra caracterização de convexidade.

Proposição 5. Seja $f \in C^2$. Então f é convexa em um conjunto convexo Ω contendo um ponto interior*, se e somente se, a matriz **hessiana** $\mathbf{H(x)}=\dfrac{\partial^2 f(x)}{\partial x_i \partial x_j}$ é semi-definida positiva através de Ω.

Prova: Do teorema de Taylor , tem-se:

$$f(x_2)= f(x_1)+\nabla f^T(x_1)(x_2-x_1)+\frac{1}{2}(x_2-x_1)^T H(x_1+\lambda(x_2-x_1))(x_2-x_1) \quad (2.38)$$

para algum λ, $0\leq\lambda\leq 1$. Da definição de matriz semi-definida positiva, $(x_2-x_1)^T H(x_1+\lambda(x_2-x_1))(x_2-x_1)\geq 0$, sendo que na desigualdade estrita, H é definida positiva.

Vê-se claramente que se H em (2.38) é semi-definida positiva, resulta que :

---

* Um ponto é interior quando todas as direções são factíveis.

$$f(\mathbf{x}_2) \geq f(\mathbf{x}_1) + \nabla f^T(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) \qquad (2.39)$$

e com vista na proposição 4, implica que f é convexa.

Suponha agora que a matriz hessiana não seja semi-definida positiva em algum ponto $\mathbf{x}_1 \in \Omega$. Pela continuidade da hessiana pode ser assumido sem perda da generalidade, que $\mathbf{x}_1$ é um ponto interior de Ω. Existe um ponto $\mathbf{x}_2 \in \Omega$ tal que $(\mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{H}(\mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1))(\mathbf{x}_2 - \mathbf{x}_1) < 0$. Novamente pela continuidade da hessiana, $\mathbf{x}_2$ pode ser selecionado de modo que λ, $0 \leq \lambda \leq 1$,

$$(\mathbf{x}_2 - \mathbf{x}_1)^T \mathbf{H}(\mathbf{x} + \lambda(\mathbf{x}_2 - \mathbf{x}_1))(\mathbf{x}_2 - \mathbf{x}_1) < 0.$$

Com vista em (2.38), implica que (2.39) não ocorre e pela proposição 4, implica que f é não-convexa.

2.6.3 Extremo de funções de valores reais

Esta secção está sendo introduzida com o objetivo de definir os conceitos sobre maximização e minimização de funções de várias variáveis.

Não se pretende aqui demonstrar as condições de otimalidade, mas tão somente relembrar essas condições para aplicação no problema que será tratado.

Definição: Se $f: U \subset \mathbf{R}^n \rightarrow \mathbf{R}$ é uma função de valor real, um ponto $\mathbf{x}^* \in U$ é chamado um *mínimo local* de f se existe uma vizinhança V de $\mathbf{x}^*$ tal que para todo ponto $\mathbf{x} \in V$, $f(\mathbf{x}) \geq f(\mathbf{x}^*)$. Similarmente, $\mathbf{x}^* \in U$ é um *máximo local* se existe uma vizinhaça V de $\mathbf{x}^*$ tal que $f(\mathbf{x}) \leq f(\mathbf{x}^*)$ para todo $\mathbf{x} \in V$. $\mathbf{x}^* \in U$ é dito ser *extremo local ou relativo* se ele for um máximo local ou um mínimo local. Um ponto $\mathbf{x}^*$ é um *ponto crítico* de f se $\nabla f(\mathbf{x}^*) = 0$. Um ponto crítico que não é extremo local é chamado *ponto de sela*.

Um ponto $\mathbf{x}^*$ é crítico se $\nabla f(\mathbf{x}^*)=0$, isto quer dizer que ele poderá ser ou um ponto de máximo, ou um ponto de mínimo, ou um ponto de sela.

Para saber qual a classe que o ponto $\mathbf{x}^*$ pertence, pode ser feita a seguite análise na matriz de derivas segunda (**hessiana**) da função.

1. O ponto $\mathbf{x}^*$ será um ponto de mínimo se a matriz hessiana for definida positiva (autovalores todos positivos).
2. O ponto $\mathbf{x}^*$ será um ponto de máximo se a matriz hessiana for definida negativa (autovalores todos negativos).
3. O ponto $\mathbf{x}^*$ será um ponto de sela se a matriz hessiana for indefinida (autovalores positivos e negativos).
4. Se a matriz hessiana for ou semi-definida positiva ou semi-definida negativa, o ponto crítico $\mathbf{x}^*$ é chamado **degenerado** e uma análise da função nas vizinhanças do ponto $\mathbf{x}^*$ deve ser feita para saber que tipo extremo será.

2.6.4 Minimização de função convexa

Proposição 1. Seja f uma função convexa definida em um conjunto convexo Ω. Então, o conjunto Γ onde f alcança seu ponto de mínimo, é convexo e qualquer mínimo local de f é mínimo global .

Prova: Se f não tem mínimo local, a proposição é aceita por vacuidade. Assumindo ser C_0 o mínimo de f, pela proposição 3 da secção (2.6.2.1), o conjunto $\Gamma_C = \{\mathbf{x}:\mathbf{x}\in\Omega, f(\mathbf{x})\leq C_0\}$ é convexo.

Suponha agora, $\mathbf{x}^* \in \Omega$ ser um ponto de mínimo local de f e que exista outro ponto $y \in \Omega$, tal que $f(y)<f(\mathbf{x}^*)$. Então, na linha $\lambda y+(1-\lambda)\mathbf{x}^*$, $0<\lambda<1$, tem-se :

$$f(\lambda y+(1-\lambda)\mathbf{x}^*)\leq \lambda f(y)+(1-\lambda)f(\mathbf{x}^*)< f(\mathbf{x}^*)$$

contradizendo ao fato de que $\mathbf{x}^*$ é mínimo local.

Proposição 2. Seja $f \in C^1$ e convexa em conjunto convexo Ω. Então, todo ponto de mínimo local em f é ponto de mínimo global.

Prova: Suponha que $\mathbf{x}$ e $\mathbf{y} \in \Omega$ sejam pontos de mínimo local de f. Além disso, suponha ser $f(\mathbf{y})<f(\mathbf{x})$. Assim, pela proposição 4 da secção (2.6.2.2),

$f(\mathbf{y})-f(\mathbf{x}) \geq \nabla f^T(\mathbf{x})(\mathbf{y}-\mathbf{x})$.

Isto implica que $\nabla f^T(\mathbf{x})(\mathbf{x}-\mathbf{y})<0$, o que significa que a função f pode ser decrescida a partir do ponto $\mathbf{x}$, contradizendo ao fato de que $\mathbf{x}$ é ponto de mínimo local.

Maiores detalhes a respeito da secção 2.6, podem ser encontrados em LUENBERGER (1973) e RAO (1979).

2.7 CONVERGÊNCIA DE SEQUÊNCIAS DE NÚMEROS REAIS

seja $x^* \in \mathbf{R}$ e $x_i \in \mathbf{R}$, i=0,1,2,... , então , a sequência de números reais $\{x_i\}=\{x_0,x_1,x_2,...\}$ é dita convergir para x^* se

$$\lim_{i\to\infty}|xi - x^*| = 0$$

Além disso, se existe uma constante $c \in \mathbf{R}$ e um inteiro $\hat{i} \geq 0$ tal que $\forall\ i \geq \hat{i}$

$$| x_{i+1}-x^*| \leq c|x_i-x^*| \qquad (2.40)$$

então, $\{x_i\}$ é dita ser q-linearmente convergênte para x^*.

Se para alguma sequência $\{c_i\}$ que converge para zero

$$|x_{i+1}-x^*| \leq c_i|x_i-x^*| \qquad (2.41)$$

então $\{x_i\}$ converge q-superlinearmente para x^*.

Se existe $p>1$, $c\geq 0$ e $\hat{i} \geq 0$ tal que $\{x_i\}$ converge para x^* e $\forall\ i \geq \hat{i}$

$$|x_{i+1}-x^*| \leq c|x_i-x^*|^p \qquad (2.42)$$

então $\{x_i\}$ é dita convergir para x_i com pelo menos q-ordem p.

No caso particular de p=2, a convergência da sequência $\{x_i\}$ é dita ser de ordem q-quadrática.

Mais detalhes a respeito desta seção são encontrados em DENIS e SCHNABEL (1983).

3. SOLUÇÃO DO PROBLEMA DE MÍNIMOS QUADRADOS LINEAR

O problema de mínimos quadrados linear foi definido na introdução como: minimizar o quadrado do resíduo (Ax-b) na norma euclidiana, ou seja,

$$\min_{x \in R^n} \|Ax - b\|_2^2 = \min \|V\|_2^2 \qquad (3.1)$$

onde:

$V = Ax-b$

$A \in R^{m \times n}$;

$b \in R^m$.

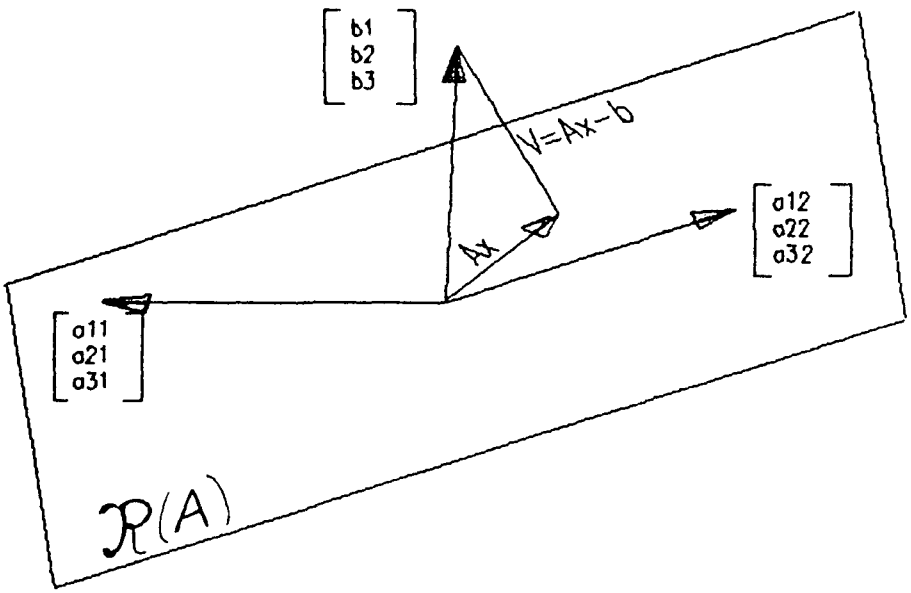
3.1 PRELIMINARES

Antes de encontrar a solução em si, será feito um breve comentário em torno do significado do resíduo Ax-b. Já se sabe do capítulo anterior que um sistema de equacões lineares só possui solução se o vetor b for combinação linear das colunas de A, ou seja, se o vetor $b \in \Re(A)$. Caso $b \in \Re(A)$, o critério de mínimos quadrados tornar-se-ia irrelevante, pois o problema resultaria em apenas resolver um sistema de equações lineares compatível.

Uma vez que $b \notin \Re(A)$, o sistema é impossível e a aplicação do critério de minímos quadrados torna-se relevante, pois trata-se de escolher dentre as soluções aproximadas existentes, aquela que minimiza o quadrado de uma certa norma .

Como qualquer vetor $Ax \in \Re(A)$, $b \in R^m$, e $\Re(A)$ é subespaço próprio do R^m, então em geral $b \notin \Re(A)$. Ao vetor Ax-b denomina-se resíduo.

FIGURA 07 - REPRESENTAÇÃO DO RESÍDUO



Da própria ilustração pode ser observado que existe infinitos resíduos **V** . A questão é escolher aquele de menor norma.

Eis uma outra questão; qual a norma a utilizar e por que ? já na própria definição do problema foi proposto a norma-2. Abaixo segue algumas justificativas da escolha dessa norma.

a) A norma-2 é diferenciavel para todo $\mathbf{x} \in \mathbf{R}^n$, tal que $\mathbf{Ax-b} \neq 0$.

b) Esta norma tem interpretação geométrica simples e as leis de projeção são válidas para ela.

c) Tem justificação estatística; é um estimador não-tendencioso e fornece variância mínima. Além disso, se as observações seguem uma distribuição normal, a estimativa de mínimos quadrados, é também estimativa de máxima verossimilhança (máxima probabilidade).

Cabe apenas citar que minimizar **Ax-b** na norma-1 ou na norma infinita, é equivalente a resolver um problema da programação linear.

3.2 SOLUÇÃO DO PROBLEMA

Seja $f: \mathbf{R}^n \rightarrow \mathbf{R}$ uma função definida por $f(x)=\|Ax-b\|_2^2$. O problema (3.1) pode ser escrito como:

$$\min_{x \in R^n} f(x)= \|Ax-b\|_2^2 \qquad (3.2)$$

usando a definição da norma-2 tem-se:

$f(x) = \langle Ax-b, Ax-b \rangle = \langle Ax, Ax \rangle - 2\langle Ax, b \rangle + \langle b, b \rangle.$ (3.3)

aplicando a condição de otimalidade em f(x) ($\nabla f(x)=0$), tem-se:

$\nabla f(x)=2(A^TA)x-2(A^Tb)=0$, donde

$$(A^TA)x^* = A^Tb \qquad (3.4)$$

onde: x^*, é a solução de ótimo para o problema (3.2) . A expressão (3.4) é conhecida como *equação normal de mínimos quadrados.*

Uma vez aplicada a condição de otimalidade ao problema e obtida a solução x^* de ótimo, deve ser feita uma análise para saber de qual ponto esta solução se refere (máximo, mínimo, ou sela). Baseado nas afirmações da secção (2.6.3), deve ser calculada a matriz hessiana da função f(x), a qual será:

$$H(x)= 2(A^TA). \qquad (3.5)$$

Agora ficou bastante simples para analisar x^*, haja visto que H(x) é no mínimo semi-definida positiva.

Prova: Por definição, uma matriz $H \in \mathbf{R}^{n \times n}$ é semi-definida positiva se $\forall\, x \in \mathbf{R}^n$, com $x \neq 0$ $x^THx \geq 0$, sendo que na desigualdade estrita, H é definida positiva.

Assim

$$x^TA^TAx = \langle Ax, Ax \rangle = \|Ax\|_2^2 \geq 0. \qquad (3.6)$$

Então, $Ax=0$ se e somente se, as colunas de A são linearmente dependentes, o que significa que A não tem posto completo. Logo, para qualquer sistema de equações lineares redundantes, onde a matriz dos

coeficientes das incógnitas tiver posto completo, e nele for aplicado o critério de mínimos quadrados, a solução x^* será ponto de mínimo, salvo em sistemas "mal-condicionados" (será definido adiante os termos mal-condicionados e bem-condicionados), onde devido a erros de arredondamento do computador a matriz hessiana pode tornar-se semi-definida positiva ou até indefinida.

A esta altura já se sabe que x^* corresponde ao ponto que minimiza a função residual f(x) definida em (3.2) e (3.3). Resta saber se este ponto é ponto de mínimo local ou mínimo global.

Nas proposições 1 e 2 da secção (2.6.4) ficou demonstrado que se uma função f é convexa em um conjunto convexo, então, todo ponto de mínimo local é ponto de mínimo global. Em virtude desse fato, deve-se provar que f(x) é convexa em um conjunto convexo.

Prova: o R^n de fato é convexo, para verificar basta ligar dois pontos quaisquer do conjunto por uma reta, que qualquer ponto tomado ao longo dessa reta pertencerá ao conjunto.

A convexidade de f(x) é assegurada pela proposição 5 da secção (2.6.2.2), uma vez que a hessiana é semi-definida positiva através do R^n .

Resta para concluir esta secção, saber quanto da unicidade de x^*. Isto pode ser verificado pela proposição a seguir:

Proposição 1. Numa função convexa $f \in C^2$, onde $f: R^n \rightarrow R$, se existir mais de um ponto de mínimo x^*, então nesta função existirão infinitos pontos de mínimo.

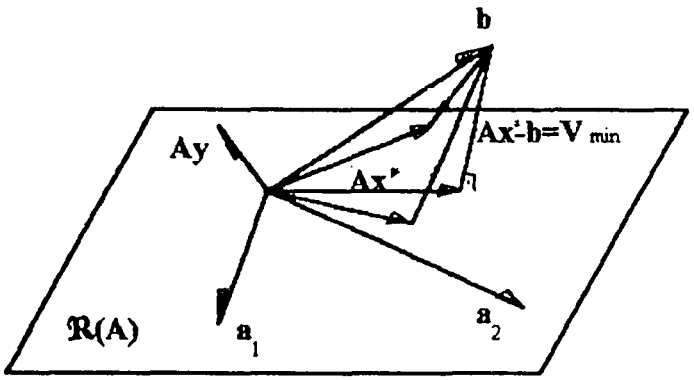
Prova:

Se x_1 e x_2 são pontos de mínimo então toda combinação convexa de x_1 e x_2 é ponto de mínimo, pois o conjunto dos pontos de mínimo como foi visto na proposição 3 (secção 2.6.2.1) é um conjunto convexo.

3.2.1 Interpretação geométrica

A solução do problema de mínimos quadrados linear também pode ser obtida por meio da geometria.

FIGURA 08 - INTERPRETAÇÃO GEOMÉTRICA DA SOLUÇÃO DE MÍNIMOS QUADRADOS.



A (fig.08) mostra três dos infinitos resíduos que podem ocorrer. Na própria figura pode ser observado que o resíduo de menor comprimento é aquele no qual é obtido pela projeção ortogonal de **b** em $\Re(A)$, ou seja, $V_{min} = \|Ax^* - b\|$. Como qualquer vetor do tipo $Ay \in \Re(A)\ \forall\ y \in R^n$, e para que **Ay** seja perpendicular ao resíduo de comprimento mínimo, deve-se ter

$$< Ay, Ax^* - b >= \|Ay\|\|Ax^* - b\|\cos 90° = 0 \quad , \quad \text{ou}$$

$$y^T A^T Ax^* - y^T A^T b = y^T (A^T Ax^* - A^T b) = 0 \qquad (3.7)$$

como y pode ser escolhido arbritrário, então

$$A^T Ax^* - A^T b = 0 \quad \Rightarrow \quad A^T Ax^* = A^T b \qquad (3.8)$$

que é a mesma (3.4) encontrada anteriormente.

Quanto a questão de uniciade de x^*. Na secção anterior foi visto que se A^TA for definida positiva, x^* é único ponto de mínimo da função $f(x)=\|Ax-b\|_2^2$ e se A^TA for semi-definida positiva (extremo degenerado), existiram infinitos pontos de mínimo x^* para a função $f(x)$.

Por causa da solução através das decomposições ortogonais(resolvem o problema diretamente sem usar as equações normais), é importante demonstrar que posto (A) e posto(A^TA) são iguais.

Proposição 1. Para qualquer matriz $A_{(mxn)}$ de posto(A)=k, o produto matricial A^TA resulta numa matriz simétrica e seu posto também é k.

Prova:

Primeira parte - verificação da simetria de A^TA.

Se A^TA é simétrica, então $(A^TA)^T=A^TA$.. pelas propriedades de transposição das matrizes, tem-se

$(A^TA)^T=A^T(A^T)^T=A^TA$.

Segunda parte - verificação do posto(A) e posto(A^TA).

Se for verificado que A e A^TA tem o mesmo espaço nulo, então fica demonstrado a igualdade entre o posto(A) e posto(A^TA), já que a dimensão do espaço nulo é o número de colunas menos o posto, e tanto A como A^TA tem (n) colunas.

Se x está no espaço nulo de A, então $Ax=0$ e se x está no espaço nulo de A^TA , então $A^TAx=0$. Como $Ax=0$ implica que $A^T0=0$, mostrando que x está em ambos espaços nulos.

Outro caminho para a verificação seria tomando o produto interno com x^T, assim

$x^TA^TAx=x^T0 \Rightarrow x^TA^TAx=0$, o que nada mais é que

$\|Ax\|_2^2=0$, logo $Ax=0$.

Se posto(A)=k=n , então A^TA é não-singular e com isso inversivel, podendo assim fornecer uma solução alternativa para (3.8), ou seja,

$$x^* = (A^T A)^{-1} (A^T b) \; . \qquad (3.9)$$

3.2.2 Propriedades da solução de mínimos quadrados

Nesta secção será feito o estudo do problema de mínimos quadrados utilizando os subespaços envolvidos, a fim de tirar alguns resultados para a análise do problema.

Recordando que o espaço coluna de A ($\Re(A)$),consiste de todos os vetores m-dimensionais que são combinações lineares das colunas de A. O subespaço complementar é o espaço nulo de A^T ($\pi(A^T)$), ou seja, contém todos os vetores m-dimensionais que são ortogonais as colunas de A ($A^T y = 0$).

A dimensão do espaço coluna de A é igual ao posto(A)=k e a dimensão do espaço nulo de A^T é igual a m-k.

3.2.2.1 Caracterização do ótimo residual

Dada uma matriz A, qualquer vetor não nulo c m-dimensional pode ser expresso como a soma de um vetor $c_R \in \Re(A)$ e um $c_N \in \pi(A^T)$, ou mais simplesmente

$$c = c_R + c_N \qquad (3.10)$$

onde:

c_R = componente de c em $\Re(A)$

c_N = componente de c em $\pi(A^T)$

com , $c_R^T c_N = 0$.

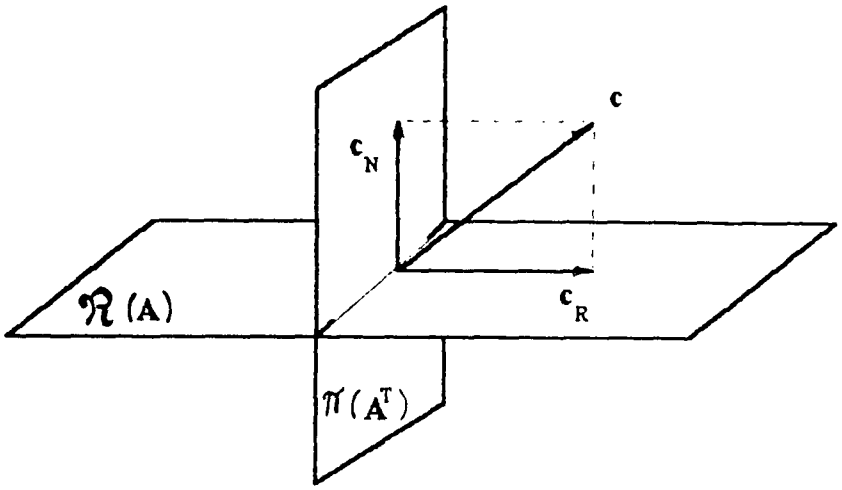
Os componentes c_R e c_N são únicos e satisfazem

$c_R = A c_A$, para algum $c_A \in R^n$;

$A^T c_N = 0$.

Mais fácil seria compreender a unicidade dos vetores $\mathbf{c}_R$ e $\mathbf{c}_N$, observando a figura abaixo.

FIGURA 09 - COMPONENTES DE UM VETOR EM $\Re(A)$ E $\pi(A^T)$



$\mathbf{c}_R$ é a projeção de **c** no espaço coluna de **A** e $\mathbf{c}_N$ é a projeção de **c** no espaço nulo de A^T, desta maneira só existem um $\mathbf{c}_R$ e um $\mathbf{c}_N$.

Embora $\mathbf{c}_R$ seja único, $\mathbf{c}_N$ só será único se as colunas de **A** forem linearmente independentes.

Devido as propriedades que a norma-2 tem na geometria euclidiana, (3.10) pode ser escrita como:

$$\|\mathbf{c}\|_2^2 = \|\mathbf{c}_R\|_2^2 + \|\mathbf{c}_N\|_2^2 \ . \qquad (3.11)$$

Estas observações são de extrema relevância para o problema de mínimos quadrados, uma vez que **b** e **V** são vetores m-dimensionais e estão relacionados com os subespaços mensionados. Desta forma, **b** e **V** podem ser escritos como:

$$\mathbf{b} = \mathbf{b}_R + \mathbf{b}_N \ , \ \text{com} \ \mathbf{b}_R = \mathbf{A}\mathbf{b}_A \ ; \qquad (3.12)$$

$$\mathbf{V} = \mathbf{v}_R + \mathbf{v}_N \ , \ \text{com} \ \mathbf{v}_R = \mathbf{A}\mathbf{v}_A \ . \qquad (3.13)$$

Escrevendo o residual V em termos de V=Ax-b

$$\mathbf{V} = \mathbf{v}_R + \mathbf{v}_N = \mathbf{Ax} - \mathbf{b}_R - \mathbf{b}_N \qquad (3.14)$$

como $\mathbf{Ax} \in \Re(\mathbf{A})$, $\forall\ \mathbf{x} \in R^n$, então

$$\mathbf{v}_R = \mathbf{A}_X - \mathbf{b}_R \quad \text{e} \quad \mathbf{v}_N = -\mathbf{b}_N \,. \qquad (3.15)$$

Observando (3.14), pode-se notar que ao sutrair Ax de b, não se altera o componente de b que está no espaço nulo de A^T, assim

$$\|\mathbf{Ax} - \mathbf{b}\|_2^2 = \|\mathbf{V}\|_2^2 = \|\mathbf{v}_R\|_2^2 + \|\mathbf{v}_N\|_2^2 = \|\mathbf{Ax} - \mathbf{b}_R\|_2^2 + \|\mathbf{b}_N\|_2^2 \qquad (3.16)$$

o que implica ser

$$\|\mathbf{V}\|_2^2 \geq \|\mathbf{b}_N\|_2^2 \qquad (3.17)$$

logo, minimizar $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ corresponde em minimizar $\|\mathbf{Ax} - \mathbf{b}_R\|_2^2$. Como $\mathbf{b}_R \in \Re(\mathbf{A})$, então deve existir algum vetor x n-dimensional tal que $\mathbf{Ax}-\mathbf{b}_R=\mathbf{0}$ e desta forma o resíduo (V) atinge seu mínimo quando $\|\mathbf{V}\|_2^2 = \|\mathbf{b}_N\|_2^2$.

Isto conduz a duas importântes caracterizações equivalentes, as quais são:

x minimiza $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ se e somente se $\mathbf{A}^T(\mathbf{Ax} - \mathbf{b}) = \mathbf{0}$;

x minimiza $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ se e somente se $\mathbf{Ax} = \mathbf{b}_R$ e $\|\mathbf{Ax} - \mathbf{b}\|_2^2 = \|\mathbf{b}_N\|_2^2$.

Um problema de mínimos quadrados é dito ser compatível se o residual ótimo ($\mathbf{b}_N$) é zero, isso significa exatamente que o vetor b está no espaço coluna de A. Caso seja $\mathbf{b}_N \neq \mathbf{0}$ o problema é dito ser incompativel. Para ilustrar essas propriedades considere o seguinte exemplo:

Seja

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad , \quad \mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 1 \end{bmatrix}$$

Um vetor pertencente ao espaço coluna de A tem a forma

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 + c_2 \\ c_1 \\ c_2 \end{bmatrix}$$

para qualquer escalar c_1 , c_2 .

Um vetor pertencente ao espaço nulo de $\mathbf{A}^T$ é dado por

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

sendo que $d_2 = -d_1$ e $d_3 = -d_1$. Assim para um escalar c_3 o vetor pertencente ao espaço nulo de $\mathbf{A}^T$ será $(c_3 \ -c_3 \ -c_3)^T$. Como $\mathbf{b} = \mathbf{b}_R + \mathbf{b}_N$, então

$$\begin{bmatrix} 3 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 + c_2 \\ c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} c_3 \\ -c_3 \\ -c_3 \end{bmatrix}$$

ou

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 1 \end{bmatrix}$$

cuja solução para c_1 , c_2 e c_3 é : $c_1 = 4$, $c_2 = 0$ e $c_3 = -1$. Assim os componentes de $\mathbf{b}$ são: $\mathbf{b}_R = (4 \ 4 \ 0)^T$ e $\mathbf{b}_N = (-1 \ 1 \ 1)^T$.

A solução de $\mathbf{A}\mathbf{x}^* = \mathbf{b}_R$, trará o ótimo residual $\mathbf{V} = \mathbf{b}_N = (-1 \ 1 \ 1)^T$ de modo que

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix}$$

resulta em

$\mathbf{x}^* = (4 \ 4)^T$.

3.3 SOLUÇÃO DE COMPRIMENTO MÍNIMO DO PROBLEMA DE MÍNIMOS QUADRADOS LINEAR

Foi visto anteriormente que no sistema $Ax=b$ incompatível, a solução de mínimos quadrados é obtida através das equações normais de mínimos quadrados, cuja uniciadade só se verificará se A tiver posto completo, caso contrário ter-se-a infinitas soluções.

O estudo desta secção trata exatamente de quando A não tem posto completo e busca escolher dentre as infinitas soluções existentes, aquela que tem menor comprimento na norma-2.

O problema pode ser formulado como:

Dado $A \in R^{mxn}$, $b \in R^m$, sendo posto(A)<n

$$\min_{x \in R^n} \{ \|x\|_2 : \|Ax-b\|_2^2 \leq \|A\hat{x}-b\|_2^2 \quad \forall \hat{x} \in R^n \} \qquad (3.18)$$

Uma vez formulado o problema, será fornecido alguns conceitos para facilitar a compreensão da solução do mesmo.

3.3.1 Projeções

Definição: Seja E um espaço vetorial, E' e E'' dois subespaços vetoriais complementares em E e P a projeção sobre E paralelamente a E''. A imagem por P de um vetor b de E denomina-se projeção de b sobre E' paralelamente a E''.

Particularizando a definição geral acima para:

Seja $R^m = E$, $S = E'$ e $S^{\perp} = E''$, onde $S^{\perp}$ pode ser melhor representado por

$S^{\perp} = \{ t \in R^m / t^T s = 0 \ \forall \ s \in S \}$.

Uma vez que S e $S^{\perp}$ são complementares em R^m , então

$S + S^{\perp} = R^m$ e $S \cap S^{\perp} = \{\ \}$.

Qualquer vetor **b** m-dimensional pode ser representado como $\mathbf{b} = \mathbf{b}_s + \mathbf{b}_{s^{\perp}}$, onde $\mathbf{b}_s \in S$ e $\mathbf{b}_{s^{\perp}} \in S^{\perp}$.

A projeção $P_{(m,m)}$ sobre **S**, denotado por P_s é a única matriz que possui as três seguintes propriedades.

i) Qualquer vetor do subespaço S pode ser escrito como combinação linear das colunas de P_s, isto é, $\mathbf{b}_s \in S$ se e somente se $\mathbf{b}_s = P_s \mathbf{v}$, para algum $\mathbf{v} \in R^m$;

ii) $P_s^T = P_s$ (a matriz P_s é simétrica);

iii) $P_s^2 = P$ (a matriz P_s é idempotente).

O significado da projeção P_s conforme definição, diz que para qualquer vetor **b** m-dimensional, a aplicação $P_s \mathbf{b} = \mathbf{b}_s$ e $P_s \mathbf{b}_s = \mathbf{b}_s$ e ainda $P_s \mathbf{b}_{s^{\perp}} = \mathbf{0}$.

A projeção sobre o complemento ortogonal de S pode ser obtido facilmente, fazendo

$\mathbf{b} = \mathbf{b}_s + \mathbf{b}_{s^{\perp}}$ $\therefore$ $\mathbf{b}_{s^{\perp}} = \mathbf{b} - \mathbf{b}_s$, mas

$(I - P_s)\mathbf{b} = \mathbf{b} - P_s\mathbf{b} = \mathbf{b} - \mathbf{b}_s = \mathbf{b}_{s^{\perp}}$, logo conclui-se que a matriz $(I - P_s)$ é a projeção sobre $S^{\perp}$, a qual também cumpre com as três propriedades acima descritas.

3.3.1.1 Projeções relacionadas ao problema de mínimos quadrados

Através da secção (3.2.2.1) foi visto que **x** minimiza $\|A\mathbf{x} - \mathbf{b}\|_2^2$ se e somente se $A\mathbf{x} = \mathbf{b}_R$ e $V = \mathbf{b}_N$, sendo $\mathbf{b}_R$ o componente de **b** em $\Re(A)$ e $\mathbf{b}_N$ o componente de **b** em $\pi(A^T)$.

Em vista disso, pode-se dizer que existe uma projeção P sobre o espaço coluna $\Re(A)$ que produz $\mathbf{b}_R$ e uma projeção (I-P) sobre $\pi(A^T)$ que produz $\mathbf{b}_N$.

Tais projeções podem ser facilmente obtidas. Lembrando que

$\mathbf{x}^* = (A^T A)^{-1} A^T \mathbf{b}$ e $A\mathbf{x}^* = \mathbf{b}_R$, então:

$P_{\Re(A)} = A(A^TA)^{-1}A^T$ e

$P_{\pi(A^T)} = I - A(A^TA)^{-1}A^T$, assim

$P_{\Re(A)}b = b_R$, $P_{\Re(A)}b_R = b_R$ e $P_{\Re(A)}b_N = 0$.

3.3.2 Inversas generalizadas e pseudo-inversa

Define-se **B** como sendo a inversa de uma matriz quadrada **A** quando **AB=I=BA**, onde **B** é denotada por A^{-1}.

Existem várias condições equivalentes para a existência de A^{-1}, uma delas é que $\det(A)\neq 0$.

Quando **A** é quadrada com $\det(A)=0$, ou **A** é retangular, a definição usual acima de inversa não pode mais ser usada, dando lugar a um novo conceito de inversa, as chamadas *inversas generalizadas*.

Usando uma inversa generalizada G, o sistema **Ax=b** pode ser representado explicitamente por **x=Gb** mesmo quando não existe a inversa ordinaria A^{-1}. A inversa generalizada G não é única, existindo, dependendo de suas propriedades denominações como: inversa a esquerda, inversa a direta e outras ver GEMAEL (1977).

Em particular existe uma inversa generalizada capaz de resolver o problema (3.18). Esta é única, sendo também a única que satisfaz totas as condições de Penrose (LAWSON e HANSON 1974)

i) **AGA=A** ;

ii) **GAG=G** ;

iii) $(AG)^T = AG$;

iv) $(GA)^T = GA$;

é denotada por A^+ e conhecida por pseudo-inversa. Assim a solução de (3.18) pode ser representado por

$x = A^+b$

3.3.3 Interpretação geométrica da solução de comprimento mínimo e propriedades da pseudo-inversa.

Um dos caminhos para entender o efeito da pseudo-inversa é através da geometria, o qual será visto a seguir.

Lembrando que $\mathbf{x} \in \mathbf{R}^n$ pode ser escrito como $\mathbf{x} = \mathbf{x}_R + \mathbf{x}_N$,onde $\mathbf{x}_R$ é o componente de $\mathbf{x}$ em $\Re(\mathbf{A}^T)$ e $\mathbf{x}_N$ o componente de $\mathbf{x}$ em $\aleph(\mathbf{A})$.

Considerando a solução de comprimento mínimo como $\mathbf{x}^* = \mathbf{x}_R + \mathbf{x}_N$ e sendo $\Re(\mathbf{A}^T)$ e $\aleph(\mathbf{A})$ complementos ortogonais, o teorema de pitágoras pode ser aplicado usando da norma-2. Assim,

$$\|\mathbf{x}^*\|_2^2 = \|\mathbf{x}_R + \mathbf{x}_N\|_2^2 = \|\mathbf{x}_R\|_2^2 + \|\mathbf{x}_N\|_2^2 .$$

Donde conclui-se facilmente que a solução de comprimento mínimo ($\min \|\mathbf{x}\|_2$), equivale em tornar igual a zero o componente arbitrário $\mathbf{x}_N$ do espaço nulo de A. Desta feita, a projeção $\mathbf{P}_{\Re(A)}\mathbf{b} = \mathbf{b}_R = \mathbf{A}\mathbf{x}^*$, fornece resíduo mínimo ($\mathbf{b}_N$) e $\mathbf{A}\mathbf{x}^* = \mathbf{b}_R \quad \Leftrightarrow \quad \mathbf{A}(\mathbf{x}_R^* + \mathbf{x}_N^*) = \mathbf{b}_R$, tem solução de comprimento mínimo quando $\mathbf{A}\mathbf{x}_R^* = \mathbf{b}_R$, donde $\mathbf{x}_R^* = \mathbf{A}^+\mathbf{b}_R$.

Esse é exatamente o efeito da pseudo-inversa quando se escreve $\mathbf{x}^* = \mathbf{A}^+\mathbf{b}$, ou mais claramente, sua aplicação em $\mathbf{b}$ corresponde ao efeito conjunto de projetar $\mathbf{b}$ em $\Re(\mathbf{A})$ obtendo $\mathbf{b}_R$ (fornece o resíduo mínimo), depois busca o componente de $\mathbf{x}$ que está inteiramente em $\Re(\mathbf{A}^T)$ desprezando a parte arbitrária de $\mathbf{x}^*$ em $\aleph(\mathbf{A})$, fornecendo assim a solução de comprimento mínimo $\mathbf{x}^*$. STRANG (1976) faz uma ilustração para essa interpretação.

A pseudo-inversa tem várias propriedades importantes, algumas delas são listadas abaixo:

i) Se $\mathbf{A}_{(m,n)}$ então $\mathbf{A}^+_{(n,m)}$, $\mathbf{A}^+$ quando aplicada a um vetor $\mathbf{b} \in \mathbf{R}^m$ produz um $\mathbf{x} \in \mathbf{R}^n$;

ii) O espaço coluna de A^+ é o espaço linha de A, e o espaço linha de A^+ é o espaço coluna de A, implicando ser posto(A)=posto(A^+);

iii) A pseudo-inversa de A^+ é a própria A;

iv) Em geral $AA^+ \neq I$, uma vez que A pode não ter a A^d(inversa a direita), mas AA^+ é sempre a projeção $P_{\Re(A)}$;

v) $(A^T)^+=(A^+)^T$;

vi) Se $A \in R^{nxn}$ e posto(A)=n então $A^+ = A^{-1}$.

3.3.4 Determinação da pseudo-inversa através da decomposição de valor singular

Existem vários caminhos para determinar a pseudo-inversa, em geral através de decomposições e em particular com quase unanimidade entre os autores, através da decomposição de valor singular (SVD), a qual será descrita a seguir.

Definição: Seja $A \in R^{mxn}$ e k=min{m,n}. A decomposição de valor singular de A é $A=UDV^T$, onde $U \in R^{mxm}$ e $V \in R^{nxn}$ são matrizes ortogonais, $D \in R^{mxn}$ é definida por $d_{ii} = \sigma_i \geq 0$, i=1,...,k, $d_{ij} = 0$, $i \neq j$. As quantidades $\sigma_1,...,\sigma_k$ são chamados valores singulares de A.

Proposição 1: Seja $A \in R^{mxn}$. Então a SVD de A existe, os elementos diagonal σ_i de D são as raizes quadradas não-negativas dos autovalores de AA^T se $m \leq n$, ou de A^TA se $m \geq n$, e as colunas de U e V são os autovetores de AA^T e A^TA, respectivamente. O número de valores singulares não-nulos é posto(A).

Prova: A prova da existência da SVD de A pode ser encontrada em (LAWSON e HANSON 1974). O restante segue; uma vez que $AA^T = UDD^TU^T$ e

$\mathbf{A^TA = VD^TDV^T}$ tem-se $(\mathbf{AA^T})\mathbf{U} = \mathbf{U}(\mathbf{DD^T})$, $(\mathbf{A^TA})\mathbf{V} = \mathbf{V}(\mathbf{D^TD})$. Assim, se u_j e v_j são respectivamente as j-ésimas colunas de U e V , então:

$(\mathbf{AA^T})u_j = \lambda_j u_j$, j=1,...,m ;

$(\mathbf{A^TA})v_j = \lambda_j v_j$, j=1,...,n,

onde os λ_{js} são os elementos da diagonal de $\mathbf{DD^T}$ e $\mathbf{D^TD}$, $\lambda_j = (\sigma_j)^2$, j=1,...,min{m,n}; e $\lambda_j = 0$, j=min{m,n}+1,...,max{m,n}. Assim a multiplicação por uma matriz não-sigular não muda o posto, posto(A)=posto(D), o número de σ_{js} não-nulos.

Proposição 2. Seja $\mathbf{A} \in \mathbf{R}^{mxn}$ com SVD $\mathbf{A = UDV^T}$, com U,D,V como definida acima. Seja a pseudo-inversa de A definida como

$\mathbf{A^+ = VD^+U^T}$, com

$$\mathbf{D^+} = \begin{cases} d_{ii}^{+} = \begin{cases} 1/\sigma_i, & i > 0 \\ 0 & i = 0 \end{cases} \\ d_{ij}^{+} = 0 \qquad i \neq j \end{cases}$$

$\mathbf{D^+} \in \mathbf{R}^{nxm}$. Então a única solução para o problema (3.18) é $\mathbf{x = A^+b}$.

Prova: A partir de $\mathbf{A = UDV^T}$ o problema (3.18) é equivalente a

$$\min_{x \in R^n} \{\|\mathbf{V^Tx}\|_2 : \|\mathbf{DV^Tx - U^Tb}\|_2 \leq \|\mathbf{DV^T}\hat{\mathbf{x}} - \mathbf{U^Tb}\|_2 \ \forall\, \hat{\mathbf{x}} \in \mathbf{R}^n\} \qquad (3.19)$$

fazendo $\mathbf{z = V^Tx}$ vem

$$\min_{z \in R^n} \{\|\mathbf{z}\|_2 : \|\mathbf{Dz - U^Tb}\|_2 \leq \|\mathbf{D}\hat{\mathbf{z}} - \mathbf{U^Tb}\|_2 \ \forall\, \hat{\mathbf{z}} \in \mathbf{R}^n\}. \qquad (3.20)$$

Seja k o número de σ_{is} não-nulos em D. então:

$$\|\mathbf{Dz - U^Tb}\|_2^2 = \sum_{i=1}^{k} (\sigma_i z_i - (\mathbf{U^Tb})_i)^2 + \sum_{i=k+1}^{m} ((\mathbf{Ub})_i)^2 \ . \qquad (3.21)$$

O qual é minimizado por algum $\mathbf{z}$ tal que $z_i = (\mathbf{U}^T\mathbf{b})_i/\sigma_i$,i=1,...,k. Dentre todos $\mathbf{z}$, $\|\mathbf{z}\|_2$ é minimizado para aquele no qual $z_i=0$, i=k+1,...,m, como $\mathbf{z}=\mathbf{D}^+\mathbf{U}^+\mathbf{b}$ é $\mathbf{x}=\mathbf{V}\mathbf{z}$,então, $\mathbf{x}=\mathbf{V}\mathbf{D}^+\mathbf{U}^T\mathbf{b}$, logo $\mathbf{x}=\mathbf{A}^+\mathbf{b}$.

Esta proposição, além de mostrar uma maneira para obter a pseudo-inversa, serve para demosntrar que através dela o problema (3.18) é solucionado o que até aqui só se tinha afirmado.

3.4 O PROBLEMA DE MÍNIMOS QUADRADOS COM PONDERAÇÃO

Nas ciências observacionais a ponderação é muito comun, uma vez que tais pesos podem ser obtidos através da própria precisão dos instrumentos de medidas utilizados, o que evidentemente traz maior confiança ao trabalho que está sendo realizado. Porém, como este trabalho não busca ligação direta com a teoria das observações, o problema é somente tratado quanto a solução.

3.4.1 Tratamento do problema

O problema de mínimos quadrados ponderado pode ser tratado por dois cominhos diferentes; um deles é ponderar todas as equações e proceder da maneira usual. O outro é minimizar o comprimento do resíduo segundo uma determinada norma (derivada a partir da norma-2).

Por ser mais usual o primeiro caso, será iniciado por ele.

Considere o sistema $\mathbf{Ax}=\mathbf{b}$ definido como:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad (3.22)$$

Cuja solução $\mathbf{x}=(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$, é a média aritmética dos elementos de $\mathbf{b}$, assim

$$x = \frac{(b_1+b_2+b_3)}{3}. \qquad (3.23)$$

Multiplicando ambos os lados de (3.22) por um escalar diferente de zero, w, o que corresponde que todas as equações tiveram a mesma ponderação, logo (3.22) ficará:

$$w\begin{bmatrix}1\\1\\1\end{bmatrix}\begin{bmatrix}x_w\end{bmatrix} = w\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix} \qquad (3.24)$$

donde com uma simplificação de w, voltaria-se novamente em (3.22) tendo como solução (3.23).

Isto serve para mostrar que se todas as equações forem ponderadas igualmente, a soluçào não sofrerá mudança.

Uma maneira geral de ver isso é escrever (3.24) na forma

$$\begin{bmatrix}w&0&0\\0&w&0\\0&0&w\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}x_w = \begin{bmatrix}w&0&0\\0&w&0\\0&0&w\end{bmatrix}\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix} \qquad \text{ou}$$

$$w\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}x_w = w\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix} \qquad (3.25)$$

que simplificando resulta

$$\begin{bmatrix}1\\1\\1\end{bmatrix}\begin{bmatrix}x_w\end{bmatrix} = \begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix} \qquad (3.26)$$

que é o sistema original.

Suponha agora que cada equação em (3.22) seja ponderada diferentemente, ou seja:

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \begin{bmatrix} x_w \end{bmatrix} = \begin{bmatrix} w_1 b_1 \\ w_2 b_2 \\ w_3 b_3 \end{bmatrix} \qquad (3.27)$$

cuja solução será

$$x_w = \frac{w_1^2 b_1 + w_2^2 b_2 + w_3^2 b_3}{w_1^2 + w_2^2 + w_3^2} \qquad (3.28)$$

que é a média ponderada, onde os pesos são os w_{is}^2. A solução x_w, neste caso seria maior influenciada pelo maior dos w_{is}, resultando um valor diferente para x_w daquele obtido em (3.23).

Uma forma mais comun de escrever (3.27) é

$$\begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_2 & 0 \\ 0 & 0 & w_3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} x_w = \begin{bmatrix} w_1 & 0 & 0 \\ 0 & w_2 & 0 \\ 0 & 0 & w_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad (3.29)$$

a qual não pode ser simplificada como no caso anterior.

Tanto (3.25) como (3.29) podem ser representadas numa forma compactada como:

$$\mathbf{WAx=Wb} \qquad (3.30)$$

Generalizando o problema usando as dimensões, (3.30) seria dimensionada, onde:

$\mathbf{W} \in \mathbf{R}^{mxm}$, $\mathbf{A} \in \mathbf{R}^{mxn}$, $\mathbf{x} \in \mathbf{R}^{n}$, $\mathbf{b} \in \mathbf{R}^{m}$.

Suponha ser A quadrada e não-singular, então (3.30) terá como solução

$$\mathbf{x}_w = (\mathbf{WA})^{-1}\mathbf{Wb} = \mathbf{A}^{-1}\mathbf{b}. \qquad (3.31)$$

Uma vez que W é sempre quadrada e inversivel, tal ponderação não mudaria em nada a solução, mesmo sendo diferentes os valores da diagonal de W; infelizmente esse não é o caso geral. No caso geral, A é retangular com mais linhas que colunas e não possui a inversa ordinária A^{-1}, mas sim a pseudo-inversa A^{+}, logo

$$\mathbf{x}_w = (\mathbf{WA})^{+}\,\mathbf{Wb} \qquad (3.32)$$

é a solução de (3.30).

Olhando para (3.32), porque não fazer $(\mathbf{WA})^{+} = \mathbf{A}^{+}\mathbf{W}^{+}$ e como W é inversivel a solução de (3.30) ficaria

$$\mathbf{x}_w = \mathbf{A}^{+}\mathbf{b} \qquad (3.33)$$

e a ponderação novamente não influenciaria na solução.

Acontece que isto não é verdade e pode ser visto quando se compara (3.23) e (3.28). O que de fato ocorre é que em geral $(\mathbf{WA})^{+} \neq \mathbf{A}^{+}\mathbf{W}^{+}$ como ficou demonstrado acima, pois $\mathbf{x}_w \neq \mathbf{x}$, quando $w_i \neq w_{i+1}$.

Para fechar esse caminho, pode-se generalizar ainda mais o problema. Considere W em (3.31) simétrica, onde os elementos fora da diagonal representam o grau de dependência entre as equações.

Já foi visto anteriormente que em qualquer sistema inconsistente $\mathbf{b} \notin \Re(\mathbf{A})$ e a solução é obtida resolvendo para x o sistema:

$$\mathbf{A}^{T}\mathbf{Ax} = \mathbf{A}^{T}\mathbf{b}. \qquad (3.34)$$

Para o caso do sistema (3.31), se $\mathbf{Wb} \notin \Re(\mathbf{WA})$, a solução será obtida pelo mesmo caminho e poderá ser tirada a partir do sistema

$$\mathbf{A}^{T}\mathbf{W}^{T}\mathbf{WA}\,\mathbf{x}_w = \mathbf{A}^{T}\mathbf{W}^{T}\mathbf{Wb}. \qquad (3.35)$$

Denominado o produto $\mathbf{W}^{T}\mathbf{W}$ por P, tem-se :

$$\mathbf{A}^T\mathbf{P}\mathbf{A}\,\mathbf{x}_w = \mathbf{A}^T\mathbf{P}\mathbf{b} \qquad (3.36)$$

a qual representa as equações normais de mínimos quadrados com ponderação.

— O segundo caminho para tratar do problema de mínimos quadrados com ponderação é usar uma nova definição para o comprimento do vetor residual.

Seja $\mathbf{W} \in \mathbf{R}^{m\times m}$ uma matriz não-singular e **v** um vetor m-dimensional. Denotando por **x** o produto de **W** por **v** tem-se

$\mathbf{x}=\mathbf{W}\mathbf{v}$.

O comprimento euclidiano de **x** é obtido fazendo

$$\|\mathbf{x}\|_2 = \sqrt{<\mathbf{x},\mathbf{x}>} = \sqrt{<\mathbf{W}\mathbf{v},\mathbf{W}\mathbf{v}>}\ . \qquad (3.37)$$

Considere agora uma nova norma para vetores denotada por $\|\cdot\|_w$ e definida como

$$\|\mathbf{v}\|_w = \sqrt{<\mathbf{W}\mathbf{v},\mathbf{W}\mathbf{v}>}\ . \qquad (3.38)$$

Da mesma forma, podemos definir um novo produto interno, ou seja:

$$<\mathbf{x},\mathbf{y}>_w = <\mathbf{W}\mathbf{x},\mathbf{W}\mathbf{y}> \qquad (3.39)$$

onde: **x** e **y** representam vetores quaisquer de iguais dimensões.

Escrevendo novamente (3.38) através do novo produto interno tem-se

$$\|\mathbf{v}\|_w = \sqrt{<\mathbf{v},\mathbf{v}>_w} \qquad (3.40)$$

sendo que a nova norma obedece as mesmas três condições das demais

i) $\|\mathbf{x}\|_w > 0\ ,\ \forall\ \mathbf{x} \neq \mathbf{0}$;

ii) $\|\alpha\mathbf{x}\|_w = |\alpha|\,\|\mathbf{x}\|_w$;

iii) $\|x+y\|_w \leq \|x\|_w + \|y\|_w$.

Desta maneira fica generalizado o problema de mínimos quadrados linear, podendo agora ser formulado como:

$$\min \|v\|_w^2 = \min_{x \in R^n} \|Ax-b\|_w^2 = \min < v,v >_w \ . \qquad (3.41)$$

Relembrando que a perpendicularidade de dois vetores acontece quando o produto interno deles é nulo, então **x** e **y** são perpendiculares quando $< x, y >_w = 0$.

Voltando ao problema original de mínimos quadrados. Lá tinha-se que o ótimo residual acontece quando **b** é projetado ortogonalmente no espaço coluna de A. Aqui será o mesmo só que o comprimento do resíduo deve ser medido por uma nova escala ($\|\cdot\|_w$) e a perpendicularidade por um novo produto interno ($< \cdot , \cdot >_w$).

Denotando por $p_w = Ax_w$ a projeção ortogonal de **b** em $\Re(A)$ medida por $< \cdot , \cdot >_w$, qualquer vetor $Ay \in \Re(A)$ deve ser perpendicular a $b - Ax_w$. Assim

$< Ay, b - Ax_w >_w = 0$ ou

$(WAy)^T(Wb - WAx_w) = 0$ e

$y^T(A^TW^TWb - A^TW^TWAx_w) = 0$ $\therefore$ $\forall$ $y \neq 0$, implica que

$A^TW^TWb = A^TW^TWAx_w$, denotando $W^TW = P$, resulta

$$A^TPAx_w = A^TPb \qquad (3.42)$$

que é a mesma equação normal obtida anteriormente.

Como observação final cabe dizer que a matriz **P** é conhecida nas ciências observacionais como matriz dos pesos, a qual corresponde a inversa da matriz de covariância das observações multiplicada pela variância da unidade de peso.

4. ANÁLISE DO CONDICIONAMENTO DO PROBLEMA DE MÍNIMOS QUADRADOS

O problema do mal-condicionamento é um problema crucial da matemática aplicada. Ocorre tanto para sistemas lineares quanto para sistemas não-linerares, principalmente devido a limitações da precisão, seja dos meios de formação do sistema, seja dos meios de obtenção da solução. Por exemplo: se os dados para o sistema forem colhidos através de observações, haverá limitação na precisão dos equipamentos e na acuidade visual do operador. Se não ocorrer tal problema neste estágio, ele poderá surgir na solução do sistema (formado a partir de um determinado modelo), uma vez que calculadoras e computadores tem limitação de precisão numérica.

Neste trabalho foi desprezado o fato dos dados serem obtidos através de observações. No entanto, fica ainda a parte concernente a limitação de precisão dos computadores.

Para analisar o condicionamento do problema de mínimos quadrados, é vantagem primeiro analisar o condicionamento de sitemas de equações lineares consistentes para identificar as diferenças com o problema de mínimos quadrados.

4.1 ANÁLISE DO CONDICIONAMENTO DE SISTEMAS DE EQUAÇÕES LINEARES CONSISTENTES.

A análise para o condicionamento de sistemas lineares será dividida

em duas partes; a primeira para quando A é não-singular e a segunda para quando A é retangular.

4.1.1 Sistema consistente com matriz inversível

Seja o sistema

$$\mathbf{Ax=b} \tag{4.1}$$

onde :

$\mathbf{A} \in \mathbf{R}^{nxn}$

$\mathbf{x,b} \in \mathbf{R}^{n}$

e posto(A)=n , então a solução pode ser escrita como:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \tag{4.2}$$

ou através do determinante e adjunta, como:

$$\mathbf{x} = \frac{\mathrm{Adj(A)}}{|\mathbf{A}|}\,\mathbf{b} \tag{4.3}$$

Em (4.3), por imposição de que posto(A)=n, implica que $|A|\neq 0$ ou que $\lambda_i \neq 0$, i=1,...,n .

Por definição, quando |A|=0, significa que a matriz A é singular e que x tem infinitas soluções ou que o sistema é incompatível. Geométricamente falando; Suponha primeiramente o sistema de duas equações a duas incógnitas, então as retas formadas no plano a partir das equações se sobreporiam para o caso indeterminado ou seriam paralelas para o caso incompatível.

Suponha agora ser o sistema formado por três equações e três incógnitas, uma equação linear em duas variaveis representa um plano no $\mathbf{R}^3$. Para o caso indeterminado, os três planos se interceptariam através de uma linha comum e qualquer ponto dessa linha seria solução para o sistema. Para o caso incompatível, qualquer dois planos se interceptariam ao longo de uma linha paralela a linha de interseção com qualquer outros dois planos e desta

forma não haveria nenhum ponto em comum para os três planos simultaneamente.

Suponha agora que |A| seja próximo de zero e imagine que isso significa ser as linhas ou colunas de **A** quase linearmente dependentes. Se fosse penssado nas duas retas, essas seriam quase paralelas havendo assim uma dificuldade muito grande em determinar o ponto de interseção das duas retas e qualquer erro de arredondamento na busca da solução do sistema traria consequências na determinação exata desta interseção.

Olhando por esse lado, analisar se um sistema merece ou não confiança seria muito simples, ou seja, se o determinante é próximo de zero a solução do sistema não é confiável.

Relembrando que o determinante de uma matriz quadrada triangular ou de uma matriz escalar é dado pelo produto dos elementos de sua diagonal principal (para este caso são seus autovalores). Faça, então o seguinte raciocínio; suponha uma matriz quadrada escalar onde os seus elementos da diagonal são: $a_{ij} = 10^{-10} \ \forall \ i=j$, o $\det(A)=10^{-10n}$ que é praticamente zero e o menor autovalor é $\lambda_1=10^{-10}$.

Através dessa simples análise é possível afirmar que este sistema não merece confiança? Se a resposta for sim, experimente multiplicar todas as equações por uma constante 10^{10}, ocorreria que a matriz dos coeficientes das incógnitas se tornaria identidade. Sabe-se da álgebra linear que a matriz identidade forma a base canônica do espaço que gera sendo portando perfeitamente condicionada (todas as linhas e colunas ortogonais umas com as outras). Desta forma uma simples adequação de escala tornaria um sistema mal-condicionado (a solução não merece confiança) em um sistema perfeitamente condicionado.

Em resumo; toda essa discução foi no sentido de mostrar que nem o determinante e nem o menor autovalor são bons indicadores do condicionamento de um sistema de equações lineares.

Uma definição um tanto grosseira para o condicionamento de um sistema pode ser enunciada como:

Um sistema de equações lineares é dito ser mal-condicionado se "pequenas" variações nos elementos do termo independente ou nos elementos da matriz dos coeficientes, gerar uma " grande" variação na solução comparada com a solução não pertubada. O sistema é dito ser bem-condicionado se essas variações não provocarem discrepâncias significativas entre as soluções.

4.1.1.1 Variação no termo independente.

Fazendo uma variação no termo independente de $\delta\mathbf{b}$, o sitema $\mathbf{Ax=b}$ pode ser escrito como :

$$\mathbf{A(x+\delta x)=(b+\delta b)} \qquad (4.4)$$

operando vem

$$\mathbf{Ax+A\delta x=b+\delta b}$$

subtraindo $\mathbf{b}$ de $\mathbf{Ax}$ resulta

$$\mathbf{A\delta x=\delta b} \quad \text{ou} \quad \delta\mathbf{x} = \mathbf{A}^{-1}\delta\mathbf{b}. \qquad (4.5)$$

Deseja-se saber qual é o valor máximo que o erro relativo da solução $\|\delta\mathbf{x}\|/\|\mathbf{x}\|$ pode alcançar. Para conhecer isso, $\|\delta\mathbf{x}\|$ deve ser tanto maior quanto possível.

Para $\|\delta\mathbf{x}\|$ ser máximo basta aplicar em (4.5) uma norma compatível (tendo em mente a definição de norma, eq.(2.8)), ou seja

$$\|\delta\mathbf{x}\| \leq \|\mathbf{A}^{-1}\|\|\delta\mathbf{b}\| \qquad (4.6)$$

e para $\|\mathbf{x}\|$ mínimo

$$\|\mathbf{b}\| \leq \|\mathbf{A}\|\|\mathbf{x}\| \qquad (4.7)$$

donde

$$\|x\| \geq \frac{\|b\|}{\|A\|} . \qquad (4.8)$$

De (4.6) e (4.8) resulta

$$\frac{\|\delta x\|}{\|x\|} \leq \|A^{-1}\|\|A\|\frac{\|\delta b\|}{\|b\|} . \qquad (4.9)$$

4.1.1.2 Variação na matriz dos coeficientes das incógnitas

Para uma variação em A de δA sem mudar o posto(A), $Ax=b$ torna-se

$$(A+\delta A)(x+\delta x)=b \qquad (4.10)$$

$$Ax+A\delta x+\delta A(x+\delta x)=b \qquad (4.11)$$

subtraindo Ax de b tem-se:

$$A\delta x+\delta A(x+\delta x)=0 \qquad (4.12)$$

onde

$\delta x = -A^{-1}\delta A(x+\delta x)$, assim para qualquer norma subordinada

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

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

multiplicando e dividindo o segundo membro por $\|A\|$ resulta

$$\frac{\|\delta x\|}{\|x+\delta x\|} \leq \|A^{-1}\|\|A\|\frac{\|\delta A\|}{\|A\|} . \qquad (4.13)$$

Chamando $\|A^{-1}\|\|A\| = C(A)$, então (4.9) e (4.13) tornam-se:

$$\frac{\|\delta x\|}{\|x\|} \leq C(A)\frac{\|\delta b\|}{\|b\|} \qquad (4.14)$$

$$\frac{\|\delta x\|}{\|x+\delta x\|} \leq C(A)\frac{\|\delta A\|}{\|A\|} . \qquad (4.15)$$

A quantidade "C(A)" é chamada *número de condição de* A e fornece a máxima ampliação que o erro relativo da solução pode sofrer baseado na mudança relativa do vetor b ou da matriz A.

O valor do número de condição C(A) vai depender da norma utilizada, mas pela equivalência na magnitude das normas, os valores entre os números de condições são comparáveis. Existe também certas denominações para o número de condição dependendo da norma utilizada, ver LUGNANI (1975).

Facilmente pode ser demosntrado que o número de condição C(A) é maior ou igual a unidade, pois para qualquer norma de matriz subordinada a matriz **I** tem norma igual a 1. De fato, como $\mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$ e para uma norma consistente $\|\mathbf{A}^{-1}\mathbf{A}\| \leq \|\mathbf{A}^{-1}\|\|\mathbf{A}\|$, logo $C(A) \geq 1$.

Em vista de que a matriz **I** é perfeitamente bem-condicionada e seu número de condição $C(I)=1$, pode-se por analogia deduzir que toda matriz bem-condicionada possui o número de condição próximo da unidade.

Observando (4.14) e (4.15), pode-se tirar duas importântes conclusões, quais são:

a) Se for conhecido o número de condição da matriz **A** pode-se saber qual será o limite superior que o erro relativo vai alcançar em função da variação relativa em **A** ou em **b**.

b) Pode-se num sistema estipular qual será o valor mínimo que o número de condição da matriz **A** pode ter, a partir do erro relativo e da variação relativa na matriz **A** ou vetor **b**.

Uma outra forma de identificar se a matriz é mal-condicionada é verificar se para **x** *diferentes mas de mesmo comprimento*, o comprimento do vetor **Ax** varia "dramaticamente". Se essas variações forem "dramaticamente diferentes" então a matriz **A** é mal-condicionada. Se a matriz dos coeficientes das incónitas de um sistema de equações lineares é mal-condicionada, então o sistema é dito ser mal-condicionado.

4.1.2 Caso geral do condicionamento de sistemas consistentes

Seja o sistema

$$\mathbf{Ax=b} \qquad (4.16)$$

onde

$\mathbf{A} \in \mathbf{R}^{m \times n}$, com posto(A)≤n ;

$\mathbf{x} \in \mathbf{R}^{n}$;

$\mathbf{b} \in \mathbf{R}^{m}$, com $\mathbf{b} \in \Re(\mathbf{A})$.

Este problema só difere do problema de mínimos quadrados por ser $\mathbf{b} \in \Re(\mathbf{A})$ e uma rápida olhada em seu condicionamento não poderia passar em branco, já que é o passo antecedente a análise do problema de mínimos quadrados.

Uma vez que posto(A)≤n, a solução de comprimento mínimo de (4.16) será:

$$\mathbf{x}^{*} = \mathbf{A}^{+}\mathbf{b} \qquad (4.17)$$

onde: $\mathbf{A}^{+}$ é a pseudo-inversa de A.

4.1.2.1 Variação do termo independente

Considere $\delta\mathbf{b}$ um vetor m-dimensional de pertubação do vetor b. Suponha que $\mathbf{b}+\delta\mathbf{b}$ mantenha a compatibilidade de (4.16), então o novo sistema será

$$\mathbf{A}(\mathbf{x}+\delta\mathbf{x})=\mathbf{b}+\delta\mathbf{b} \qquad (4.18)$$

subtraindo Ax de b resulta

$$\delta\mathbf{x} = \mathbf{A}^{+}\delta\mathbf{b} \qquad (4.19)$$

Considerando as desigualdades

$$\|\delta\mathbf{x}\| \leq \|\mathbf{A}^{+}\|\|\delta\mathbf{b}\| \qquad (4.20)$$

e

$$\|\mathbf{b}\| \leq \|\mathbf{A}\|\|\mathbf{x}^{*}\| \qquad (4.21)$$

com uma simples combinação resulta

$$\frac{\|\delta x\|}{\|x^*\|} \leq \|A^+\|\|A\|\frac{\|\delta b\|}{\|b\|} \qquad (4.22)$$

ou

$$\frac{\|\delta x\|}{\|x^*\|} \leq C(A)\frac{\|\delta b\|}{\|b\|}. \qquad (4.23)$$

Que é o mesmo resultado obtido na secção anterior só que agora a solução é a de comprimento mínimo (obviamente se A tem posto completo a solução é única) e o número de condição é obtido com auxilio da pseudo-inversa.

4.1.2.2 Pertubação na matriz A

Suponha ser δA uma matriz com mesma dimensão de A e que $A+\delta A$ tenha o mesmo espaço coluna de A. Então o novo sistema será

$$(A+\delta A)(x+\delta x)=b \qquad (4.24)$$

de maneira que

$$\delta A\delta x=b \qquad \Rightarrow \qquad \delta x = \delta A^+ b. \qquad (4.25)$$

Um procedimento análogo ao da secção anterior conduz a

$$\frac{\|\delta x\|}{\|x^* + \delta x\|} \leq C(A)\frac{\|\delta A\|}{\|A\|} \qquad (4.26)$$

sendo x^* a solução de comprimento mínimo e $C(A)=\|A^+\|\|A\|$.

As interpretações de tais condições são idênticas ao da secção (4.1.1), apenas com a observação de que um sistema pode ser mal-condicionado para quando posto(A)=k mas pode não ser quando posto(A)=k-1.

4.2 ANÁLISE DO CONDICIONAMENTO DO PROBLEMA DE MÍNIMOS QUADRADOS LINEAR

Foi visto acima várias maneiras para identificar se um sistema de equações lineares consistente é mal-condicionado ou não. Primeiramente através das variações na matriz dos coeficientes das incógnitas e no vetor de termos independentes, depois através da magnitude do número de condição da matriz dos coeficientes das incógnitas e por último através da aplicação de A à vetores diferentes de mesmo comprimento.

O problema de mínimos quadrados é um pouco mais complexo de ser analisado, pois tanto as pertubações quanto o número de condição deixam falhas (a não ser que seja formado o sistema de equações normais).

Quando se utiliza o sistema compatível formado pelas equações normais para resolver o problema, o condicionamento fica na ordem do quadrado com relações ao sistema original.

A seguir será visto detalhes de tais afirmações.

4.2.1 Sistemas de equações normais

Já se sabe que qualquer solução do problema de mínimos quadrados é também solução do sistema de equações normais

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b} \qquad (4.27)$$

sendo este sempre compatível (o vetor $\mathbf{A}^T\mathbf{b} \in \Re(\mathbf{A}^T\mathbf{A})$).

A análise para ver o condicionamento já foi feita no capítulo anterior, basta encontrar o número de condição $C(\mathbf{A}^T\mathbf{A})$, se este for elevado com relação a unidade é quase certo o mal-condicionamento do sistema de equações normais.

O que deve ser mostrado agora é a desvantagem em se resolver o problema através das equações normais quando comparado com os métodos que resolvem o problema diretamente do sistema original.

4.2.1.1. O condicionamento quadrado das equações normais

O número de condição de uma matriz A é definido como

$$C(A)=\|A^{-1}\|\|A\| \tag{4.28}$$

para uma norma consistente tem-se

$$\|A^TA\| \le \|A^T\|\|A\| \quad , \text{logo}$$

$$C(A^TA)=\|(A^TA)^{-1}\|\|A^TA\| \le \|A^{-1}\|\|(A^T)^{-1}\|\|A^T\|\|A\| \equiv C(A)^2 \tag{4.29}$$

$$C(A^TA) \le C(A)^2. \tag{4.30}$$

Exemplo:

Seja a matriz A definida como

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.00001 \\ 1 & 1.00001 \end{bmatrix}$$

usando três digitos significativos, os valores singulares de A serão

$\sigma_1 = 2,45$ e $\sigma_2 = 5,77.10^{-6}$, a norma euclidiana será

$\|A\|_2 = 2,45$ e o número de condição será

$$C(A)=\frac{\sigma_1}{\sigma_2} = 433102,25.$$

A matriz A^TA arredondada na quinta casa decimal será

$$A^T A = \begin{bmatrix} 3 & 3.00002 \\ 3.00002 & 3.00004 \end{bmatrix}$$

seus valres singulares (autovalores nesse caso) serão

$$\sigma_1 = 6{,}00 \quad \text{e} \quad \sigma_2 = 3{,}33.10^{-11}$$

sua norma euclidiana será

$\|A^T A\|_2 = 6{,}00$ e o número de condição será

$$C(A^T A) = \frac{\sigma_1}{\sigma_2} = 1{,}80.10^{11}.$$

No exemplo acima mostrou-se através da norma-2 que o número de condição de $A^T A$ é aproximadamente o quadrado do número de condição de **A**.

Se o sistema original fosse compatível, a análise através da secção(4.1) mostra que se este fosse moderadamente mal-condicionado a formação das equações normais o deixariam terrivelmente mal-condicionado.

Esse é o principal motivo de se evitar resolver o problema de mínimos quadrados através da equações normais quando não se tem certeza do condicionamento do sistema a-priori.

O mesmo exemplo pode ser usado para mostrar a perda de informação pela formação do produto $A^T A$.

Através dos valores singulares de A pode-se ver que A tem posto completo em se utilizando um computador que tenha unidade de arredondamento $\mathbf{u} = 10^{-8}$, mas a matriz $A^T A$ com essa precisão é sigular, para comprovar basta olhar para seu menor valor singular. Uma outra forma de checar é calcular seu determinante que será zero, pois o valor exato na posição (2,2) é 3,0000400002 ultrapassando a precisão **u**.

A seguir será feita uma tentativa em fixar um valor para o número de condição de A para resolver com segurança o problema de mínimos quadrados através das equações normais.

Supondo primeiramente ser a unidade de arreondamento do computador **u**, considerando o número de condição na norma-2, então

$$C(A)=\frac{\sigma_{max}}{\sigma_{min}} \quad \text{, donde quando} \quad \sigma_{min} \to 0 \quad \Rightarrow \quad C(A) \to \infty$$

e A → posto deficiente (o símbolo → para esse caso lê-se " tende para").

O número de condição de A^TA é

$$C(A^TA)=\frac{\sigma_{max}^2}{\sigma_{min}^2}=\frac{\lambda_{max}}{\lambda_{min}}. \qquad (4.31)$$

A singularidade de A^TA vai ser afetada somente pela pequenez de λ_{min} e para ser reconhecida a não-singularidade deve ocorrer que $\lambda_{min} > \mathbf{u}$. Sendo assim, deve-se ter

$$C(A^TA)<\frac{1}{\mathbf{u}} \qquad (4.32)$$

e para o problema original

$$C(A)<\frac{1}{\sqrt{\mathbf{u}}}. \qquad (4.33)$$

Para o exemplo anterior onde C(A)=433102,25 só poderá ser resolvido através das aquações normais se for utilizado um computador com unidade de arredondamento $\mathbf{u}=10^{-12}$, pois (4.33) ficaria 433102,25<1 000 000.

4.2.2 Análise do condicionamento do problema geral

Foi visto acima que quando o problema de mínimos quadrados é tratado através das equações normais a análise pode ser feita com base no número de condição de A^TA que será sempre da ordem do quadrado do número de

condição **A**, isto sugere usar os métodos baseados em decomposição ortogonal, como o QR e o SVD. Suponha ainda que mesmo usando tais métodos deseja-se conhecer o comportamento do problema, então, qual tipo de análise deve-se proceder para detectar se existe ou não o mal-condicionamento do problema? É exatamente esta análise que será apresentada nessa secção.

A parte introdutória para tal análise já foi anunciada quando caracterizou-se o ótimo residual, secção (3.2.2.1). Lá foi visto que dependendo da direção que o vetor de pertubação, $\delta \mathbf{b}$, tomar, ou ocorre variação só na solução ou só no resíduo. Aqui será dada continuidade nessa análise com mais detalhes e incluindo o tratamento do problema de comprimento mínimo.

O problema de mínimos quadrados formulado em (3.18) é:

Seja

$$\mathbf{Ax=b} \tag{4.34}$$

onde

$\mathbf{A} \in \mathbf{R}^{m \times n}$, com posto(A)$\leq$n e m>n;

$\mathbf{b} \in \mathbf{R}^{m}$, com $\mathbf{b} \notin \Re(\mathbf{A})$;

$\mathbf{x} \in \mathbf{R}^{n}$;

cujo objetivo é

$$\min_{x \in R^n} \{\|\mathbf{x}\|_2 : \|\mathbf{Ax}-\mathbf{b}\|_2^2 \leq \|\mathbf{A}\hat{\mathbf{x}}-\mathbf{b}\|_2^2 \ \forall\ \hat{\mathbf{x}} \in \mathbf{R}^n\}.$$

Vale relembrar a caracterização de ótimo para esse problema, que é

$\mathbf{x}^{*} = \mathbf{A}^{+}\mathbf{b}$, ou em termos dos componentes

$$\mathbf{Ax}_R = \mathbf{b}_R \ ; \quad \mathbf{x}_N = \mathbf{0} \ ; \quad \mathbf{V} = \mathbf{b}_n. \tag{4.35}$$

Nota: $\mathbf{x}_R$ e $\mathbf{x}_N$ são os componentes de $\mathbf{x}$ em $\Re(\mathbf{A}^T)$ e $\pi(\mathbf{A})$ respectivamente e $\mathbf{b}_R$ e $\mathbf{b}_N$ são os componetes de $\mathbf{b}$ em $\Re(\mathbf{A})$ e $\pi(\mathbf{A}^T)$ respectivamente.

O problema pode ser tratado de várias formas: Analisar de uma só vez as consequências das pertubações em **A** e em **b** com alteração do posto(A), ou sem alteração do posto(A), ou ainda com pertubação em **b** sem pertubação em **A** ou sem pertubação em **b** com pertubação em **A**, enfim como for conveniente.

Achou-se conveniente tratar cada caso separadamente e mantendo o posto(A) sem mudança. Quando o tratamento envolve muitos efeitos de uma só vez, fica complicado uma interpretação de maneira que é preferivel uma análise mais simples com interpretação.

4.2.2.1 Pertubação do termo independente

O problema (4.34) pertubado ficará

$$\mathbf{A}(\mathbf{x}+\delta\mathbf{x})=\mathbf{b}+\delta\mathbf{b}. \tag{4.36}$$

Considerando os componentes de **b** e $\delta\mathbf{b}$ em $\Re(\mathbf{A})$ e $\aleph(\mathbf{A}^T)$ e os componentes de **x** em $\Re(\mathbf{A}^T)$ e $\aleph(\mathbf{A})$, tem-se

$$(\mathbf{x}_R+\mathbf{x}_N)+(\delta\mathbf{x}_R+\delta\mathbf{x}_N)=\mathbf{A}^+(\mathbf{b}_R+\mathbf{b}_N)+\mathbf{A}^+(\delta\mathbf{b}_R+\delta\mathbf{b}_N) \tag{4.37}$$

da caracterização de ótimo obtém-se

$$\mathbf{x}_N=0\ ;\ \mathbf{x}_R=\mathbf{A}^+\mathbf{b}_R+\mathbf{A}^+\mathbf{b}_N \tag{4.38}$$

$$\delta\mathbf{x}_N=0\ ;\ \delta\mathbf{x}_R=\mathbf{A}^+\delta\mathbf{b}_R+\mathbf{A}^+\delta\mathbf{b}_N. \tag{4.39}$$

Por ser $\rho(\mathbf{A}^+)=\rho(\mathbf{A}^T)$, implica

$$\mathbf{x}_R=\mathbf{A}^+\mathbf{b}_R \quad \text{e} \quad \delta\mathbf{x}_R=\mathbf{A}^+\delta\mathbf{b}_R. \tag{4.40}$$

Usando uma norma compatível em (4.40) e (4.35) obtém-se

$$\|\delta\mathbf{x}_R\|\le\|\mathbf{A}^+\|\|\delta\mathbf{b}_R\| \quad \text{e} \tag{4.41}$$

$$\|\mathbf{b}_R\|\le\|\mathbf{A}^+\|\|\mathbf{x}_R\| \tag{4.42}$$

sendo $\|\mathbf{b}_R\|\neq 0$, ambas podem ser combinadas fornecendo

$$\frac{\|\delta x_R\|}{\|x_R\|} \leq C(A)\frac{\|\delta b_R\|}{\|b_R\|} \qquad (4.43)$$

onde:

$C(A)=\|A^+\|\|A\|$.

O efeito da pertubação δb no residual será

$$V+\delta V=A(x+\delta x)-(b+\delta b) \qquad (4.44)$$

como

$$Ax^* = b_R \quad \Rightarrow \quad V = b_N \quad \text{e}$$

$$\delta V=A\delta x-\delta b \quad \Rightarrow \quad A\delta x^* = \delta b_R \quad \text{e} \quad \delta V = \delta b_N. \qquad (4.45)$$

Analisando (4.43) e (4.44) pode-se ver que o número de condição da matriz A, causa efeito somente no erro relativo da solução, isso ainda se o vetor de pertubação tiver componete no espaço coluna de A. Se δb estiver inteiramente em $\pi(A^T)$, a solução não sofrerá variação. Tais variações serão assimiladas somente pelo residual (ver teste 04).

4.2.2.2 Pertubação na matriz A

Suponha ser δA uma matriz de mesma dimensão de A e que $A+\delta A$ mantém o posto de A. Assim (4.34) pode ser escrito como:

$$(A+\delta A)(x+\delta x)=b. \qquad (4.46)$$

Como agora o sistema é incompatível pois $b \notin \Re(A)$, deve-se permitir ocorrer variações no espaço coluna de A e no espaço nulo de A^T, de forma que suas dimensões permanecam as mesmas, uma vez que o posto é o mesmo.

Suponha ser

$$e_R = \frac{\|\delta A_R\|}{\|A\|} \quad \text{e} \quad e_N = \frac{\|\delta A_N\|}{\|A\|} \qquad (4.47)$$

a medida do efeito relativo da pertubação δA_R no espaço coluna de A e δA_N no espaço nulo de A^T.

Se o componente b_R de b for não-nulo demonstra-se que o erro relativo da solução satisfaz (GILL et alii 1991)

$$\frac{\|\delta x\|}{\|x^*\|} \leq 2C(A)e_R + 4C(A)^2 e_N \frac{\|b_N\|}{\|b_R\|} + O(e_N^2) \; . \qquad (4.48)$$

As quantidades δA_R e δA_N em (4.47) são obtidas fazendo

$$\delta A_R = G^T \delta A \quad \text{e} \quad \delta A_N = K^T \delta A \qquad (4.49)$$

onde:

G: é uma base ortonormal de $\Re(A)$;

K: é uma base ortonormal de $\pi(A^T)$,

adiante será visto como se obtém tais bases a partir da SVD.

Do residual pertubado $V + \delta V = b - (A + \delta A)(x^* + \delta x)$, resulta

$$\frac{\|\delta V\|}{\|b\|} \leq e_N + 2C(A)e_N + O(e_N^2). \qquad (4.50)$$

A quantidade $O(e_N^2)$ em (4.48) e (4.50) significa ter ordem de magnitude igual ou possivelmente inferior ao correspondente e_N^2.

Olhando para (4.48) e (4.50) pode-se concluir, que se a matriz de pertubação afetar somente o espaço coluna de A, ou b não tiver componente no espaço nulo de A^T, então o condicionamento quadrado não afeta o erro relativo da solução e o residual permanecerá inalterado. Em compensação se δA pertubar o espaço nulo de A^T, o erro relativo da solução passa a ser ampliado pelo número de condição ao quadrado conforme mostra (4.48) e o residual sofrerá alteração.

4.3 A DECOMPOSIÇÃO DE VALOR SINGULAR E BASES PARA OS SUBESPAÇOS FUNDAMENTAIS

Foi visto através das secções antecedentes a importância em se conhecer a forma e a base para os subespaços fundamentais. No capítulo 2 foi comentado alguns meios de se obter as bases para tais subespaços mas não se falou em bases ortonormais. A decomposição de valor singular providência automaticamente tais bases. Abaixo segue uma explanação de forma resumida.

Seja

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T \qquad (4.51)$$

onde

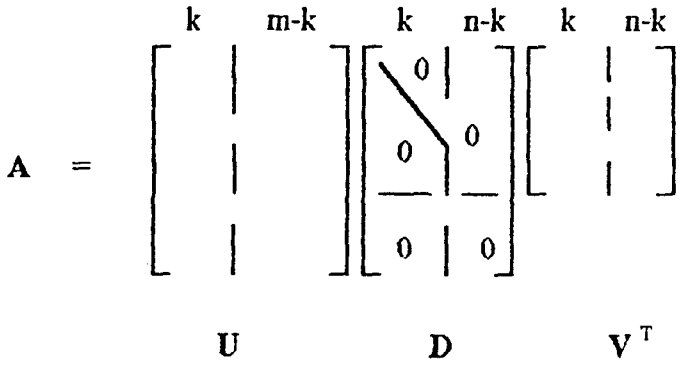
$\mathbf{U}_{(m,m)}$ ortonormal;

$\mathbf{V}_{(n,n)}$ ortonormal;

$\mathbf{D}_{(m,n)}$ diagonal com valores singulares $\sigma_i \ \forall \ i=j$.

observação: Se U ou V não estiverem normalizadas pelo algoritmo, pode-se facilmente normaliza-las dividindo cada elemento do vetor coluna por seu comprimento.

FIGURA 10 - BASES PARA OS SUBESPAÇOS FUNDAMENTAIS A PARTIR DA SVD



Se posto(A)=k<n, então (4.53) tem a forma da fig.10 e a interpretação é caracterizada pelas condições:

i) As colunas de $U_{(m,k)}$ fornecem uma base ortonormal para $\Re(A)$;

ii) As coluna de $U_{(m,m-k)}$ formam uma base ortonormal para $\pi(A^T)$;

iii) As colunas de $V_{(n,k)}$ formam uma base ortonormal para $\Re(A^T)$;

iv) As colunas de $V_{(n,n-k)}$ formam uma base ortonormal para $\pi(A)$.

4.4 ESTIMAÇÃO DO POSTO DE UMA MATRIZ

Quando se fala que uma matriz tem posto=n, isto significa que existe n linhas (ou colunas) linearmente independente na matriz o que consequentemente é um número inteiro. Supondo que exista uma "quase dependência" entre qualquer dessas linhas ou colunas, como ficaria a consideração desse "quase", já que o posto diz é ou não é, uma vez que é um número inteiro.

Foi visto também que para problemas com posto deficiente o que interessa é a solução de comprimento mínimo, ou melhor dizendo, se a escolha do posto estiver errada as soluções podem ser completamente diferentes daquelas esperadas.

4.4.1 Dificuldades na determinação do posto

A indagação acima só se constitui devido ao fato da precisão finita, pois caso contrário qualquer informação por menor que fosse seria de grande importância. O que acontece é que não se sabe a origem dessa pequena informação, sua causa pode ter ocorrido devido a um arredondamento, desprezo de certas quantidades, ou mesmo serem verdadeiras, ou ainda

podendo se originar da incerteza de observações se os dados assim fossem obtidos.

O simples exemplo abaixo pode elucidar melhor isso.

Considere

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2+\delta \end{bmatrix} \qquad (4.52)$$

se $\delta=0$ a terceira coluna é a soma das duas primeira e posto(A)=2, mas se δ for relativamente pequeno, como fica o posto(A)? se os elementos de A fossem coletados de observações e essas tivessem precisão menor que δ, obviamente poderia-se desprezar δ e considerar posto(A)=2. δ poderia ainda ser proveniente da soma das duas primeiras colunas calculadas com pouca precisão. Isto já traria dificuldade na decisão do posto de A, pois ficaria a dúvida se realmente aconteceu alguma imprecisão no cálculo ou se este estaria realmente correto.

4.4.2 A decomposição de valor singular na decisão do posto

A decomposição de valor singular é atualmente a técnica mais segura na determinação do posto de uma matriz, sendo utilizada para esse fim em importantes pacotes de análise numérica, como por exemplo: O MATLAB.

Já se sabe que o posto de uma matriz pode ser obtido pelo número de valores singulares não-nulos fornecidos pela SVD.

Suponha existir um desses valores singulares relativamente pequeno com relação a unidade e discrepando grandemente dos demais. Por exemplo: na matriz (4.52) dois valores singulares são de ordem 1 e o terceiro na ordem

de δ. Dependendo se δ é considerado "negligênciavel" ou não é como vai ser caracterizado o posto.

O termo negligênciavel pode ser relacionado com a precisão que serão efetuados os cálculos mas mesmo assim é inerente ao problema em questão. Por exemplo: suponha $\sigma_1 = 1$, $\sigma_2 = 10^{-1}$, $\sigma_3 = 10^{-2}$, . . ., $\sigma_n = 10^n$ os valores singulares de uma matriz mxn. Se fosse desprezado aqueles valores singulares abaixo da unidade de arredondamento poderia-se estar desprezando informações preciosas e a solução seria prejudicada.

Com a breve discussão acima já se pode ter uma idéia de quão ambiguo é este assunto, para o primeiro caso seria razoavel desprezar δ se este fosse relativamente pequeno, mas para o segundo é muito questionável desprezar qualquer informação.

Para essa análise elaboramos um programa no qual foi dividido em duas partes. Primeiro fornece a decomposição em si para ser analisada, depois permite calcular a solução com valores alterados se efetivamente forem alterados. Esta situação é mostrada nos testes 4 e 5 do capítulo 6.

5. MÉTODOS NUMÉRICOS PARA A SOLUÇÃO DO PROBLEMA DE MÍNIMOS QUADRADOS LINEAR

Neste capítulo será descrito 5 dos métodos que podem ser usados para a obtenção da solução do problema de mínimos quadrados linear. Os três primeiros resolvem o sistema consistente formado pelas equações normais de mínimos quadrados e os dois últimos se utilizam da propriedade especial das matrizes ortogonais (conservam o comprimento de vetores sob a norma-2) e resolvem diretamente o problema inconsistente (formado por equações redundantes) original.

5.1 ELIMINAÇÃO DE GAUSS-JORDAN

Este método foi escolhido para fazer parte do conjunto de métodos que resolvem o problema de mínimos quadrados linear pelos motivos: primeiro; ser simples, direto e tão estável quando aplicado o pivotamento completo quanto qualquer outro método direto que resolva um sistema de equações lineares consistente. Segundo; se utiliza da eliminação de Gauss que é a base da maioria dos métodos de solução de sistemas de equações lineares. Uma outra vantagem é que ao final do procedimento dos cálculos, o método fornece a inversa da matriz dos coeficientes das incógnitas que por certo pode ser aproveitada (matriz de covariância, ou resolver novamente o sistema com outros vetores de termos independentes).

O método de Gauss-Jordan pode solucionar um sistema de equações lineares com vários vetores de termos independentes simultaneamente. Isto é

uma vantagem, mas se for olhado em termos de armazenamento computacional, torna-se-ia uma desvantagem.

O desenvolvimento do método é feito sob forma geral e segue-se abaixo.

Considere os sitemas abaixos

$$\begin{aligned} Ax_1 &= b_1 \\ Ax_2 &= b_2 \\ &\vdots \\ Ax_n &= b_n \\ AY &= I \end{aligned} \qquad (5.1)$$

onde:

$A \in R^{nxn}$;

$x's \in R^n$;

$b's \in R^n$;

$Y \in R^{nxn}$;

Cujas soluções podem ser escritas como:

$$\begin{aligned} x_1 &= A^{-1}b_1 \\ x_2 &= A^{-1}b_2 \\ &\vdots \\ x_n &= A^{-1}b_n \\ Y &= A^{-1}I \end{aligned} \qquad (5.2)$$

Considere agora os sistemas (5.1) escritos na forma:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \cdot \left[\begin{array}{ccccccc} x_{11} & & x_{12} & & & x_{1m} & \\ \vdots & U & \vdots & U \cdots U & & \vdots & U \\ x_{n1} & & x_{n2} & & & x_{nm} & \end{array}\right. \left.\begin{array}{ccc} y_{11} & y_{12} & \cdots y_{1n} \\ \vdots & & \\ y_{n1} & y_{n2} & \cdots y_{nm} \end{array}\right] =$$

$$= \left[\begin{array}{ccccccc} b_{11} & & b_{12} & & & b_{1m} & \\ \vdots & U & \vdots & U \cdots & U & \vdots & U \\ b_{n1} & & b_{n2} & & & b_{nm} & \end{array}\right. \left.\begin{array}{ccc} 1 & 0 & \cdots 0 \\ \vdots & 1 & \ddots \vdots \\ 0 & 0 & \cdots 1 \end{array}\right] \qquad (5.3)$$

ou de maneira compacta, como:

$$[\mathbf{A}]\cdot[\mathbf{x}_1 U \mathbf{x}_2 U \cdots U \mathbf{x}_m U Y] = [\mathbf{b}_1 U \mathbf{b}_2 U \cdots U \mathbf{b}_m U I] \qquad (5.4)$$

onde U significa o agrupamento das colunas. É facil ver em (5.3) que x_{ij} , i=1,...,n significa o i-ésimo elemento do vetor solução **x** associado ao j-ésimo , j=1,...,m , vetor **b**. A matriz **y** é claramente vista ser A^{-1} quando tiver seus elementos determinados.

5.1.1 Operações elementares

As operações elementares aplicadas ao método de Gauss-Jordan são:

i) Permutando duas linhas de **A** e suas correspondentes linhas em **b's** e **I**, tanto as soluções **x's** como **Y** não sofrerão alterações (isto corresponde em apenas trocar a ordem das equações lineares no sistema).

ii) O conjunto solução não muda no processo de escalonamento quando multiplica-se qualquer linha de **A** por um escalar diferente de zero ou substitui-se qualquer linha de **A** por sua soma com outra linha previamente multiplicada por um escalar diferente de zero, desde de que as

correspondentes operações são também realizadas nas correspondente linhas dos **b's** e **I**.

iii) A permutação de qualquer duas colunas de A não muda o conjunto solução se simultaneamente for permutadas as correspondentes linhas dos **x's** e **y**. Em outras palavras, a solução sofrerá alteração de sua posição original. É interessante ao final voltar sua posição original e para isso devem ser guardadas as posições das permutações realizadas.

5.1.2 Solução do sistema por Gauss-Jordan

O método da eliminação de Gauss-Jordan usa uma ou mais das operações elementares acima para transformar a matriz A na matriz identidade (I). Quando isto é alcançado, I torna-se A^{-1} e os vetores **b's** tornam-se os vetores soluções **x's**.

Chama-se pivotamento a colocação através de permutações do maior elemento da linha ou coluna em sua correspondente posição na diagonal principal. Se não for realizado o pivotamento, somente a operação (ii) será usada.

Quando houver permutações somente nas de linhas A, de modo a levar o maior elemento da linha que não foi escalonada ainda para a diagonal principal, se está realizando um pivotamento parcial.

Quando as permutações são feitas nas linhas e colunas, então, diz-se estar realizando um pivotamento completo.

A eliminação de Gauss-Jordan , sem o pivotamento pode se tornar instável devido aos erros de arredondamento (com um elemento pivô grande no denominador há um maior número de digitos significativos na hora da divisão, ocasionando menor perda da precisão). Com o pivotamento completo este método torna-se bastante estavel, apesar de que com apenas o pivotamento parcial ele é quase tão bom quanto com o pivotamento completo.

5.1.3 Subrotina para a eliminação de Gauss-Jordan

Abaixo segue uma subrotina em linguagem "Fortran" da eliminação de Gauss-Jordan, tirada a partir de PRESS et alii (1986)

```
SUBROUTINA GAUSSJ(A,N,NP,B,M,MP)
      A solução do sistema de equações pela eliminação de Gauss-Jordan, equação (5.4). A é a
      matriz de entrada de N x N elementos, guardados em um array de dimensão física NP x
      NP. B é uma matriz de entrada contendo N x M vetores de termos independentes,
      guardados em um array de dimensão física de NP x MP. Na saída, A é substituida por sua
      matriz inversa e B é substituido pelos correspondentes conjuntos de vetores soluções.
PARÂMETROS (NMAX=50)
DIMENSÃO
A(NP,NP), B(NP,MP),IPIV(NMAX),INDXR(NMAX),INDXC(NMAX)
      Os vetores IPIV,INDXR e INDXC, são usados para marcar o pivotamento. NMAX deve
      ser igual ao valor pré-determinado N.
DO 11 J=1,N
      IPIV(J)=0
   11 CONTINUE
DO  22  I=1,N
    BIG=0
    DO 13  J=1,N
       IF (IPIV(J),NE.1) THEN
           DO 12 K=1,N
              IF (IPIV(K).EQ.0) THEN
                 IF (ABS(A(J,K)).GE.BIG) THEN
                   BIG=ABS(A(J,K))
                   IROW=J
                   ICOL=K
                 ENDIF
              ELSE IF (IPIV(K).GT.1) THEN
                 PAUSE ' Matriz singular '
              ENDIF
            12 CONTINUE
       ENDIF
       13 CONTINUE
     IPIV(ICOL)=IPIV(ICOL)+1
      Temos agora o elemento pivô, se necessário houve permutação de linhas para colocar o
      elemento pivô na diagonal. As colunas do i-ésimo elemento pivô não foram fisicamente
      permutadas, mas sim anotadas em INDX(I), enquanto INDXR(I) marca a linha na qual o
      elemento pivô foi originalmente colocado. Se INDXR(I)≠INDXC(I) existe uma permutação
      de colunas implicita. Com esta forma de guardar as mudanças, as soluções B's e A^-1
      terminariam na ordem correta.
     IF (IROW.NE.ICOL) THEN
       DO 14 L=1,N
            DUM=A(IROW,L)
            A(IROW,L)=A(ICOL,L)
            A(ICOL,L)=DUM
          14 CONTINUE
       DO 15 L=1,M
            DUM=B(IROW,L)
```

```
            B(IROW,L)=B(ICOL,L)
            B(ICOL,L)=DUM
          15 CONTINUE
      ENDIF
      INDXR(I)=IROW        Agora estamos pronto para dividir a linha do elemento pivô,
localizado em
      INDXC(I)=ICOL        IROW e ICOL.
      IF (A(ICOL,ICOL).EQ.0) PAUSE ' Matriz singular '
      PIVINV=1/A(ICOL,ICOL)
      A(ICOL,ICOL)=1
      DO 16 L=1,N
         A(ICOL,L)=A(ICOL,L)*PIVINV
        16 CONTINUE
      DO 17 L=1,M
         B(ICOL,L)=B(ICOL,L)*PIVINV
        17 CONTINUE
      DO 21 LL=1,N          Agora reduziremos as linhas exceto para o
         IF (LL.NE.ICOL) THEN      elemento pivô.
            DUM=A(LL,ICOL)
            A(LL,ICOL)=0
            DO 18 L=1,N
               A(LL,L)=A(LL,L)-A(ICOL,L)*DUM
                18 CONTINUE
            DO 19 L=1,M
               B(LL,L)=B(LL,L)-B(ICOL,L)*DUM
               19 CONTINUE
         ENDIF
         21 CONTINUE
      22 CONTINUE        Este é o fim do loop principal de redução de colunas
   DO 24 L=N,1,-1       ele ficaria para trocar a solução no caso de permutaçao
      IF (INDXR(L).NE.INDXC(L)) THEN            de colunas. Isto é feito
        DO 23 K=1,N         permutando as colunas aos pares na ordem
           DUM=A(K,INDXR(L))   inversa de que foram permutadas.
           A(K,INDXR(L))=A(K,INDXC(L))
           A(K,INDXC(L))=DUM
          23 CONTINUE
      ENDIF
     24 CONTINUE
   RETURN
   END
```

5.2 " SUBROUTINE VERSOL "

A " Subroutine Versol " é um algoritmo para inversões de matrizes quadradas com elementos reais, cujo autor é desconhecido. O motivo da introdução deste algoritmo neste trabalho é devido ao fato de ter sido e ainda está sendo usado em muitos trabalhos de ajustamento aplicado nas Ciências Geodésicas no Brasil. Isto nos leva a confrontá-lo com os demais no intuito de tirar algumas conclusões a respeito do mesmo.

Este método será desenvolvido segundo (MODRO 1981), mas antes será feita uma breve recordação da regra de Chió para reduzir a ordem do determinante.

5.2.1 Regra de Chió

i) Escolher um elemento igual a 1 (caso não exista, transformar por operações elementares).

ii) Suprimir a linha e a coluna que se cruzam no elemento de valor 1 considerado, obtendo-se o menor complementar do referido elemento.

iii) Subtrair de cada elemento do menor complementar obtido, o produto dos elementos que ficam nos pés das perpendiculares traçadas do elemento considerado às filas suprimidas.

iv) Multiplicar o determinante obtido no (iii) itém por $(-1)^{i+j}$, onde i e j designam a ordem da linha e da coluna as quais pertence o elemento 1.

Exemplo:

$$A = \begin{bmatrix} a_{11} & 1 & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$\det(A) = (-1)^{i+j} * \det \begin{bmatrix} a_{21} - a_{11}\, a_{22} & a_{23} - a_{22}\, a_{13} \\ a_{31} - a_{11}\, a_{32} & a_{33} - a_{13}\, a_{32} \end{bmatrix}$$

5.2.2 Desenvolvimento do método

O processo da " Subroutine Versol " será desenvolvido para uma matriz A de dimensão 1x1 e posteriormente com A de dimensão 2x2. Com isso espera-se ser suficiente a compreensão do algoritmo.

Considere $A_{(1,1)}$, ou seja

$$A = [a_{11}]. \tag{5.5}$$

Suponha agora uma nova matriz da forma:

$$A^{I} = \begin{bmatrix} a_{11} & 1 \\ -1 & 0 \end{bmatrix} \tag{5.6}$$

dividindo a primeira linha por a_{11} , tem-se:

$$A^{II} = \begin{bmatrix} 1 & 1/a_{11} \\ -1 & 0 \end{bmatrix} \tag{5.7}$$

aplicando a regra de Chió em (5.7) desenvolvido pela primeira linha e primeira coluna, resulta

$$\det(A^{II}) = (-1)^{1+1}.[\ 0-(-1).\frac{1}{a_{11}}\ ] = \frac{1}{a_{11}} \tag{5.8}$$

que é a inversa de $A=[a_{11}]$.

Considere agora $A_{(2,2)}$, ou seja

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{5.9}$$

e suponha uma nova matriz da forma

$$A^{I} = \begin{bmatrix} a_{11} & a_{12} & 1 \\ a_{21} & a_{22} & 0 \\ -1 & 0 & 0 \end{bmatrix} \qquad (5.10)$$

dividindo a primeira linha pelo elemento a_{11} , obtém-se:

$$A^{II} = \begin{bmatrix} 1 & a_{12}/a_{11} & 1/a_{11} \\ a_{21} & a_{22} & 0 \\ -1 & 0 & 0 \end{bmatrix} \qquad (5.11)$$

aplicando a regra de Chió em (5.11) desenvolvido pela primeira linha e primeira coluna, resulta

$$\det(A) = (-1)^{1+1} \begin{bmatrix} a_{22} - \frac{a_{21}\, a_{12}}{a_{11}} & 0 - a_{21}\frac{1}{a_{11}} \\ 0 - (-1)\frac{a_{12}}{a_{11}} & 0 - (-1)\frac{1}{a_{11}} \end{bmatrix} =$$

$$= \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = A^{III} \qquad (5.12)$$

Fazendo novamente

$$A^{IV} = \begin{bmatrix} c_{11} & c_{12} & 1 \\ c_{21} & c_{22} & 0 \\ -1 & 0 & 0 \end{bmatrix} \qquad (5.13)$$

dividindo a primeira linha por c_{11} , obtém-se

$$A^{V} = \begin{bmatrix} 1 & \frac{c_{12}}{c_{11}} & \frac{1}{c_{11}} \\ c_{21} & c_{22} & 0 \\ -1 & 0 & 0 \end{bmatrix} \qquad (5.14)$$

aplicando a regra de Chió desenvolvido pela primeira linha e primeira coluna, resulta

$$\det(A^{V}) = (-1)^{1+1} \cdot \begin{bmatrix} c_{22} & c_{21} - \frac{c_{21}\, c_{12}}{c_{11}} & 0 - c_{21}\frac{1}{c_{11}} \\ 0 & -(-1)\frac{c_{12}}{c_{11}} & 0 - (-1)\frac{1}{c_{11}} \end{bmatrix}$$

$$= \begin{bmatrix} c_{22} - \frac{c_{21}\, c_{12}}{c_{11}} & -\frac{c_{21}}{c_{11}} \\ \frac{c_{12}}{c_{11}} & \frac{1}{c_{11}} \end{bmatrix} \qquad (5.15)$$

a qual pode facilmente ser demonstrado ser $A^{-1}_{(2,2)}$. Procedendo-se desta forma, pode ser encontrada a matriz inversa de qualquer matriz não-singular $A_{(n,n)}$, aplicando n vezes o processo acima descrito.

5.2.3 Subrotina " versol "

Esta subrotina em linguagem "Fortran" foi tirada de MODRO (1981).

```
      SUBROUTINE VERSOL (A, B, I)
      IMPLICIT REAL (A-H, 0-3)
      DIMENSION A(I,I); B(I)
      IF (I.EQ.1) GO TO 10
      IM=I-1
      DO 5 K=1,I
      DO 2 J=1,IM
2     B(J)=A(1,J+1)/A(1,1)
      B(I)=1/A(1,1)
      DO 4  L=1 ,IM
      DO 3  J=1,IM
3     A(L,J)=A(L+1,J+1)-A(L+1,1)*B(J)
4     A(L,I)=-A(L+1,1)*B(I)
      DO 5 J=1,I
5     A(I,J)=B(J)
      RETURN
10    A(1,1)=1/A(1,1)
      RETURN
      END.
```

5.3 DECOMPOSIÇÃO DE CHOLESKY

Quando A é simétrica e definida positiva* , A pode ser fatorada na forma $A = LL^T$, onde L é uma matriz triangular inferior com elementos reais. A fatoração de $A = LL^T$ é chamada fatoração de Cholesky (ou decomposição de Cholesky). Existem várias maneiras de se obter a decomposição de Cholesky (por exemplo: usando a decomposição LU e QR). Aqui será descrito o método da raiz quadrada o qual aplica diretamente a definição, sendo que a matriz L é encontrada simplesmente escrevendo $(n^2+n)/2$ equações expressando cada elemento da parte triangular inferior de A em termos dos elementos de L. Assim:

---

* A definição de matriz definida positiva pode ser encontrada em (STRANG 1976), (GILL et alii 1991), (LUENBERGER 1973), (RAO 1978) e outros.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & & l_{nn} \end{bmatrix} \cdot \begin{bmatrix} l_{11} & l_{21} & \cdots & l_{n1} \\ & l_{22} & \cdots & l_{n2} \\ & & \ddots & \vdots \\ & & & l_{nn} \end{bmatrix}$$

onde:

$a_{11} = l_{11}^2$

$a_{21} = l_{11} . l_{21}$

.

.

.

$a_{n1} = l_{11} . l_{n1}$

$a_{22} = (l_{21})^2 + (l_{22})^2$

$a_{32} = l_{21} l_{31} + l_{22} l_{32}$

.

.

.

$a_{n2} = l_{21} l_{n1} + l_{22} l_{n2}$

$a_{33} = (l_{31})^2 + (l_{32})^2 + (l_{33})^2$

.

.

.

Resolvendo na ordem para $l_{11}, l_{22}, \ldots, l_{n1}$, $l_{22}, l_{32}, \ldots, l_{n2}$, $l_{33}, \ldots, l_{n3}$, ... , . L desta forma ficaria determinada.

Para conduzir a alguns detalhes importantes em torno dessa fatoração, serão citadas algumas propriedades das matrizes simétricas e definidas positivas.

Suponha ser A simétrica e definida positiva, então:

i) $a_{ii} > 0$, para todo i;

ii) Os maiores elementos estão na diagonal;

iii) Todos autovalores de A são reais e estritamente positivos.

iv) Qualquer matriz selecionada simétricamente dentro de A deve ser definida positiva.

v) Se a eliminação de Gauss sem permutação de linhas ou colunas for aplicada em A, a matriz equivalente será definida positiva em qualquer passo.

De posse dessas propriedades, já se pode analisar alguns passos dessa fatoração.

Como $l_{11} = \pm\sqrt{a_{11}}$, pela propriedade i , o elemento l_{11} fica bem determinado (por convenção toma-se o valor positivo da raiz).

Os demais elementos da primeira coluna de L podem ser calculados pela fórmula

$$l_{i1} = \frac{a_{i1}}{l_{11}}, \quad i = 2 , \ldots , n. \qquad (5.16)$$

O elemento l_{22} é fornecido pela expressão:

$$l_{22} = \sqrt{a_{22} - (l_{21})^2} \qquad (5.17)$$

Com isso para que l_{22} fique determinado $a_{22} - (l_{21})^2$ deve ser positivo, isso é assegurado devido ao fato desta quantidade ser exatamente o elemento da diagonal da matriz equivalente quando aplicado o primeiro passo da eliminação de Gauss em A (propriedade v).

Os demais elementos da segunda coluna de L são cálculados pela fórmula

$$l_{i2} = \frac{a_{i2} - l_{21} l_{i1}}{l_{22}} \quad , i = 3, \ldots, n \qquad (5.18)$$

Continuando desta forma coluna por coluna obtem-se a forma fatorada $A = LL^T$.

Para resolução do sistema de equações normais usando a fatoração de Cholesky procede-se:

$$Ax = b \quad \Leftrightarrow \quad LL^T x = b \qquad (5.19)$$

fazendo $Ly=b$, resolve-se para y e $L^T x = y$ resolve-se para x, que é a solução do sistema.

Devido A ser definida positiva não se faz necessário permutações para reter a estabilidade na formação da fatoração de Cholesky. Isto vem do fato de que existe uma limitação para os elementos de L. Para comprovar isso, considere que o elemento l_{kk} no k-ésimo passo da fatoração é obtido pela expressão

$$l_{kk}^2 = a_{kk} - l_{k1}^2 - l_{k2}^2 - \cdots - l_{k,k-1}^2 \qquad (5.20)$$

e assim

$$a_{kk} = l_{k1}^2 + l_{k2}^2 + \cdots + l_{k,k-1}^2 + l_{kk}^2 \; . \qquad (5.21)$$

Como todos os termos da soma do lado direito são positivos, implica que

$$\sqrt{a_{kk}} \geq |l_{ki}| \; . \qquad (5.22)$$

Desse modo todos os elementos da k-ésima coluna de L ficam limitados pela magnitude da raiz quadrada do k-ésimo elemento da diagonal de A.

Mesmo quando A é mal-condicionada, os elementos de L se comportam bem no sentido da satisfazer o limite a-priori (5.22).

Apesar desse fato favorável, casos extremos podem acontecer. O elemento l_{kk} pode não ser definido por causa de erros de arrendondamento, ou seja, em (5.20) o segundo membro torna-se negativo.

Mesmo garantida pela teoria, a estabilidade do método de Cholesky não é mantida na prática (ver testes). Para que exista a estabilidade neste método, A além de ser definida positiva deve ser bem-condicionada.

5.3.1 Subrotina " método de Cholesky "

Abaixo um "procedure" em "Algol" tirado a apartir de GASTINEL (1971) para calcular a decomposisão de Cholesky e solução do sistema de equações lineares.

```
procedure CHOLESKY(A) LADO DIREITO:(B) ORDEM:(N) RESULTADO: (X) FALHA E SAÍ:
(IMPOSSÍVEL); VALORES A,B;
array real A,B,X; integer N; label IMPOSSÍVEL;
comentário RESOLVE AX=B (ONDE A É SIMÉTRICA E DEFINIDA POSITIVA) PELA
DECOMPOSIÇÃO A=R.R^T (ONDE R É TRIANGULAR INFERIOR) E ENTÃO
SOLUCIONANDO RY=B E R^TX=Y. A MATRIZ R É DEPOSITADA NA PARTE TRIANGULAR
INFERIOR DE A;
begin integer I,J,K ; real S,TX; if A[1,1]≤0 then go to
IMPOSSÍVEL ; A[1,1]:=SQRT(A[1,1]);
      for I:=2 step 1 until N do A[I,1]:=A[I,1]/A[1,1];
      for j:=2 step 1 until N do
                  begin S:=0; for I:=1 step 1 until j-1 do
                        S:=S+A[J,I]*A[J,I];
                        S:=A[J,J]-S;
                  if S≤ 0 then go to IMPOSSÍVEL;
                  A[J,J]:=SQRT(S);
            for I:=J+1 step 1 until N do
                  begin S:=0; for K=1 step 1 until J-1 do
                        S:=S+A[I,K]*A[J,K];
                        S:=A[I,J]-S;
                        A[I,J]:=S/A[J,J];
                  end;
            end;
                  for I:=1 step 1 until N do
                        begin TX:=0; for J:=1 step 1 until I-1 do
                              TX:=TX+A[I,J]*X[J];
                              X[I]:=(B[I]-TX)/A[I,I]; B[I]:=X[I];
                        end;
                  for I:=N step -1 until 1 do
                        begin TX:=0; for J:=N step -1 until I +1 do
                              TX:=TX+A[J,I]*X[J];
                              X[I]:=(B[I]-TX)/A[I,I];
                        end;
end;
```

5.4 O MÉTODO DA DECOMPOSIÇÃO QR

O método da decomposição QR consiste em transformar a matriz dos coeficientes das incógnitas em uma triangular superior através de transformações ortogonais, este método vem sendo últimamente bastante utilizado na solução do problema de mínimos quadrados. Não é tão potente quanto o SVD mas em estabilidade são praticamente equivalentes. As transformações ortogonais podem ser obtidas ou pelo processo de Gram-Schmidt ou pelas transformações de Householder, sendo a última mais recomendada por questões de economia computacional e estabilidade.

5.4.1 Transformação de Householder

A transformação de Householder tem utilidade para muitos propósitos da álgebra linear. Consiste basicamente em transformar um vetor em outro de mesmo comprimento euclidiano. A formalização pode ser obtida da seguinte proposição.

Proposição 1: Dado um vetor não-nulo **a** m-dimensional, existe uma matriz simétrica e ortogonal Q tal que

$$\mathbf{Qa} = -\sigma \mathbf{e}_1 \qquad (5.23)$$

onde:

$\mathbf{e}_1 = (1\ 0 \cdots 0)^T$ e $\sigma = \|\mathbf{a}\|$.

Primeira parte: prova de (5.23)

defina Q como

$$Q = I_m - 2\frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|_2^2} \qquad (5.24)$$

e tomando

$$\mathbf{v} = \mathbf{a} + \sigma \mathbf{e}_1 \qquad (5.25)$$

então

$$\mathbf{Qa} = (\mathbf{I}_m - 2\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}})\mathbf{a} = \mathbf{a} - \mathbf{v}\frac{2\mathbf{v}^T\mathbf{a}}{\mathbf{v}^T\mathbf{v}} =$$

$$= \mathbf{a} - \mathbf{v}\frac{2(\mathbf{a}^T\mathbf{a} + \sigma\mathbf{e}_1^T\mathbf{a})}{\mathbf{a}^T\mathbf{a} + \mathbf{a}^T\sigma\mathbf{e}_1 + \sigma\mathbf{e}_1^T\mathbf{a} + \sigma^2} = \mathbf{a} - \mathbf{v}\frac{(\mathbf{a}^T\mathbf{a} + \mathbf{a}^T\mathbf{a} + 2\sigma\mathbf{e}_1^T\mathbf{a})}{\mathbf{a}^T\mathbf{a} + 2\sigma\mathbf{e}_1^T\mathbf{a} + \sigma^2}$$

como

$$\sigma^2 = \mathbf{a}^T\mathbf{a}$$

então

$$\mathbf{Qa} = \mathbf{a} - \mathbf{v} = \mathbf{a} - \mathbf{a} - \sigma\mathbf{e}_1 = -\sigma\mathbf{e}_1. \qquad (5.26)$$

observação: O valor de Qa pode também ser representado por

$$\mathbf{Qa} = -\sigma\|\mathbf{a}\|\mathbf{e}_1 \text{ , com } \quad \mathbf{v} = \mathbf{a} + \sigma\|\mathbf{a}\|\mathbf{e}_1 \qquad (5.27)$$

onde

$$\sigma = \begin{cases} +1 & \text{se } a_1 \geq 0 \\ -1 & \text{se } a_1 < 0 \end{cases}$$

sendo a_1 o elemento de **a** correspondente ao elemento 1 de $\mathbf{e}_1$, desta forma o componte de **v** correspondente será sempre acrescido conservando o sinal de a_1 sem o risco de tornar-se nulo.

Segunda parte: prova da ortogonalidade e simetria de Q

Considere o vetor normalizado **u** como $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|$, então Q pode ser definida como

$$\mathbf{Q} = \mathbf{I}_m - 2\mathbf{u}\mathbf{u}^T . \qquad (5.28)$$

Por definição de simetria $\mathbf{Q} = \mathbf{Q}^T$, então

$$\mathbf{I}_m - 2\mathbf{u}\mathbf{u}^T = (\mathbf{I}_m - 2\mathbf{u}\mathbf{u}^T) = (\mathbf{I}_m - \mathbf{u}\mathbf{u}^T 2) = \mathbf{I}_m - 2\mathbf{u}\mathbf{u}^T. \qquad (5.29)$$

Por definição de ortogonalidade $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$, então

$$(\mathbf{I}_m - 2\mathbf{u}\mathbf{u}^T)(\mathbf{I}_m - 2\mathbf{u}\mathbf{u}^T) = \mathbf{I}$$

$$\mathbf{I}_m^2 - 2\mathbf{u}\mathbf{u}^T - 2\mathbf{u}\mathbf{u}^T + 4\mathbf{u}\mathbf{u}^T = \mathbf{I} \quad \therefore \quad \mathbf{I} = \mathbf{I}. \qquad (5.30)$$

A matriz Q definida na proposição acima é conhecida como transformação de Householder por ter sido utilizada por A.S. Householder

pela primeira vez em problemas de autovalores. O vetor particular v é chamado vetor de Householder.

5.4.2 A decomposição QR

Proposição 1: Seja $A \in R^{m \times n}$ como posto(A)=n. Então existe uma matriz ortogonal $Q_{(m,m)}$ tal que $Q^T A = R$, sendo $R_{(m,n)}$ com zeros abaixo da diagonal principal.

Prova: Permita $Q_1 = I_m - 2\frac{v_1 v_1^T}{\|v\|_2^2}$, onde $v_1 = a_1 + \sigma\|a_1\|e_1$, sendo σ definido como em (5.27), a_1 e e_1 a primeira coluna da matriz A e matriz I respectivamente, então

$$Q_1 v_1 = \begin{bmatrix} a_{11} + ||a_1|| = r_{11} \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix} \qquad (5.31)$$

A aplicação de Q_1 sobre A resultará $Q_1 A = A^{(2)}$, da forma

$$A^{(2)} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & a^{(2)}_{22} & \cdots & a^{(2)}_{2n} \\ . & . & \cdots & \\ . & . & \cdots & \\ . & . & \cdots & \\ 0 & a^{(2)}_{m2} & \cdots & a^{(2)}_{mn} \end{bmatrix} \qquad (5.32)$$

Formando uma matriz ortogonal $P_{2_{(m-1,m-1)}}$ com a segunda coluna de $A^{(2)}$ a partir do elemento pivô, forma-se assim Q_2 da forma

$$Q_2 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & P_2 & \\ 0 & & & \end{bmatrix} \qquad (5.33)$$

aplicando $Q_2A^{(2)} = A^{(3)}$, da forma

$$A^{(3)} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ 0 & r_{22} & r_{23} & \cdots & r_{2n} \\ 0 & 0 & a^{(3)}_{33} & \cdots & a^{(3)}_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & a^{(3)}_{m3} & \cdots & a^{(3)}_{mn} \end{bmatrix} \qquad (5.34)$$

Procedendo desta maneira com no máximo n aplicaçoes da transformação de Householder pode-se transformar A em R, isto é:

$$Q_nQ_{n-1}\cdots Q_1A = R \qquad (5.35)$$

chamando $Q^T = Q_nQ_{n-1}\cdots Q_1$, tem-se

$$A=QR\ . \qquad (5.36)$$

5.4.3 O método QR aplicado ao problema de mínimos quadrados linear

Proposição 1: Seja $m \geq n > 0$, $b \in R^m$, $A \in R^{m \times n}$ com posto completo. Então, existe uma decomposição $A=QR$, onde $Q_{(m,m)}$ matriz ortogonal, $R \in R^{m \times n}$ com zeros abaixo da diagonal principal. Seja ainda R_n a matriz formada pelas n primeiras linhas de R de modo que R_n é não-singular e triangular superior.

Assim a solução única de $\mathbf{x}\in\mathbf{R}^n$ é dada por $\mathbf{x}=\mathbf{R}_n^{-1}(\mathbf{Q}^T\mathbf{b})_n$, onde $(\mathbf{Q}^T\mathbf{b})_n=[(\mathbf{Q}^T\mathbf{b})_1,\ldots,(\mathbf{Q}^T\mathbf{b})_n]$ são os n primeiros componentes de $\mathbf{Q}^T\mathbf{b}$; o vetor residual $\mathbf{V}=\|\mathbf{Ax}-\mathbf{b}\|_2$ será $\mathbf{V}=\sum_{i=n+1}^{m}[(\mathbf{Q}^T\mathbf{b})_i]^2$.

Prova: A existência da decomposição QR segue a partir da transformação de Householder e a não-singularidade de $\mathbf{R}_n$ a partir da independência das linhas de A. Da propriedade da norma euclidiana $\|\mathbf{QV}\|_2=\|\mathbf{V}\|_2$ vem

$$\|\mathbf{Ax}-\mathbf{b}\|_2=\|\mathbf{QRx}-\mathbf{b}\|_2=\|\mathbf{Rx}-\mathbf{Q}^T\mathbf{b}\|_2 \qquad (5.37)$$

e o problema pode ser escrito como

$$\min_{\mathbf{x}\in\mathbf{R}^n}\|\mathbf{Rx}-\mathbf{Q}^T\mathbf{b}\|_2^2 . \qquad (5.38)$$

Denotando $(\mathbf{Q}^T\mathbf{b})_L=[(\mathbf{Q}^T\mathbf{b})_{n+1},\ldots,(\mathbf{Q}^T\mathbf{b})_m]$, os componentes de $\mathbf{Q}^T\mathbf{b}$ a partir de n+1 até m, tem-se

$$\|\mathbf{Rx}-\mathbf{Q}^T\mathbf{b}\|_2^2=\|\mathbf{R}_n\mathbf{x}-(\mathbf{Q}^T\mathbf{b})_n\|_2^2+\|(\mathbf{Q}^T\mathbf{b})_L\|_2^2 \qquad (5.39)$$

o qual é minimizado quando

$$\mathbf{R}_n\mathbf{x}-(\mathbf{Q}^T\mathbf{b})_n=\mathbf{0} \quad\Leftrightarrow\quad \mathbf{R}_n\mathbf{x}=(\mathbf{Q}^T\mathbf{b})_n \qquad (5.40)$$

e o ótimo residual terá comprimento

$$\|\mathbf{V}\|_2^2=\|(\mathbf{Q}^T\mathbf{b})_L\|_2^2 \qquad (5.41)$$

concluindo-se assim a prova da proposição.

5.4.4 Subrotina do método QR

A subrotina abaixo foi tirada a apartir de (DENNIS e SCHNABEL 1983) é para matrizes quadradas e não-singulares, mas pode facilmente ser implementada para o problema de mínimos quadrados como descrito acima.

DECOMPOSIÇÃO-QR (QRDECOMP)

Propósito: calcula a decomposição QR de uma matriz quadrada M usando o algoritmo de Stewart[1973] no final do algoritmo a decomposição é armazenada em M, M1 e M2 como descrito abaixo.

Parâmetros de entrada: n ∈ Z (conjunto dos números inteiros)

Parâmetros de entrada e saída: M $\in \mathbf{R}^{nxn}$(na saída, Q e R são embutidas em M, M1 e M2 como descrito abaixo)

Parâmetros de saída: M1 $\in \mathbf{R}^n$, M2 $\in \mathbf{R}^n$, sing $\in$ boolean

Considerações sobre o armazenamento

1) Não é necessário nenhum vetor ou matriz adicional

2) Na saída Q e R são guardadas como descrito por Stewart: R esta contida na parte triangular superior de M exceto a diagonal principal que esta em M2, e $Q^T = Q_{n-1},\ldots,Q_1$ onde $Q_j = I-(u_j u_j^T/\pi_j)$, $u_j[i]=0$, i=1,...,j-1, $u_j[i]= M[i,j]$, i=j,...,n , $\pi_j = M1[j]$.

Descrição:

A decomposição QR de M é realizada usando a transformação de Householder pelo algoritmo de Stewart[1973]. A decomposição retorna sing=true se a singularidade de M é detectada, sing=false caso contrário. A decomposição se completa mesmo com a singularidade detectada.

Algoritmo:

1 sing ← false (sing torna-se true se a singularidade é detectada)

2 for k=1 ro n-1 do

2.1 $\eta \leftarrow \max_{k\le i\le n}\{|M[i,k]|\}$

2.2 If η=0

2.2.T Then (*matriz singular*)

2.2.T.1 M1[k] ← 0

2.2.T.2 M2[K] ← 0

2.2.T.3 sing ← true

2.2.E Else (*Forma Q_k e premultiplica M por ela *)

2.2.E.1 For i=k to n do M[i,k] ←M[i,k]/η

2.2.E.2 $\sigma \leftarrow$ sign (M[k,k])*$(\sum_{i=k}^{n} M[i,k]^2)^{1/2}$

2.2.E.3 M[k,k] ← M[k,k]+σ

2.2.E.4 M1[k] ← σ * M[k,k]

2.2.E.5 M2[k] ← - η * σ

2.2.E.6 For j=k+1 to n do

2.2.E.6.1 $\tau \leftarrow (\sum_{i=k}^{n} M[i,k]*M[i,j]/M1[k])$

2.2.E.6.2 For i=k to n do

M[i,j] ← M[i,j]-τ*M[i,k]

3 M2[n] ← M[n,n]

(* Retorno do algoritmo QRDECOMP *)

(* END do algoritmo *)

Subrotina RSOLVE

Propósito: Resolve Rx=b para b, onde R é triangular superior armazenada como descrito na subrotina QRDECOMP

Parâmetros de entrada: n ∈ Z, M ∈ $\mathbf{R}^{nxn}$, M2 ∈ $\mathbf{R}^{n}$ (M2 contém a diagonal de R, o restante de R esta contida na parte superior de M)

Parâmetros de entrada e saída: b ∈ $\mathbf{R}^{n}$ (na saída , b é sobreposto pela solução x)

Parâmetros de saída: nenhum

Consideração de armazenamento:

1) Não é necessário nenhum vetor ou matriz adicionais

Algoritmo

1. b[n] ← b[n]/M2[n]

2. For i=n-1 downto 1 do

$$b[i] \leftarrow \frac{b[i]-\sum_{j=i+1}^{n} M[i,j]*b[j]}{M2[i]}$$

(* Retorno do algoritmo RSOLVE *)

(* END do algoritmo RSOLVE *)

Subrotina QRSOLVE

Propósito: Resolve (QR)x=b para x, onde a matriz ortogonal Q e a triangular superior R são armazenadas como descrito na subrotina QRDECOMP.

Parâmetros de entrada: n ∈ Z, M ∈ $\mathbf{R}^{nxn}$, M1 ∈ $\mathbf{R}^{n}$, M2 ∈ $\mathbf{R}^{n}$ (Q e R são embutidas em M, M1 e M2 como descrito na subrotina QRDECOMP)

Parâmetros de saída: nenhum

Considerações sobre o armazenamento:

1) Nenhuma matriz ou vetor adicionais são necessários

Descrição:

1) Multiplica b por Q^T, Q^T é armazenado como $Q^T = Q_{n-1},\ldots,Q_1$, onde cada Q_j é uma transformação de Householder implicita como descrito na subrotina QRDECOMP. Assim este passo consiste de pré-multiplicar b por Q_j , j=1,...,n-1.

Algoritmo

(* b ← Q^Tb*)

1. For j=1 to n-1 do

(* b ← Q_jb*)

1.1 $\tau \leftarrow \left(\sum_{i=j}^{n} M[i,j]*b[j]/M1[j]\right)$

1.2 For i=j to n do

b[i] ← b[i]-τ*M[i,j]

```
      (* b ← R^-1 b *)
2.    CALL RSOLVE(n,M,M2,b)      (* alg. RSOLVE*)
(* Retorno do QRSOLVE *)
(* END, QRSOLVE *)
```

5.5 DECOMPOSIÇÃO DE VALOR SINGULAR

Nesta secção será descrito os procedimentos de como obter a decomposição de valor singular e uma subrotina em linguagem Fortran para calculá-la. Sua aplicação ao problema de mínimos quadrados linear já foi mostrada pelas proposições da secção (3.3.4).

Antes de estudar este fantástico algoritmo é necessário primeiro dar uma olhada no algoritmo QR para o problema de autovalores e por causa da estrutura especial das matrizes que será utilizada, também será usada a transformação de Givens* .

5.5.1 Transformação de Givens

A transformação de Givens é uma matriz ortogonal que é utilizada para quando se deseja tornar nulo apenas um dos elementos do vetor. As modificações no vetor ocorrerão somente em dois de seus componentes.

A transformação de Givens pode ser definida pela seguinte proposição.

Proposição 1: Seja o vetor $v=(v_1\ v_2\ v_3\ \ldots\ v_m)^T$ com pelo menos dois componentes não nulos, existe uma matriz $G_{(m,m)}$ ortogonal

---

* Transformação de Givens : devido a Givens 1954.

$$G = \begin{bmatrix} c & s & 0 & \cdots & 0 \\ -s & c & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \qquad (5.42)$$

tal que

$$Gv = (\sqrt{v_1^2+v_2^2} \;\; 0 \;\; v_3 \ldots v_m)^T. \qquad (5.43)$$

Prova

primeira parte: fazendo em G

$$c = \frac{v_1}{\sqrt{v_1^2+v_2^2}}, \quad s = \frac{v_2}{\sqrt{v_1^2+v_2^2}} \qquad (5.44)$$

e multiplicando G por v (como definido na proposição) resultará em (5.43).

segunda parte: Ortogonalidade de G.

Pela definição de ortogonalidade $GG^T = I$ e tomando só a parte superior esquerda de G já que o restante já esta sob a forma de I, tem-se

$$GG^T = \begin{bmatrix} c^2 + s^2 & 0 \\ 0 & s^2 + c^2 \end{bmatrix} \qquad (5.45)$$

como $c^2+s^2 = \frac{v_1^2}{v_1^2+v_2^2} + \frac{v_2^2}{v_1^2+v_2^2} = 1$ implica

$$GG^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad (5.46)$$

Observação: G pode ser escolhida para ser simétrica e ortogonal, para isso basta em (5.42) colocar o sinal negativo em um dos c's deixando ambos os s's com sinais positivos.

5.5.2 O algoritmo QR

Foi visto na secção (5.4) como decompor uma matriz $A_{(m,n)}$ com $m \geq n$, posto completo e resolver desta forma o problema de mínimos quadrados. Aqui como uma preparação para a decomposição de valor singular, será mostrado que uma sequência de decomposições do tipo QR com $A \in R^{nxn}$ leva a obtenção do autovalores de A.

5.5.2.1 O algoritmo QR

Suponha $A_0 = A$ ser decomposta em

$$A_0 = Q_0 R_0 \tag{5.47}$$

e uma nova matriz semelhante a A_0, definida invertendo a ordem do segundo membro (5.47),ou seja

$$A_1 = R_0 Q_0 \tag{5.48}$$

essa nova matriz conserva os autovalores de A_0 .

Para provar essa afirmação basta fazer

$$Q_0^T A_0 Q_0 = Q_0^T Q_0 R_0 Q_0 = R_0 Q_0 = A_1 \tag{5.49}$$

lembrando que o produto no primeiro membro da igualdade não altera os autovalores.

Montando uma sequência e denotando k e k+1 o algoritmo fica

$$A_k = Q_k R_k \tag{5.50}$$

$$A_{k+1} = R_k Q_k. \tag{5.51}$$

Procedendo desta maneira, o algoritmo sob circunstâncias gerais convergirá, ou seja, A_k se aproxima da forma triangular superior e os elementos de sua diagonal se aproximam de seus autovalores.

5.5.2.2 O algoritmo QR - modificado

O puro e simples algoritmo QR para encontrar os autovalores é bom e terá um acréscimo adicional na velocidade de convergência se for dada a seguinte implementação.

$$A_0 = A \tag{5.52}$$

$$(A_k - \sigma_k I_n) = Q_k R_k \tag{5.53}$$

$$A_{k+1} = R_k Q_k + \sigma_k I_n \tag{5.54}$$

onde k=0,1,2,3,... e σ_k um escalar. A matriz A_{k+1} conserva os autovalores de A_k pelo mesmo motivo de (5.48), para verificar faça

$$Q_k^T A_k Q_k = Q_k^T (Q_k R_k + \sigma_k I_n) Q_k = R_k Q_k + \sigma_k I_n = A_{k+1} \tag{5.55}$$

a quantidade σ_k é escolhido para ser um autovalor aproximado de A_k, na prática o que se faz é tomar o elemento no canto inferior direito da matriz que esta sendo utilizada para ser essa aproximação. Após alguns passos do algoritmo esse elemento corresponderá a um autovalor aproximado de A_k. Assim que este elemento for designado para ser um autovalor, toma-se a submatriz de A_k desprezando a linha e coluna deste autovalor, como mostra a figura abaixo.

FIGURA 11 - CONVERGÊNCIA DO ALGORITMO QR

$$A_k = \left[\begin{array}{ccc|c} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ \hline 0 & 0 & \varepsilon & \lambda \end{array}\right]$$

Repete-se o processo acima descrito até restar a submatriz 2x2 no canto superior esquerdo. Se ainda for necessário faz-se ainda mais alguns passos, com isso os dois elementos da diagonal tornam-se os dois últimos autovalores aproximados de A .

Demostra-se que o algoritmo QR-modificado converge quadraticamente (LARSON e HANSON 1974) para os autovalores de A.

Para se ter uma idéia da razão de convergência de cada método, considere a seguinte matriz

$$\mathbf{A} = \begin{bmatrix} 4 & 2 & 1 \\ 5 & 7 & 2 \\ 5 & 2 & 1 \end{bmatrix}$$

cujos autovalores com 14 decimais são:

$\lambda_1 = 9{,}41781432379456$;

$\lambda_2 = 2{,}70015847343159$;

$\lambda_3 = -0.11797279722616$.

Para o simples algoritmo QR começando com k=0,1,2,3,... , e truncando na 6 casa decimal, os elementos da diagonal de A_9 foram:

$\lambda_1' = 9{,}417825$;

$\lambda_2' = 2{,}700147$;

$\lambda_3' = -0.117972$.

Para o algoritmo QR-modificado, começando com k=0,1,2,3,..., e truncando na 6 casa decimal o elemento a_{33} de A_5 foi $\lambda_3' = -0.117972$ e os elementos da submatriz A_6 foram $\lambda_2' = 2{,}700158$ e $\lambda_1' = 9{,}417814$. Se for comparado os resultados e a rapidez na convergência, nota-se claramente a maior eficiência por parte do algoritmo QR-modificado.

Através do exemplo pode-se ver a melhora trazida ao algoritmo QR com o parâmetro de variação σ, mas essa melhora pode ser ampliada ainda mais se a matriz A for preparada para tal problema. Tal preparação pode ser realizada pela transformação de Householder deixando A na forma Hessenberg ou tridiagonal. A forma de Hessenberg e tridiagonal são mostradas na figura 12.

FIGURA 12 - FORMAS PREPARATÓRIAS PARA O PROBLEMA DOS AUTOVALORES DO ALGORITMO QR

$$\begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{bmatrix}$$

12.a Forma de Hessenberg

$$\begin{bmatrix} * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ 0 & * & * & * & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{bmatrix}$$

12.b Forma tridiagonal

Considerando A_0 uma matriz com qualquer uma dessas formas obtida fazendo $A_0 = Q^T A Q$, onde Q é uma matriz ortogonal (alcançada através do produto das transformações de Householder) e aplicando o algoritmo QR a essa matriz cujos autovalores são os mesmos de A (garantido por construção) a sequência de matrizes A_k mantém a forma de A_0 , convergindo para triângular superior para o caso de A_0 ter a forma Hessenberg e para diagonal (se A é simetrica) ou bidiagonal superior (será definida adiante) para o caso de A_0 ser tridiagonal. Os autovalores serão os elementos da diagonal dessas matrizes.

5.5.3 Cálculo da decomposição de valor singular

O procedimento para a obtenção da decomposição de valor singular será desenvolvido segundo descrito por LAWSON e HANSON (1974) e GOLUB e REINSCH (1970), a subrotina para o cálculo será tirada a apartir de PRESS et alii (1986).

O cáculo é feito em dois estágios. Primerio , faz-se a redução de A a forma bidiagonal (ver fig. 13) e depois através duma variante no algoritmo QR-modificado se obtém os valores singulares da forma bidiagonal.

5.5.3.1 Redução a forma bidiagonal

Foi dito na secção (5.5.2.2) como preparar uma matriz para o problema dos autovalores deixando-a sob a forma de Hessenberg ou tridiagonal. Aqui a matriz $A \in R^{mxn}$ deve ser reduzida a forma bidiagonal superior para que posteriormente a pré-multiplicação por sua transposta gere uma matriz simétrica e tridiagonal. A forma bidiagonal superior é mostrada na figura abaixo.

FIGURA 13 - FORMA BIDIAGONAL SUPERIOR DE UMA MATRIZ

$$\begin{bmatrix} * & * & 0 & 0 \\ 0 & * & * & 0 \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

A forma bidiagonal consiste de elementos não-nulos na diagonal principal e uma diagonal imediatamente acima da principal, os demais elementos da matriz são nulos.

A redução de A a forma bidiagonal é alcançada por uma sequência de não mais do que 2n-1 transformação de Householder do tipo.

$$Q_k = I_m - 2x_k x_k^T \qquad (k=1,2,\ldots,n) \qquad (5.56)$$

$$H_k = I_n - 2y_k y_k^T \qquad (k=1,2,\ldots,n-2) \qquad (5.57)$$

onde os vetores x e y já estão normalizados. Então

$$B = Q_n \ldots ((Q_2((Q_1 A)H_1))H_2) \ldots H_{n-2} \qquad (5.58)$$

cuja forma será representada por

$$B = \begin{bmatrix} q_1 & e_2 & & & \\ & q_2 & e_3 & 0 & \\ & & \ddots & \ddots & \\ & & & \ddots & e_n \\ & 0 & & & q_n \\ \hline & & 0 & & \end{bmatrix} \begin{matrix} \\ \\ \\ \\ m \\ \\ (m-n) \end{matrix} \qquad (5.59)$$

$$\underbrace{\qquad\qquad}_{n}$$

Denotando $A_1 = A$ e definindo

$$A^{(k+\frac{1}{2})} = Q^k A^k \qquad (k=1,2,\ldots,n) \qquad (5.60)$$

$$A^{(k+1)} = A^{(k+\frac{1}{2})} H^k \qquad (k=1,2,\ldots,n-2) \qquad (5.61)$$

então Q_k é definido tal que

$$A_{ik}^{(k+\frac{1}{2})} = 0 \qquad (i=k+1,\ldots,m) \qquad (5.62)$$

e H_k como

$$A_{kj}^{(k+1)} = 0 \qquad (j=k+2,\ldots,n). \qquad (5.63)$$

Procedendo desta maneira os valores singulares de **B** serão os mesmos de **A**. Para verificar façamos

$$B = Q_n Q_{n-1} \ldots Q_1 A H_1 \ldots H_{n-2} \qquad (5.64)$$

$$B^T = H_{n-2}^T \ldots H_1^T A^T Q_1^T \ldots Q_{n-1}^T Q_n^T \qquad (5.65)$$

o produto B^TB será

$$B^TB = H_{n-2}^T \ldots H_1^T A^T A H_1 \ldots H_{n-2} \; . \qquad (5.66)$$

Observe que (5.66) conserva os autovalores de A^TA mostrando assim a veracidade da afirmação acima.

Uma vez que os valores singulares de **A** são os mesmos de **B**, a decomposição de valor singular de **A** pode ser obtida através da decomposição de valor singular de **B**.

Considere a decomposição de valor singular de **B** como

$$B = \hat{U} S \hat{V}^T \qquad (5.67)$$

onde $\hat{U}$ e $\hat{V}$ são ortogonal e S é diagonal. Então pela condição de ortogonalidade e simetria de H, a decomposição de valor singular de A será

$$A = (Q\hat{U})S(H\hat{V})^T \qquad (5.68)$$

ou

$$A = USV^T. \qquad (5.69)$$

5.5.3.2 Decomposição de valor singular da matriz bidiagonal

Na secção (5.5.2.2) foi visto a eficiência do algoritmo QR-modificado na determinação dos autovalores de uma matriz quadrada. Ora, sendo B bidiagonal superior, o produto B^TB é tridiagonal, caindo exatamente na forma cobiçada para a aplicação do algoritmo QR-modificado o qual ficaria mais econômico se fosse utilizadas as transformações de Givens em vez das trasformações de Householder, pois restaria só um componente abaixo da diagonal principal para ser zerado.

Pensando desta forma nosso problema já estaria resolvido, bastaria fazer o produto B^TB e aplicar o algoritmo QR-modificado para a matriz tridiagonal simétrica e ainda com garantia da convergência quadrática. O algoritmo desta forma já seria ótimo, mas melhor ainda se não housse a necessidade de explicitar B^TB , pois este produto pode conduzir a perda de informações por causa de erros de arreondamento. A SVD previne-se também desse fato, sendo um dos motivos por ser a técnica recomendada para a determinação dos valores singulares de uma matriz.

O procedimento para obter a decomposição de valor singular de B é um processo iterativo da forma

$$B_{(0)} \rightarrow B_{(1)} \rightarrow \cdots \rightarrow S \qquad (5.70)$$

$$B_{k+1} = U_k^T B_k V_k \ , \qquad k=1,2,3,... \qquad (5.71)$$

onde U_k e V_k são ortogonal. As matrizes V_k são escolhidas de modo que a sequência $M_k = B_k^T B_k$ convergem para a forma diagonal, enquanto as matrizes

U_k são escolhidas de modo que B_k permaneça bidiagonal superior para todo k.

Um passo deste algoritmo é descrito como segue.

Dada B_k, o algoritmo determina os autovalores λ_1 e λ_2 da submatriz inferior direita $M_{(2,2)}$ de M_k, aquele que mais se aproximar do elemento inferior direito de M_k será escolhido para ser σ_k do algoritmo QR-modificado .

A matriz V_k é determinda pelo processo usual do algoritmo QR-modificado

$$(M_k - \sigma_k I_n) = V_k R_k \qquad (5.72)$$

onde R_k é triangular superior.

A matriz U_k é determinada de maneira que

$$B_{k+1} = U_k^T B_k V_k \qquad (5.73)$$

seja bidiagonal superior.

O passo detalhado deste algoritmo não é feito como sugere (5.72) e (5.73), na verdade M_k nunca é formada e os autovalores de $M_{(2,2)}$ são calculados implicitamente.

Abandonando o índice k por comodidade de notação e formando $M = B^T B$ onde B tem a forma (5.59), a submatriz $M_{(2,2)}$ será

$$\begin{bmatrix} q_{n-1}^2 + e_{n-1}^2 & e_n q_{n-1} \\ e_n q_{n-1} & q_n^2 + e_n^2 \end{bmatrix} \qquad (5.74)$$

cuja equação característica será

$$(q_{n-1}^2+e_{n-1}^2-\lambda)(q_n^2+e_n^2-\lambda)-(e_n q_{n-1})^2 = 0. \qquad (5.75)$$

Como procura-se o λ_i que mais se aproxima de $q_n^2+e_n^2$, é conveniente fazer a substituição

$$\delta = q_n^2+e_n^2 - \lambda \qquad \Rightarrow \qquad \lambda = q_n^2+e_n^2 - \delta \tag{5.76}$$

e substituindo λ de (5.76) em (5.75) resulta

$$\delta^2 +(q_{n-1}^2+e_{n-1}^2-q_n^2-e_n^2)\delta-(e_n q_{n-1})^2=0 \tag{5.77}$$

dividindo toda a expressão por $(e_n q_{n-1})^2$ tem-se

$$\frac{\delta^2}{(e_n\, q_{n-1})^2}+\frac{(q_{n-1}^2+e_{n-1}^2-q_n^2-e_n^2)\delta}{(e_n q_{n-1})(e_n q_{n-1})}-1=0 \tag{5.78}$$

chamando

$$\gamma=\frac{\delta}{(e_n q_{n-1})} \tag{5.79}$$

e

$$f=\frac{(q_n^2-q_{n-1}^2+e_n^2-e_{n-1}^2)}{2(e_n q_{n-1})} \tag{5.80}$$

e substituindo em (215) obtem-se

$$\gamma^2-2f\gamma-1=0\ . \tag{5.81}$$

Aplicando Baskara para obter as raizes tem-se

$$\hat{\gamma}=\frac{2f\pm\sqrt{4f^2+4}}{2}=f\pm\sqrt{1+f^2}\ . \tag{5.82}$$

Analisando (218), percebe-se que uma raiz será sempre negativa e outra positiva, além disso, uma vai ser menos o inverso da outra, com base nisso pode-se escrever

$$\hat{\gamma}=\frac{1}{t} \tag{5.83}$$

onde

$$t=\begin{cases} -f-(1+f^2)^{1/2}\ , & \text{se } f\geq 0 \\ -f+(1+f^2)^{1/2}\ , & \text{se } f<0 \end{cases}$$

levando (5.83) em (5.79) e posteriormente em (5.76), resulta

$$\hat{\lambda}=q_n^2+e_n^2-\frac{e_n q_{n-1}}{t}=q_n^2+e_n(e_n-\frac{q_{n-1}}{t}) \tag{5.84}$$

$$\sigma=\hat{\lambda}\ . \tag{5.85}$$

observação: com t escolhido da forma acima o $\hat{\lambda}$, será sempre o valor mais próximo do elemento inferior direito de $M_{(2,2)}$.

A dedução acima forneceu o valor de σ a ser usado em (5.72). A seguir será dado o procedimento para oter U e V, mas antes porém deve ser enunciada uma proposição (LAWSON e HANSON 1974).

Proposição 1. Se M é tridiagonal com todos elementos não-nulos, V é ortogonal, σ é um escalar arbitrário,

$V^T M V$ é tridiagonal

e a primeira coluna de $V^T(M-\sigma I)$ é zero abaixo do primeiro elemento, então $V^T(M-\sigma I)=R$ é triangular superior.

Note que a primeira coluna de (M-σI) é

$$\begin{bmatrix} q_1^2 - \sigma \\ q_1 e_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (5.86)$$

e pela condição da proposição 1 de $V^T(M-\sigma I)$ ter zeros embaixo do primeiro elemento e $V^T(M-\sigma I)=R$, implica que V^T deve anular o segundo componente de (5.86). Para isso basta V^T ser uma trasformação de Givens.

Acontece que esta trasformação vai tirar B da forma bidiagonal, pois isto faria gerar um elemento na posição b_{21} . Para manter B na forma bidiagonal e conseqüentemente as condições da proposição 1, uma sequência de transformações de Givens devem ser aplicadas na ordem

$$\bar{B} = U_n(...U_3((U_2(BV_2))V_3)...V_n \qquad (5.87)$$

onde

$\bar{B}=B_{k+1}$;

$B=B_k$;

$V_k = V_2 V_3 \ldots V_n$;

$U_k^T = U_n \ldots U_3 U_2$

com

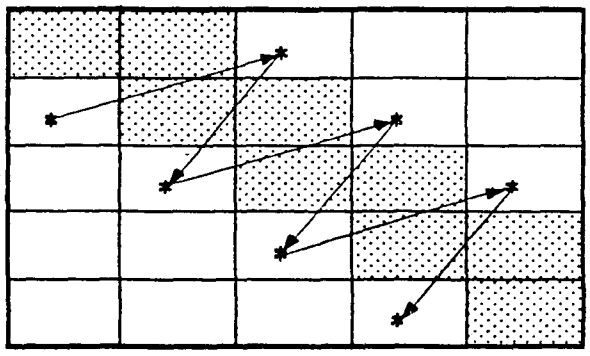
$$U_{i\ mxm} = \begin{bmatrix} 1 & 0 & & & & & \\ 0 & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & s & & \\ & & & -s & c & & \\ & & & & & 1 & \\ & & & & & & \ddots \ 0 \\ & & & & & 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ (i-1) \\ (i) \\ \\ \\ \end{matrix} \quad , i = 2,3,\ldots,n$$

$$\begin{matrix} & & & & & & (i-1)(i) \end{matrix}$$

e V_i n-dimensional é definida de forma análoga para i=2,3,4,...,n. A V_2 é determinada para zerar o segundo componete de (5.86), as demais para manter as condições mensionadas anteriormente.

O comportamento dessas transformações no processo de transição de $B_k \rightarrow B_{k+1}$ são mostrados pela figura abaixo.

FIGURA 14 - PROCESSO " CHASING "



as quais tem a seguinte interpretação:

V_2 não aniquila nada e gera b_{21} ;

$\mathbf{U}_2^T$ aniquila b_{21} e gera b_{13} ;

$\mathbf{V}_3$ aniquila b_{13} e gera b_{32} ;

.

.

.

$\mathbf{U}_n^T$ aniquila $b_{n,n-1}$ e não gera nada.

O procedimento descrito acima é um passo do algoritmo da decomposição de valor singular de **B**, a sua aplicação resulta na transição de $\mathbf{B}_k \rightarrow \mathbf{B}_{k+1}$, lembrando que a convergência se dá quando (5.70) é alcançada.

Existe ainda algumas observações com respeito ao processo de convergência e serão tratados a seguir.

O procedimento descrito acima é um passo do algoritmo da decomposição de valor singular de **B**, a sua aplicação resulta na transição de $\mathbf{B}_k \rightarrow \mathbf{B}_{k+1}$, lembrando que a convergência se dá quando (5.70) é alcançada. Existe ainda algumas observações com respeito ao processo de convergência e serão tratados a seguir.

5.5.3.3 Teste de convergência

Suponha ser δ uma certa tolerância, se $|e_n| \leq \delta$ aceita-se $|q_n|$ como um valor sigular de **A** e toma-se a submatriz de ordem n-1 de $\mathbf{B}_k$ (no estágio em questão) para prosseguir os cálculos. Entretanto, se $|e_k| \leq \delta$ para $k \neq n$, matriz é dividida em duas e os valores singulares de cada bloco podem ser tratados independentemente.

Suponha que em algum estágio $|q_k| \leq \delta$, então por convenientes transformações de Givens a matriz pode ser dividida em dois blocos. O procedimento é mostrado por:

$$\mathbf{B}_k' = \mathbf{T}_{k,n}\mathbf{T}_{k,n-1}\cdots\mathbf{T}_{k,k+1}\mathbf{B}_k \qquad (5.88)$$

sendo que

e_{k+1} é aniquilado por $\mathbf{T}_{k,k+1}$, mas $b_{k,k+2}$ e $\delta_{k+1,k}$ são gerados ;

$b_{k,k+2}$ é aniquilado por $T_{k,k+2}$, $b_{k,k+3}$ e $\delta_{k+2,k}$ são gerados ;

.

.

.

$b_{k,n}$ é aniquilado por $T_{k,n}$ mas $\delta_{n,k}$ é gerado.

A matriz $\mathbf{B}_k^{'}$ ficará com a forma

$$\mathbf{B}_k^{'} = \begin{bmatrix} q_1 & e_2 & & & & & \\ & \ddots & \ddots & & 0 & & \\ & & \ddots & e_k & & & \\ & & & \bar{q}_k & 0 & \cdots & 0 \\ & & & & \bar{q}_{k+1} & \bar{e}_{k+2} & \\ & 0 & & & & \ddots & \\ & & & & 0 & & \bar{e}_n \\ & & & & & & \bar{q}_n \end{bmatrix}^{(k)} \qquad (5.89)$$

Se for escolhido $\delta = \mathbf{u}\|\mathbf{B}_0\|_\infty$ (**u** unidade de arredondamento do computador) então os elementos $\bar{q}_k, \delta_{k+1}, \ldots, \delta_n$ podem ser desprezados e $\mathbf{B}_k^{'}$ pode ser dividida em dois blocos como

$$\mathbf{B}_k^{'} = \begin{bmatrix} \mathbf{B}_1^{'} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix} \qquad (5.90)$$

e tratados independentemente.

sendo $\bar{q}_k$ de $\mathbf{B}_1^{'}$ considerado nulo e_k poderá ser zerado por uma sequência de transformação de Givens da forma

$$\mathbf{B}_1^{'}\mathbf{T}_{k-1}\ldots\mathbf{T}_1 = \mathbf{B}_1^{*} \qquad (5.91)$$

e $\bar{q}_k$ passa a ser um valor singular de **A**.

Se $q_k=0$, então pelo menos um valor singular é zero e a mesma sequência (5.88) pode ser aplicada (a diferença é a não geração dos δ_{ik} , i=k,k+1,...,n) para reduzir a ordem da matriz ou parcioná-la conforme seja a posição de k.

5.5.3.4 Formação da decomposição de valor singular de A

A decomposição de valor singular de A

$$\mathbf{A}=\mathbf{USV}^T \tag{5.92}$$

é formada a partir da SVD de B ,ver (5.67)

$$\mathbf{B}=\hat{\mathbf{U}}\mathbf{B}\hat{\mathbf{V}}^T. \tag{5.93}$$

Uma vez que os elementos da diagonal de B tende para seus valores singulares, ou seja, a diagonal de S a cada passo do processo iterativo pela aplicação do produto das transformações de Givens, então o produto de todas as transformações aplicadas tenderão para os autovetores de B.

Um detalhe é que através da aplicações das transformações de Givens os elementos da diagonal de S podem vir com sinais positivos ou negativos e por ser $\mathbf{B}^T\mathbf{B}$ no mínimo semi-definida positiva esses deveriam ser ou nulos ou positivos para tornar positivos os que estão negativos, deve-se usar uma matriz diagonal D com elementos +1 e -1 conforme seja o respectivo sinal do elemento de S. Então

$$\mathbf{S}=\mathbf{SD} \tag{5.94}$$

a matriz $\hat{\mathbf{U}}$ e $\hat{\mathbf{V}}$ de (5.93) será formada por

$$\hat{\mathbf{U}}=\mathbf{U}_n^T\ldots\mathbf{U}_1^T \tag{5.95}$$

$$\hat{\mathbf{V}}=\mathbf{V}_1\ldots\mathbf{V}_v \tag{5.96}$$

a matriz Q e H de (5.68) será formada por

$$\mathbf{Q}=\mathbf{Q}_n\mathbf{Q}_{n-1}\cdots\mathbf{Q}_1 \tag{5.97}$$

$$\mathbf{H}=\mathbf{H}_1\mathbf{H}_2\ldots\mathbf{H}_{n-2} \tag{5.98}$$

de (5.68): $\mathbf{A}=(\mathbf{Q}\hat{\mathbf{U}})\mathbf{S}(\mathbf{H}\hat{\mathbf{V}})^T$ (5.99)

assim

$$\mathbf{U} = \mathbf{Q}_n\mathbf{Q}_{n-1}\ldots\mathbf{Q}_1\mathbf{U}_n^T\ldots\mathbf{U}_1^T \qquad (5.100)$$

$$\mathbf{V} = \mathbf{H}_1\mathbf{H}_2\ldots\mathbf{H}_{n-2}\mathbf{V}_1\ldots\mathbf{V}_v \qquad (5.101)$$

e finalmente

$$\mathbf{A}=\mathbf{U}\mathbf{S}\mathbf{V}^T \qquad (5.102)$$

5.5.3.5 Subrotina para a decomposição de valor singular

Como mensionado anteriormente essa subrotina foi tirada a apartir de PRESS et alii (1986). Nenhuma modificação foi feita .

```
SUBROUTINE SVDCMP(A,M,N,MP,NP,W,V)
      dada a matirz A, com dimensão lógica M por N e dimensão fisíca MP por NP, esta rotina
      cálcula sua decomposição de valor singular. A=U.W.V . A matriz U substitui A na saída. A
      matriz diagonal de valores singulares W é armazenada no vetor W. A matriz V(não V)é
      armazenada como V. M deve ser maior ou igual a N. Se for menor então A deve ser completada
      até ficar quadrada com linhas zeros.
PARAMETER (NMAX=100)                máximo valor antecipado de N
DIMENSION  A(MP,NP),W(NP),V(NP,NP),RV1(NMAX)
IF (M.LT.N) PAUSE ' voce deve aumentar linhas extras em A'
      redução de Householder para a forma bidiagonal.
G=0.0
SCALE=0.0
ANORM=0.0
DO 25 I=1,N
      L=I+1
      RV1(I)=SCALE+G
      G=0.0
      S=0.0
      SCALE=0.0
      IF (I.LE.M) THEN
            DO 11 K=I,M
                  SCALE=SCALE+ABS(A(K,I))
             11 CONTINUE
            IF (SCALE.NE.0.0) THEN
                  DO 12 K=I,M
                        A(K,I)=A(K,I)/SCALE
                        S=S+A(K,I)*A(K,I)
                   12 CONTINUE
                  F=A(I,I)
                  G=-SIGN(SQRT(S),F)
                  H=F*G-S
                  A(I,I)=F-G
                  IF (I.NE.N) THEN
                        DO 15 J=L,N
                              S=0.0
                              DO 13 K=I,M
```

```
                        S=S+A(K,I)*A(K,J)
                    13 CONTINUE
                    F=S/H
                    DO 14 K=I,M
                        A(K,J)=A(K,J)+F*A(K,J)
                    14 CONTINUE
                15 CONTINUE
                ENDIF
                DO 16 K=I,M
                    A(K,I)=SCALE*A(K,I)
                16 CONTINUE
        ENDIF
ENDIF
W(I)=SCALE*G
G=0.0
S=0.0
SCALE=0.0
IF ((I.LE.M). AND. (I.NE.N)) THEN
        DO 17 K=L,N
            SCALE=SCALE+ABS(A(I,K))
        17 CONTINUE
        IF (SCALE.NE.0.0) THEN
            DO 18 K=L,N
                A(I,K)=A(I,K)/SCALE
                S=S+A(I,K)*A(I,K)
            18 CONTINUE
            F=A(I,L)
            G=-SIGN(SQRT(S),F)
            H=F*G-S
            A(I,L)=F-G
            DO 19 K=L,N
                RV1(K)=A(I,K)/H
            19 CONTINUE
            IF (I.NE.M) THEN
                DO 23 J=L,M
                    S=0.0
                    DO 21 K=L,N
                        S=S+A(J,K)*A(I,K)
                    21 CONTINUE
                    DO 22 K=L,N
                        A(J,K)=A(J,K)+S*RV1(K)
                    22 CONTINUE
                23 CONTINUE
                ENDIF
                DO 24 K=L,N
                    A(I,K)=SCALE*A(I,K)
                24 CONTINUE
            ENDIF
        ENDIF
        ANORM=MAX(ANORM,(ABS(W(I))+ABS(RV1(I))))
  25 CONTINUE
        acumulação das transformações a direita
DO 32 I=N,1,-1
        IF (I.LT.N) THEN
            IF (G.NE.0.0) THEN
                DO 26 J=L,N dupla divisão p/ evitar subfluxo
            V(J,I)=(A(I,J)/A(I,L))/G
                26 CONTINUE
```

```
                DO 29 J=L,N
                        S=0.0
                        DO 27 K=L,N
                                S=S+A(I,K)*V(K,J)
                           27 CONTINUE
                        DO 28 K=L,N
                                V(K,J)=V(K,J)+S*V(K,I)
                           28 CONTINUE
                     29 CONTINUE
                ENDIF
                DO 31 J=L,N
                        V(I,J)=0.0
                        V(J,I)=0.0
                   31 CONTINUE
                ENDIF
                V(I,I)=1.0
                G=RV1(I)
                L=I
             32 CONTINUE
                acumulação das trasformações da esquerda
          DO 39 I=N,1,-1
                L=I+1
                G=W(I)
                IF (I.LT.N) THEN
                        DO 33 J=L,N
                                A(I,J)=0.0
                           33 CONTINUE
                ENDIF
                IF (G.NE.0.0) THEN
                        G=1.0/G
                IF (I.NE.N) THEN
                        DO 36 J=L,N
                                S=0.0
                                DO 34 K=L,M
                                        S=S+A(K,I)*A(K,J)
                                   34 CONTINUE
                                F=(S/A(I,I))*G
                                DO 35 K=I,M
                                        A(K,J)=A(K,J)+F*A(K,I)
                                   35 CONTINUE
                             36 CONTINUE
                ENDIF
                DO 37 J=I,M
                        A(J,I)=A(J,I)*G
                   37 CONTINUE
          ELSE
                DO 38 J=I,M
                        A(J,I)=0.0
                   38 CONTINUE
          ENDIF
          A(I,I)=A(I,I)+1.0
       39 CONTINUE
          diagonalização da forma bidiagonal
     DO 49 K=N,1,-1  loop sobre valores singulares
      DO 48 ITS=1,30        núm. de iterções permissiveis
          DO 41 L=K,1,-1
                NM=L-1  rv1(1) é permitdo ser zero
                IF((ABS(RV1(L))*ANORM).E.Q.ANORM)          GO TO 2
```

```
                    IF ((ABS(W(NM))*ANORM).EQ.ANORM)        GO TO 1
                  41 CONTINUE
                 C=0.0
                 S=1.0
                 DO 43 I=L,K
                  F=S*RV1(I)
                  IF ((ABS(F)+ANORM).NE.ANORM) THEN
                        G=W(I)
                        H=SQRT(F*F+G*G)
                        W(I)=H
                        H=1.0/H
                        C=(G*H)
                        S=-(F*H)
                        DO 42 J=1,M
                              Y=A(J,NM)
                              Z=A(J,I)
                              A(J,NM)=(Y*C)+(Z*S)
                              A(J,I)=-(Y*S)+(Z*C)
                           42 CONTINUE
                  ENDIF
               43 CONTINUE
2       Z=W(K)
        IF (L.EQ.K) THEN  convergência
              IF (Z.LT.0.0) THEN  valor singular é feito positivo
                    W(K)=-Z
                    DO 44 J=1,N
                          V(J,K)=-V(J,K)
                       44 CONTINUE
              ENDIF
              GO TO 3
        ENDIF
        IF (ITS.EQ.30) PAUSE 'Não converge em 30 itirações'
        X=W(L)          variação a partir da matriz 2x2
        NM=K-1
        Y=W(NM)
        G=RV1(NM)
        H=RV1(K)
        F=((Y-Z)*(Y+Z)+(G-H)*(G-H))/(2.0*H*Y)
        G=SQRT(F*F+1.0)
        F=((X-Z)*(X+Z)+H*((Y/(F+SIGN(G,F)))-H))/X
        próxima transformação QR
        C=1.0
        S=1.0
        DO 47 J=L,NM
              I=J+1
              G=RV1(I)
              Y=W(I)
              H=S*G
              G=C*G
              Z=SQRT(F*F+H*H)
              RV1(J)=Z
              C=F/Z
              S=H/Z
              F=(X*C)+(G*S)
              G=-(X*S)+(G*C)
              H=Y*S
              Y=Y*C
              DO 45 JJ=1,N
```

```
                X=V(JJ,J)
                Z(JJ,J)=(X*C)+(Z*S)
                V(JJ,J)=-(X*S)+(Z*C)
             45 CONTINUE
          Z=SQRT(F*F+H*H)
          W(J)=Z rotação pode ser arbitrária se Z=0
          IF (Z.NE.0.0) THEN
                Z=1.0/Z
                C=F*Z
                S=H*Z
          ENDIF
          F=(C*G)+(S*Y)
          X=-(S*G)+(C*Y)
          DO 46 JJ=1,M
                Y=A(JJ,J)
                Z=A(JJ,I)
                A(JJ,J)=(Y*C)+(Z*S)
                A(JJ,I)=-(Y*S)+(Z*C)
             46 CONTINUE
       47 CONTINUE
    RV1(L)=0.0
    RV1(K)=F
    W(K)=X
  48 CONTINUE
3  CONTINUE
  49 CONTINUE
RETURN
END
```

6. TESTES REALIZADOS PARA O PROBLEMA DE MÍNIMOS QUADRADOS LINEAR

Este é dedicado exclusivamente aos testes para comprovar o que foi descrito ao longo do desenvolvimento teórico. Para isso escolheu-se nove testes a fim de elucidar os principais pontos a enfatizar.

Os resultados dos testes foram obtidos através da execução de um programa computacional em linguagem "Turbo Pascal - versão 6.0" em uma computador 386-Sx. Este programa foi confeccionado utilizando as subroutinas descritas ao longo do capítulo 5, algumas sofreram algum tipo de implementação, mas nada que mude a velocidade e a estabilidade das subroutinas originais. O programa não será anexado neste, mas será fornecida uma cópia do mesmo ao Curso de Pós-Graduação em Ciências Geodésicas que poderá ser consuntado quando solicitado.

6.1 DESCRIÇÕES PRELIMINARES

Nesta secção será feita uma breve descrição dos objetivos de cada teste.

1°. teste: Este teste tem o simples objetivo de mostrar a variação na solução quando as equações são ponderadas por pesos iguais e por pesos diferentes .

2°. teste: Este teste mostra a equivalência das soluções entre os métodos descritos neste trabalho, quando o problema é bem-condicionado.

3°. teste: Este teste mostra a perda de informações ao se formar as equações normais de mínimos quadrados para a solução do problema.

4°. teste: Este teste tem por finalidade mostrar o comportamento da solução e resíduo quando o vetor **b** é perturbado no problema de mínimos quadrados linear.

5°. teste: Este teste tem por objetivo mostrar o comportamento da solução e resíduo quando a matriz dos coeficientes é perturbada.

6°. teste: Este teste caracteriza vários pontos; ele mostra a desvantagem em se formar as equações normais, caracteriza a dificuldade na determinação do posto de uma matriz, mostra a equivalência na solução para os métodos baseados na decomposição ortogonal, QR e SVD (quando não se cogita a solução de comprimento mínimo).

7°. teste: Neste teste, Cholesky detecta a deficiência de posto(enquanto que os demais fornecem uma solução errada). A SVD é usada na análise e fornece a solução de comprimento mínimo.

8°. teste: Neste teste a matriz dos coeficientes tem posto deficiente, Cholesky nao detecta a deficiência, enquanto que os demais dectectam. A análise é feita pela SVD que fornece a solução de comprimento mínimo.

9°. teste: Este teste não será impresso completamente no trabalho, é colocado para mostrar o tempo gasto em cada método para obter a solução e matriz de covariância em um sistema inconsistente de 38 equação e 34 incógnitas.

6.2 TESTES

TESTE 01: Diferença entre as soluções do problema ponderado por pesos iguais e por pesos diferentes.

a) pesos iguais:

Matriz A	vetor b	desv. padrões
3 2 1	1	5
4 5 9	2	5
2 1 0	4	5
3 4 5	7	5

Vetor x (solução por QR)	$\|\|V\|\|_w$ =comp. de resíduo
-2.193 548 387 1	0.660 400 660 40
5.870 967 741 9	
-2.064 516 129 0	

b) pesos diferentes

desv. padrões	Vetor x (solução por QR)	comp. resíduo
5	-3.093 515 358 3	0.480 329 807 06
7	6.779 522 184 3	
8	-2.169 283 276 5	
1		

TESTE 02: Equivalência dos métodos para um problema bem-condicionado.

MATRIZ A	Vetor b
3 5 1	1
2 3 9	2
1 7 3	5
4 2 1	3

Os resultados obtidos pelo programa estão na tabela abaixo:

	Gauss-J	Versol	Cholesky	QR	SVD
x	0.14393627249	0.14393627249	0.14393627249	0.14393627249	0.14393627249
	0.50089273451	0.50089273451	0.50089273451	0.50089273451	0.50089273452
	0.05892047795	0.05892047795	0.058920477956	0.058920477935	0.058920477955
||V||	2.7053742035	2.7053742035	2.7053742035	2.7053742035	2.7053742035

TESTE 03: Perda de informações em se formar as equações normais de mínimos quadrados e cuidados na solução.

Matriz A		Vetor b
1	1.02	7
1	1	3
1	1	2

a) Precisão com unidade de arredondamento u=1.0E-12

Matriz A^TA		Vetor A^Tb	Vetor solução
3.0	3.02	12	-222.50
3.02	3.0404	12.14	225.00

b) Precisão com unidade de arredondamento u=1.0E-04

Matriz A^TA		Vetor A^Tb	Vetor solução
3.0	3.02	12	457
3.02	3.04	12.14	-450

Comentário: Para a precisão do primeiro caso, a matriz A^TPA é definida positiva e a solução verdadeira, isso já não acontece com a precisão reduzida do segundo caso, a matriz A^TPA torna-se indefinida pois seus autovalores com três digitos significativos são: λ_1=-6,62E-05 e λ_2=6,04 e a solução completamente errada.

Este teste mostra o cuidado que se deve ter ao selecionar um método de solução para o problema de mínimos quadrados linear. Para o segundo caso recomendar-se-ia utilizar, QR ou o SVD. Por Cholesky o problema seria detectado mas por Gauss-Jordan e Versol a solução fornecida seria errada.

TESTE 04: Este teste tem a finalidade de mostrar o comportamento da solução e do residual quando o vetor **b** é pertubado inteiramente no espaço coluna de **A** e no espaço nulo de A^T.

a) Problema original (sem pertubação)

MATRIZ A	Vetor b
1 1	1
1 0	0
0 1	-5

Vetor x (solução por QR)	Comp. do resíduo
2.000	3.464 101 615 1
-3.000	

b) O problema com o vetor b pertubado inteiramente no espaço coluna de A.

Vetor b+δb	Vetor x (solução por QR)	Comp. do resíduo
1.002000	2.001	3.464 101 615 1
0.001000	-2.999	
-4.999000		

comentário: Ocorreu modificação na solução mas não no resíduo.

c) Pertubação no vetor b inteiramente no espaço nulo de A^T.

Vetor b	Vetor x (solução por QR)	Comp. do resíduo
1.001	2.000	3.465 833 665 9
-0.001	-3.000	
-5.001		

Cometário: Não ocorreu mudança na solução, mas houve mudança no resíduo.

Comentário final: Este teste mostra o cuidado que se deve ter ao analisar se um problema é ou não mal-condicionado. Se for voltada a atenção somente para a solução, viu-se que dependendo da variação feita no vetor b esta não sofre mudança, mas isto não garante que o problema é bem-condionado, o resíduo poderá ficar muito elevado (nas ciências observacionais, a consequência seria na variância da unidade de peso a-posteriori).

TESTE 05: Este teste tem por objetivo mostrar o comportamento da solução e do resíduo quando a matriz A é pertubada.

a) Problema original sem pertubação

```
MATRIZ A                Vetor b

1  1                    2.000 00
1  1                    0.000 01
1  1                    4.000 01

METODO DA DECOMP. DE VALOR SINGULAR

MATRIZ U                                    MATRIZ V

-8.1649794175E-01 -5.7734834469E-01    -7.0710913820E-01 -7.0710442416E-01
4.0824692964E-01 -5.7735123144E-01     7.0710442416E-01 -7.0710913820E-01
4.0824692964E-01 -5.7735123144E-01

Vetor W (valores singulares)          Vetor X (solução)

5.7734822257E-06                       9.9999911726E-01
2.4494979078E+00                       1.0000008827E+00

Comprim. do resíduo = 2.8284271248E+00
```

b) Matriz A pertubada somente no seu espaço coluna de A (a coluna não nula da matriz de pertubação é multipla da segunda coluna de U no problema (a)).

```
MATRIZ A+δA

1.00001  1.00000
1.00001  1.00001
1.00001  1.00001

METODO DA DECOMP. DE VALOR SINGULAR

MATRIZ U                                    MATRIZ V

-5.7734834471E-01  8.1649794174E-01   -7.0710795968E-01  7.0710560269E-01
-5.7735123143E-01 -4.0824692965E-01   -7.0710560269E-01 -7.0710795968E-01
-5.7735123143E-01 -4.0824692965E-01

Vetor W (valores singulares)              Vetor x (solução)

2.4495101552E+00                          9.9998909219E-01
5.7735110931E-06                          1.0000009079E+00

Comprim. do resíduo = 2.8284271248E+00
```

c) Pertubação em A somente no espaço coluna (a coluna não nula da matriz de pertubação é multipla da primeira coluna de U do problema (a)).

MATRIZ A+δA

```
100000  0.99998
100000  1.00002
100000  1.00002
```

METODO DA DECOMP. DE VALOR SINGULAR

```
MATRIZ U                                    MATRIZ V

-8.1650202420E-01 -5.7734257116E-01  -7.0710913826E-01  -7.0710442410E-01
 4.0824284714E-01 -5.7735411816E-01   7.0710442410E-01  -7.0710913826E-01
 4.0824284714E-01 -5.7735411816E-01

Vetor W (valores singulares)                Vetor x (solução)

2.3093932540E-05                            1.7500046995E+00
2.4494979079E+00                            2.5000030047E-01

Comprim. do resíduo  =  2.8284271248E+00
```

d) Pertubação em A somente no espaço nulo de A^T (a coluna não nula da matriz de pertubação é multipla da última coluna de U, a qual não é calculada pelo algoritmo apresentado. Tal pertubação foi obtida conforme exposto no exemplo da pg. 36).

MATRIZ A

```
1.00000  1.00000
1.00000  1.00000
1.00000  1.00002
```

METODO DA DECOMP. DE VALOR SINGULAR

```
MATRIZ U                                    MATRIZ V

-4.0825095508E-01  5.7734834468E-01  -7.0710913821E-01  7.0710442415E-01
-4.0825106914E-01  5.7734834468E-01   7.0710442415E-01  7.0710913821E-01
 8.1649385926E-01  5.7735411818E-01

Vetor W (valores singulares)                Vetor X (solucao)

1.1546966064E-05                           -1.4999926775E+05
2.4494979078E+00                            1.5000026776E+05

Comprim. do resíduo  =  1.4142064913E+00
```

comentário: quando a pertubação se dá somente no espaço coluna de A, o resíduo não sofre alteração (ver equação (4.50), mas a solução sim. A maior variação na solução é quando a pertubação em A multipla do menor valor singular. Quando a pertubação ocorre no espaço nulo de A^T, ambos sofrem alterações, a solução e o resíduo.

TESTE 06: Este teste reune várias características de uma só vez. Ele mostra a desvantagem em se formar as equações normais, caracteriza a dificuldade na determinação do posto e mostra a equivalência das soluções pelo método QR e SVD.

obs.: neste e nos próximos dois testes a seguir será impressa a matriz de covariâncias dos parâmetros ajustados (ver GEMAEL 1974) a fim de tornar este trabalho mais aplicativo e mostrar uma característica a mais na análise. O programa foi implementado para fazer esses cálculos.

MATRIZ A

-1.3405550000E-01	-2.0162830000E-01	-1.6930780000E-01	-1.8971990000E-01	-1.7387230000E-01
-1.0379480000E-01	-1.5766340000E-01	-1.3346260000E-01	-1.4848550000E-01	-1.3597690000E-01
-8.7795970000E-02	-1.2883870000E-01	-1.0683010000E-01	-1.2011800000E-01	-1.0932970000E-01
2.0585540000E-02	3.3533100000E-03	-1.6412700000E-02	7.8606000000E-04	2.7165900000E-03
-3.2480930000E-02	-1.8767990000E-02	4.1063900000E-03	-1.4058940000E-02	-1.3843910000E-02
5.9676620000E-02	6.6677140000E-02	4.3521530000E-02	5.7404380000E-02	5.0249620000E-02
6.7124570000E-02	7.3524370000E-02	4.4897700000E-02	6.4718620000E-02	5.8764550000E-02
8.6871860000E-02	9.3682960000E-02	5.6723270000E-02	8.1410430000E-02	7.3023200000E-02
2.1496620000E-02	6.2226620000E-02	7.2134860000E-02	6.2000690000E-02	5.5709310000E-02
6.6874070000E-02	1.0344510000E-01	9.1538490000E-02	9.5082230000E-02	8.3936670000E-02
1.5879070000E-01	1.8088340000E-01	1.1540690000E-01	1.6160730000E-01	1.4796480000E-01
1.7642890000E-01	2.0361830000E-01	1.3057860000E-01	1.8385730000E-01	1.7005550000E-01
1.1414080000E-01	1.7259610000E-01	1.4816470000E-01	1.6007470000E-01	1.4374100000E-01
7.8460380000E-02	1.4669560000E-01	1.4365800000E-01	1.4003840000E-01	1.2571180000E-01
1.0803180000E-01	1.6994620000E-01	1.4971520000E-01	1.5885310000E-01	1.4301550000E-01

Vetor b	desv. padroes
-0.436100	1.000000
-0.343700	1.000000
-0.265700	1.000000
-0.039200	1.000000
0.019300	1.000000
0.074700	1.000000
0.093500	1.000000
0.107900	1.000000
0.193000	1.000000
0.205800	1.000000
0.260600	1.000000
0.314200	1.000000
0.352900	1.000000
0.361500	1.000000
0.364700	1.000000

M E T O D O D A E L I M. D E G A U S S - J O R D A N

VETOR X (solucao)
-6.3554977127E-01
-2.4536948396E+00
 2.0545839088E+00
-2.4429301733E+00
 6.5086492249E+00

Comprim. do resíduo = 1.3926311208E-04

Variancia da unidade peso a-posteriori = 1.9394214386E-09

MATRIZ COVARIANCIA

-1.1572290549E+03 1.6070657777E+03 -1.2752070849E+03 1.4832983693E+03 -1.3481476116E+03
1.6092387117E+03 -2.2322645150E+03 1.7740784638E+03 -2.0692511446E+03 1.8782342029E+03
-1.2745334034E+03 1.7707484479E+03 -1.4042286634E+03 1.6316147406E+03 -1.4837181109E+03
1.4776132562E+03 -2.0585709253E+03 1.6262135403E+03 -1.8767367706E+03 1.7122187088E+03
-1.3451201753E+03 1.8714985226E+03 -1.4811668562E+03 1.7149612776E+03 -1.5621537109E+03

M E T O D O D A 'S U B R O T. V E R S O L'

VETOR X (solução)

-2.2500000000E+00
 0.0000000000E+00
 2.5000000000E-01
-1.0000000000E+00
 4.7500000000E+00

Comprim. do resíduo = 1.4788362216E-01

Variancia da unidade peso a-posteriori = 2.1869565704E-03

MATRIZ COVARIÂNCIA

-9.3282762505E+08 1.2968386265E+09 -1.0274931333E+09 1.1919963570E+09 -1.0847694497E+09

1.2955086700E+09 -1.7982087964E+09 1.4278595897E+09 -1.6628652123E+09 1.5104783275E+09

-1.0279054639E+09 1.4298977448E+09 -1.1319469553E+09 1.3111882164E+09 -1.1941082220E+09

1.1954759624E+09 -1.6694021007E+09 1.3144940514E+09 -1.5081938694E+09 1.3798557355E+09

-1.0866224090E+09 1.5146009385E+09 -1.1956697317E+09 1.3781771307E+09 -1.2581069067E+09

M É T O D O D E C H O L E S K Y

A matriz ATPA nao e definida positiva e a solucao foge ao alcance do metodo de CHOLESKY. Para este caso, a solucao do problema de minimos quadrados linear atraves das equacoes normais nao e confiavel.

M E T O D O Q R

Vetor X (solução)

-8.3743427234E+00
8.2922509419E+00
-6.4734983020E+00
7.4791793889E+00
-2.5083616919E+00

Comprim. do resíduo = 1.3923310988E-04

Variancia da unidade peso a-posteriori = 1.9385858887E-09

MATRIZ COVARIÂNCIA

1.3909023107E+04 -1.9307891586E+04 1.5329458090E+04 -1.7848618743E+04 1.6214630675E+04

-1.9307891586E+04 2.6804875492E+04 -2.1278897545E+04 2.4770089112E+04 -2.2504932802E+04

1.5329458090E+04 -2.1278897545E+04 1.6895194108E+04 -1.9673416645E+04 1.7871607401E+04

-1.7848618743E+04 2.4770089112E+04 -1.9673416645E+04 2.2921269324E+04 -2.0816428075E+04

1.6214630675E+04 -2.2504932802E+04 1.7871607401E+04 -2.0816428075E+04 1.8907302042E+04

M É T O D O D A D E C O M P. D E V A L O R S I N G U L A R

Vetor W (valores singulares)

1.0000000192E+00
1.0000000362E-01
9.9999558899E-03
1.3963999608E-07
9.9989606793E-06

Vetor X (solucao)

-8.3740741624E+00
8.2918782634E+00
-6.4732022758E+00
7.4788344319E+00
-2.5080484386E+00

Comprim. do resíduo = 1.3923310924E-04

Variância da unidade peso a-posteriori = 1.9385858708E-09

MATRIZ COVARIÂNCIA

1.3908996548E+04 -1.9307854817E+04 1.5329428788E+04 -1.7848584403E+04 1.6214599577E+04
-1.9307854817E+04 2.6804824588E+04 -2.1278856979E+04 2.4770041570E+04 -2.2504889748E+04
1.5329428788E+04 -2.1278856979E+04 1.6895161780E+04 -1.9673378759E+04 1.7871573090E+04
-1.7848584403E+04 2.4770041570E+04 -1.9673378759E+04 2.2921224926E+04 -2.0816387867E+04
1.6214599577E+04 -2.2504889748E+04 1.7871573090E+04 -2.0816387867E+04 1.8907265629E+04

TESTE 07: Neste teste, Cholesky detecta a deficiência de posto na matriz A, enquanto que Gauss-Jordan e Versol fornecem uma solução errada para o problema. A SVD é usada para a determinação do posto e fornece a solução de comprimento mínimo.

MATRIZ A			vetor b
1	2	2	6
7	6	10	6
4	4	6	8
1	0	1	3

METODO DA ELIM. DE GAUSS-JORDAN

VETOR X (solucao)

-8.5714285714E-01
2.5714285714E+00
-1.4285714284E-01

Comprim. do resíduo = 5.2915026221E+00

Variancia da unidade peso a-posteriori = 2.8000000000E+01

MATRIZ COVARIANCIA

2.7487790694E+11 1.3743895346E+11 -2.7487790694E+11
1.3743895346E+11 6.8719476755E+10 -1.3743895348E+11
-2.7487790694E+11 -1.3743895348E+11 2.7487790695E+11

M E T O D O D A 'S U B R O T. V E R S O L'

VETOR X (solucao)

2.0000000000E+00
3.0000000000E+00
-2.0000000000E+00

Comprim. do resíduo = 7.0000000000E+00

Variancia da unidade peso a-posteriori = 4.9000000000E+01

MATRIZ COVARIANCIA

8.4181359004E+11 4.2090679499E+11 -8.4181359002E+11
4.2090679499E+11 2.1045339753E+11 -4.2090679501E+11
-8.4181359002E+11 -4.2090679501E+11 8.4181359002E+11

M E T O D O D E C H O L E S K Y

A matriz ATPA nao e definida positiva e a solucao foge ao alcance do metodo de CHOLESKY. Para este caso, a solucao do problema de minimos quadrados linear atraves das equacoes normais nao e confiavel.

M E T O D O Q R

Vetor X (solucao)

1.5315294006E+10
7.6576470058E+09
-1.5315294007E+10

Comprim. do resíduo = 5.2760930814E+00

Variancia da unidade peso a-posteriori = 2.7837158203E+01

MATRIZ COVARIANCIA

5.2531604756E+23 2.6265802378E+23 -5.2531604756E+23
2.6265802378E+23 1.3132901189E+23 -2.6265802378E+23
-5.2531604756E+23 -2.6265802378E+23 5.2531604755E+23

M E T O D O D A D E C O M P. D E V A L O R S I N G U L A R

Vetor W (valores singulares)

1.6207964877E+01
7.7397363334E-12
1.1409971671E+00

Num. de condicao na norma 2 = 2.0941236470E+12

Vetor X (solucao)

1.0486926357E+11
5.2434631785E+10
-1.0486926357E+11

```
Comprim. do residuo  =  6.1096747049E+00

Variancia da unidade peso a-posteriori =  3.7328125000E+01

 MATRIZ COVARIANCIA

 2.7694997206E+23  1.3847498603E+23 -2.7694997206E+23
 1.3847498603E+23  6.9237493013E+22 -1.3847498603E+23
-2.7694997206E+23 -1.3847498603E+23  2.7694997206E+23

 METODO DA DECOMP. DE VALOR SINGULAR

 Vetor W (valores singulares)

 1.6207964877E+01
 7.7397363334E-12
 1.1409971671E+00

 Num. de condicao na norma 2 =  2.0941236470E+12

 Vetor W (valores singulares)

 1.6207964877E+01
 0.0000000000E+00
 1.1409971671E+00

 Vetor X (solucao de comprimento minimo)

-1.1111111111E+00
 2.4444444445E+00
 1.1111111110E-01

Comprim. do resíduo  =  5.2915026221E+00

Variancia da unidade peso a-posteriori =  2.8000000000E+01

 MATRIZ COVARIANCIA

 6.5224171539E+00 -9.7153996100E+00  1.6647173489E+00
-9.7153996100E+00  1.4627680312E+01 -2.4015594542E+00
 1.6647173489E+00 -2.4015594542E+00  4.6393762185E-01
```

TESTE 08: Este teste é aplicado a uma matriz de posto deficiente para mostrar que nem sempre Cholesky detecta o problema, também mostra a vantagem da SVD para a análise e fornece a solução de comprimento mínimo.

```
MATRIZ A              Vetor b

  5  5                  6
  5  5                  4
  5  5                  4
```

E L I M I N. D E G A U S S - J O R D A N

Pausa 2 em GaussJ - A matriz A nao tem posto completo e a eliminacao nao pode continuar.

M E T O D O D A 'S U B R O T. V E R S O L'

Este problema nao pode ser resolvido pela 'subrotina versol', pois a matriz A nao tem posto completo.

M E T O D O D E C H O L E S K Y

Vetor X (solucao)

-6.6666666666E-02
1.0000000000E+00

Comprim. do resíduo = 1.6329931619E+00
Variancia da unidade peso a-posteriori = 2.6666666667E+00

MATRIZ COVARIANCIA

2.2906492245E+10 -2.2906492245E+10
-2.2906492245E+10 2.2906492245E+10

M E T O D O Q R

Esta matriz nao possui posto completo e nao sera possivel continuar ate a solucao.

M E T O D O D A D E C O M P. D E V A L O R S I N G U L A R

Vetor W (valores singulares)

1.4551915228E-11
1.2247448714E+01.

Num. de condicao na norma 2 = 8.4163826697E+11

Vetor X (solucao)

-7.9350416784E+10
7.9350416785E+10

Comprim. do resíduo = 2.5980762114E+00

Variancia da unidade peso a-posteriori = 6.7500000000E+00

MATRIZ COVARIANCIA

1.5937986880E+22 -1.5937986880E+22
-1.5937986880E+22 1.5937986880E+22

```
METODO DA DECOMP. DE VALOR SINGULAR

Vetor W (valores singulares)

1.4551915228E-11
1.2247448714E+01

Num. de condicao na norma 2 =  8.4163826697E+11

Vetor W (valores singulares)

0.0000000000E+00
1.2247448714E+01


Vetor X (solucao de comprimento minimo)

4.6666666667E-01
4.6666666667E-01

Comprim. do residuo  =  1.6329931619E+00

Variancia da unidade peso a-posteriori =  2.6666666667E+00

MATRIZ COVARIANCIA

8.8888888888E-03  8.8888888888E-03
8.8888888888E-03  8.8888888888E-03
```

TESTE 09: Este teste faz comparação do tempo gasto pelos métodos em se obter a solução e posteriormente para obter a matriz de covariâncias dos parâmetros. Foi colocado no intuito de comparar a velocidade de cada um dos métodos. Os tempos na obtenção dos resultados são medidos através de uma subroutina colocada no programa (subroutina VER_HORAS). Sua localização no programa principal determina a igualdade de condições na comparação da medição dos tempos. Neste teste a matriz A tem 38 linhas por 34 colunas, os tempos obtidos estão relacionados na tabela a seguir

Método	Tempo p/ obter a solução e ‖V‖	Tempo p/ obter a matriz covariância	Tempo Total
Gauss-Jordan	3.52 s	1.98 s	5.50 s
Versol	3.29 s	2.04 s	5.33 s
Cholesky	1.76 s	2.64 s	4.40 s
QR	1.97 s	2.64 s	4.61 s
SVD	12.86 s	8.02 s	20.88 s

7. PROBLEMA DE MÍNIMOS QUADRADOS NÃO-LINEAR

Até aqui foi discutido amplamente o problema de mínimos quadrados linear. Conclui-se que para manter a estabilidade na solução quando o sistema é suspeito de ser mal-condicionado deve-se utilizar a decomposição de valor singular (SVD) ou a decomposição QR. Também viu-se que na solução do problema, o método mais rápido é o método utilizando a decomposição de Cholesky.

Tais conclusões devem serem retidas na mente quando for tratado o problema de mínimos quadrados não-linear, pois este é solucionado iterativamente e em cada iteração deve ser resolvido um problema de mínimos quadrados linear.

O problema de mínimos quadrados equação (1.1) generalizado através da ponderação pode ser definido como:

$$\min_{x \in R^n} f(x) = \|V(x)\|_W^2 \qquad (7.1)$$

onde:

$f : R^n \rightarrow R$, função objetivo

$V : R^n \rightarrow R^m$, não-linear em x ;

$\|\cdot\|_W^2$: é a norma como definida em (3.40).

Uma maneira equivalente em definir o problema é:

$$\min \phi = V^T P V \qquad (7.2)$$

onde:

P : matriz dos pesos;

V : função de x (fica implicito).

Para a solução de (7.1) será proposto dois métodos tirados da progamação não-linear. O método de Gauss-Newton e o método de Levenberg-Marquardt.

O método de Gauss-Newton é o método mais básico e mais comumente usado, principalmente em Geodésia. O método de Levenberg-Marquardt porque é um método de convergência global* e utiliza o que de mais recente vem sendo pesquisado para minimização de funções não-lineares, que é a região de confiança.

A seguir será descrito cada um deles e ao final serão feitas algumas comparações quanto a convergência e número de iterações que cada leva para alcançar a solução dentro de uma determinada precisão, aplicados a um problema simples da Geodésia.

7.1 MÉTODO DE GAUSS-NEWTON

O método de Gauss-Newton utiliza um modelo afim de V em um ponto aproximado, x_0. Permita que $V(x_0)$ represente a equação residual avaliada no ponto aproximado xo. O modelo afim de V(x) é então definido por:

$$V_C(x) = V(x_0) + J(x_0)(x - x_0) \qquad (7.3)$$

onde:

$V_C(x)$: é a aproximação linear para V(x) no ponto(x_0);

$J(x_0)$: é a matriz jacobiano calculada no ponto x_0.

---

* Se entende por método de convergência global, aquele método que a partir de qualquer ponto inicial converge para um mínimo local, que pode não ser o mínimo global da função (mínimos dos mínimos).

$(x-x_0)=\Delta x$: vetor das correções aos parâmetros aproximados.

O probema de mínimos quadrados não-linear tornou-se agora um problema de mínimos quadrados linear, pois (7.3) representa um sistema de equações lineares inconsistente e o caminho agora é resolve-lo usando o critério de mínimos quadrados.

$$\min \|V_C\|_w^2 = V_C^T P\, V_C \qquad (7.4)$$

A solução para (7.4) pode ser obtida pelas decomposições QR ou SVD ou ainda por qualquer dos métodos que se utilizam das equações normais, para essas a solução para o problema torna-se:

$$\Delta x=-[J(x_0)^T P J(x_0)]^{-1} J(x_0)^T P\, V(x_0) \qquad (7.5)$$

desde que $[J(x_0)^T P J(x_0)]$ tenha inversa ordinária. x é então calculado como:

$$x=x_0+\Delta x \qquad (7.6)$$

O parâmetro x só será ponto de mínimo de f(x) se V(x) for linear, caso contrário x corresponde apenas a um valor melhorado de x_0, numa direção de decrescimento f(x) (desde de que $J(x_0)$ tenha posto completo).

Em muitas aplicações da Geodésia essa primeira iteração já fornece parâmetros considerados como bons, mas isso vai depender muito do valor aproximado inicial x_0.

Quando busca-se a solução para que a condição de otimalidade de f(x) no ponto de mínimo seja satisfeito, ou seja, $\nabla f(x^*)=0$, é necessário um processo iterativo.

Para realizar esse processo iterativo, considere x_i como sendo o valor de x numa iteração i e x_{i+1}, o valor de x na iteração consequente.

Esse passo é calculado por:

$$x_{j+1} = x_j - [J(x_j)^T P J(x_j)]^{-1} J(x_j)^T P\, V(x_j) \qquad (7.7)$$

para quando P=I, (7.7) torna-se:

$$x_{j+1}=x_j - [J(x_j)^T J(x_j)]^{-1} J(x_j)^T V(x_j) \qquad (7.8)$$

Resta saber agora, como é o comportamento desse método quanto a sua convergência. DENNIS e SCHNABEL (1983) diz isso através das seguintes vantagens e desvantagens do método, são elas:

Vantagens

1. Se o problema de mínimos quadrados é linear, o método resolve o problema com apenas uma iteração;
2. Se $V(x^*)=0$, o método de Gauss-Newton é localmente q-quadraticamente convergente;
3. Se as equações residuais V(x) são pouco não-lineares e $V(x^*)$ são pequenos, o método de Gauss-Newton é localmente rapidamente q-linearmente convergente

Desvantagens

1. Se $J(x_j)$ não tem posto completo a sequência $\{x_j\}$ não é bem definida;
2. Não tem convergência global;
3. Não converge localmente em problemas com grandes residuais ou altamente não-lineares;
4. Para problemas com residuais razoavelmente grandes ou razoavelmente não-lineares a convergência é lentamente localmente q-linearmente.

Nota: O termo grande residual é para quando a função objetivo f(x) atinge valor grande no ponto de ótimo, o inverso se dá para pequenos residuais.

Algumas dessas características serão mostradas no exemplo do final desse capítulo.

7.2. MÉTODOS DE MINIMIZAÇÃO SEM RESTRIÇÃO

Para melhor entender como funciona o método de Levenberg-Marquardt, será muito rapidamente apresentado o método de Newton, o método Steepest-Descent (ou método do gradiente) e os conceitos sobre região de confiança.

7.2.1 Método de Newton

O método de Newton para resolver o problema (1.1), utiliza-se da aproximação quadrática da série de Taylor em um ponto aproximado x_0. Isto é:

$$m(x)=f(x_0)+\nabla f(x_0)(x-x_0)+\frac{1}{2}(x-x_0)\nabla^2 f(x_0)(x-x_0) \qquad (7.9)$$

onde:

$H(x_0)=\nabla^2 f(x_0)$: matriz hessiana .

Aplicando as condições de otimalidade em (7.9) obtém-se um passo do método de Newton

$$x_{i+1}=x_i - [H(x_i)]^{-1}\nabla f(x_i) \qquad (7.10)$$

O método de Newton tem como vantagem de ser q-quadraticamente convergente próximo da solução, razão porque a grande maioria dos métodos globais de solução de equações não-lineares ou métodos de miniminização de

funções não-lineares sem restrição* procuram quando próximos da solução utilizar o passo de Newton (a demostração da convergencia quadrática do método de Newton pode ser encontrado em DENNIS e SCHNABEL (1983)).

As desvantagens do método de Newton é que a matriz hessiana **H(x)** é em geral impraticavel de ser obtida analiticamente e muito cara do ponto de vista computacional de ser avaliada por diferenças finitas.

Uma outra desvantagem do método de Newton, é que ele não é globalmente convergente.

7.2.2 Método Steepest descent ou método do gradiente

Demonstra-se que a direção $-\nabla f(x)$ é a direção de maior decrescimento da função, "direção steepest descent" (RAO 1978).

Com base nisto Cauchy em 1847, criou o método steepest descent para minimização de funções.

O algorítimo para esse método é dado abaixo:

Dado f: $\mathbf{R}^n \rightarrow \mathbf{R}$, diferenciavel continuamente em xi $\in \mathbf{R}^n$ para cada iteração i.

1. comece com x_1 , faça i=1;
2. encontre a direção $Si=-\nabla f(x_i)$;
3. encontre λ_i^* que minimiza $f(x_i+\lambda S_i)$;
4. Faça $x_{i+1}=x_i+\lambda_i^* S_i$;
5. verifique se x_i é ótimo

   se sim, faça $x_{opt}=x_{i+1}$ e pare; senão

   faça i=i+1 e volte para 2.

---

* O termo restrição é como estamos traduzindo "constraint". Em Geodésia tal termo é traduzido como injunção.

O método acima apesar de caminhar sempre na máxima direção de decrescimento da função, dependendo da topologia da função a convergência de $\{x_i\}$ pode ser extremamente lenta. Já se a função tem uma topologia regular e a função bem comportada, a sequência $\{x_i\}$ converge rapidamente para a solução.

A propriedade do método de caminhar sempre na direção de decrescimento da função é aproveitada na implementação de algoritmos de convergência global mais sofisticados.

7.2.3 Modelos aproximados pela região de confiança

Suponha que seja possível conhecer um δc no qual pode-se adequar com confiança um modelo quadrático $m(x)$ à função $f(x)$ e então, utilizando tal modelo quadrático n-dimensional, calcular a direção do passo S_c a ser dado a partir do ponto aproximado x_c e assim calcular o próximo ponto x+, restrito a $\|S_c\|_2=\delta_c$. Assim,

$$x+=x_c+S_c \qquad (7.11)$$

O problema acima pode ser enunciado como:

$$\min\ m(x_c+S) = f(x_c) + \nabla f(x_c)^T S + \frac{1}{2} S^T H_c\, S \qquad (7.12)$$

$$\text{sujeito a: } \|S\|_2 \le \delta_c$$

A seguir será dada a idéia principal para a solução do problema (7.12).

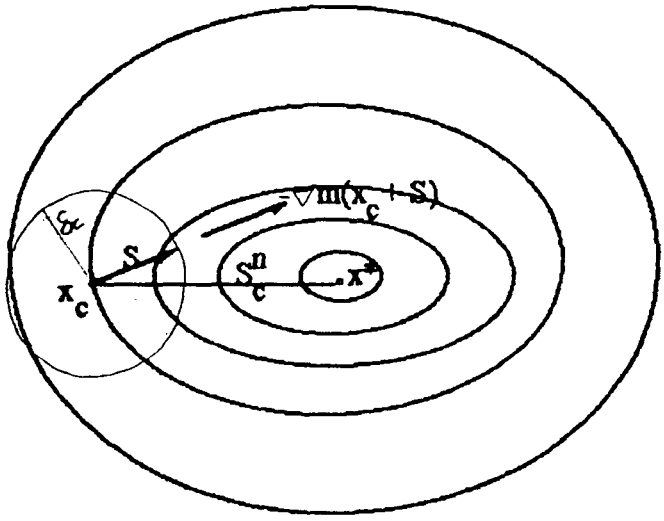
Chamando o passo de Newton de S_c^N e lembrando que ele minimiza uma função quadrática na primeria iteração, então se

$$\|S_c^N\|_2=\|- H(x_c)^{-1} \nabla f(x_c)\|_2 \ge \delta_c \qquad (7.13)$$

$x^*=x_c+S^N$, sendo x^* o ponto de mínimo da quadrática

Agora, se $\delta_c \leq \|S_c^N\|_2$, então deve-se encontrar um passo S tal que $\|S\|_2=\delta_c$.

FIGURA 15 - SOLUÇÃO DA QUADRÁTICA NUMA REGIÃO DE CONFIANÇA



Se S é solução do problema, então para qualquer distância arbitrariamente pequena a partir de x_c+S, a distância com relação a x_c deve aumentar. Se p representar uma direção de decrescimento de *m* a partir de x_c+S. Então a condição

$$p^T\nabla m(x_c+S)=p^T[H(x_c)S+\nabla f(x_c)]<0 \qquad (7.14)$$

deve ser satisfeita e ainda, para que uma direção p a partir de x_c+S aumente a distância em relação a x_c. O ângulo entre os vetores S e p deve ser menor do que $\pi/2$. daí

$$p^TS>0. \qquad (7.15)$$

Assim para S solução do problema (7.12), qualquer p que satisfaça (7.14) deve satisfazer (7.15), o que significa que a direção S é a mesma de $-\nabla m(x_c+S)$.

Agora para uma constante $\mu>0$

$\mu S=-\nabla m(x_c+S)=-[H(x_c)S+\nabla f(x_c)]$

donde

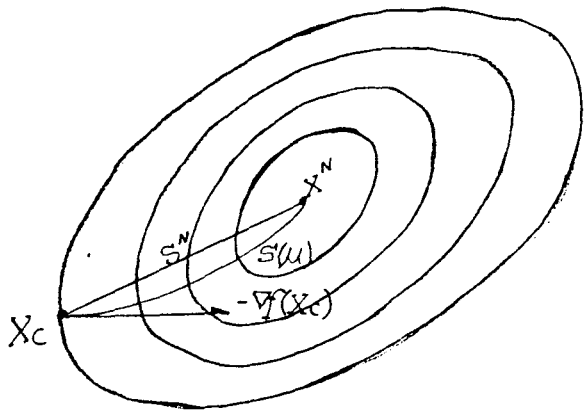
$\mu S+H(x_c)S=-\nabla f(x_c)$, resulta

$$S(\mu)=-[H(xc)+\mu I]^{-1}\nabla f(xc) \qquad (7.16)$$

que é a solução para o problema (7.12) fig. 15. Ainda pode ser demonstrado que tal solução é única.

O comportamento de $S(\mu)$ é mostrado na figura 16. Observe que, $S(\mu)$ varia da direção de Newton para a direção do gradiente conforme seja μ e o comprimento de δ_c.

FIGURA 16 - COMPORTAMENTO DE $S(\mu)$



7.2.4 Método de Levenberg-Marquardt

Considere o problema de encontrar x+ pela aproximação da região de confiança a partirde x_c

$$\min_{x^+ \in R^n} \left\| V(x_c)+J(x_c)(x^+ - x_c) \right\|_2^2 \qquad (7.17)$$

sujeito a: $\left\| x^+ - x_c \right\| \leq \delta_c$

Cuja solução (DENNIS e SCHNABEL 1983) é:

$$x+=x_C-[J(x_C)^TJ(x_C)+\mu_cI]^{-1}J(x_C)^TV(x_C) \qquad (7.18)$$

onde $\mu_c=0$ se $\delta c \geq \|(J(x_C)^TJ(x_C))^{-1}J(x_C)^TV(x_C)\|_2$ e $\mu_c>0$ caso contrário.

Quando o problema é ponderado a fórmula (7.18) torna-se:

$$x+=x_C-[J(x_C)^TP\ J(x_C)+\mu_cI]^{-1}J(x_C)^TP\ V(x_C). \qquad (7.19)$$

A utilização da fórmula (7.18) e (7.19) foi sugerida por Levenberg (1944) e Marquardat (1963), o significado do passo $S(\mu)$ já foi visto na secção anterior. O passo $S= x^+ - x_C$, varia suavemente da direção do método do gradiente "Steepest descent" quando x_C está longe da solução para o passo de Newton quando x_C está próximo da solução. Tal método é conhecido como método de Levenberg-Marquardt (também conhecido por método de marquardt).

Existem muitas estratégias para escolher μ_C e δ_C. No entanto, neste trabalho será escolhida a estratégia trazida em PRESS et alii (1986), que é mostrada pelo seguinte algoritmo.

1. adote como valor para $\mu_C=0,001$;
2. avalie $f(x_C)$ e $J(x_C)$;
3. resolva o sistema de equações lineares
   $[J(x_C)^TJ(x_C)+\mu_cI]S = - J(x_C)^TV(x_C)$ para S ; e avalie $f(x_C+S)$;
4. se $f(x_C+S)\geq f(x_C)$, aumente μ_C por um fator de 10 e volte para 3 ;

5. se $f(x_c+S)<f(x_c)$, diminua μ_c por um fator de 10, atualize a solução $x_c=x_c+S$, verifique se ela é ótima. Se sim faça $x_{opt}=x_c$, senão volte para 2.

A ordem de convergência do método de Levenberg-Marquardt, é semelhante a do método de Gauss-Newton. Para problemas grandes residuais ou problemas muito não-lineares o método de Marquardt é lentamente convergente.

Uma vantagem do método de Marquardt é que mesmo se $J(x_c)$ não tem posto completo, o método é bem definido. Outra vantagem é quando o passo de Gauss-Newton é muito longo, o passo de Marquardt é aproximadamente na direção "steepest descent"e com comprimento superior ao passo dado usando busca unidimensional. Na prática tal método segundo autores como DENNIS e SCHNABEL (1983), PRESS et alii (1986), tem tido convergência global sendo também o método atualmente recomendado por eles para resolver o problema de mínimos quadrados não-linear.

7.2.5 Critérios de parada

Existe alguns critérios especiais para o método de Levenberg-Marquardt quanto a constante μ_c e para o problema de mínimos quadrados não-linear em si, mas será colocado aqui apenas os critérios utilizados de uma maneira geral para minimização de funções não-lineares.

Os critérios mais comuns são:

1. $\|x_{i+1}-x_i\|_\infty <$ tolerância 1;

2. $\left|\frac{f(x_{i+1})-f(x_i)}{f(x_i)}\right| <$ tolerância 2;

3. $\|\nabla f(x_{i+1})\|_\infty <$ tolerância 3.

Para uma maior certeza de que realmente se encontrou o ponto de mínimo da função f(x) ou se está próximo dele o suficiente para negligenciar as demais iterações, deveria-se utilizar os três critérios acima até que os três fossem satisfeitos.

7.3. APLICAÇÃO PRÁTICA DO PROBLEMA DE MÍNIMOS QUADRADOS NÃO-LINEAR

Para a aplicação prática dos dois métodos acima será utilizado um exercício dado por GEMAEL (1974) e solucionado pelo processo iterativo em DALMOLIN (1976). Com isso, pretende-se comparar os resultados com aqueles lá encontrados.

Enunciado:

Foi medido a partir de um ponto P de coordenadas desconhecidas as distâncias l_1, l_2, l_3 e l_4 até 4 pontos de coordenadas conhecidas P1,P2,P3 e P4. Também foi medido o ângulo P1PP2. Em todas as medidas é conhecido o desvio padrão σi, ï=1,...,5. Os dados seguem abaixo:

Est.	X(m)	Y(m)	distância até o Pto P	desvio padrão
P1	842,281	925,523	244,512 m	0,012 m
P2	1337,544	996,249	321,570 m	0,016 m
P3	1831,727	723,962	773,154 m	0,038 m
P4	840,408	658,345	279,992 m	0,014 m
Est	*********	*********	ângulo	
P			123°38'01.4"	2,0"

pede-se as coordenadas de P ajustadas por mínimos quadrados.

- Equações Residuais

$$v_1=[(x_1-x_p)^2+(y_1-y_p)^2]^{1/2}-l_1;$$

$$v_2=[(x_2-x_p)^2+(y_2-y_p)^2]^{1/2}-l_2;$$

$$v_3=[(x_3-x_p)^2+(y_3-y_p)^2]^{1/2}-l_3;$$

$$v_4=[(x_4-x_p)^2+(y_4-y_p)^2]^{1/2}-l_4;$$

$$v_5 = \text{arc tg}[\frac{x_2-x_p}{y_2-y_p}]-\text{arc tg}[\frac{x_1-x_p}{y_1-y_p}]-\alpha_5$$ (v_5 deve ter como unidade segundos)

- Matriz Jacobiano

$$J(x_p) = \begin{bmatrix} (x_p-x_1)/l_{1c} & (y_p-y_1)/l_{1c} \\ (x_p-x_1)/l_{1c} & (yp-y1)/l1c \\ (x_p-x_1)/l_{1c} & (yp-y1)/l1c \\ (x_p-x_1)/l_{1c} & (yp-y1)/l1c \\ [\frac{(y_1-y_p)}{(l_{1c})^2}-\frac{(y_2-y_p)}{(l_{2c})^2}]\rho'' & [\frac{(x_2-x_p)}{(l_{2c})^2}-\frac{(x_1-x_p)}{(l_{1c})^2}]\rho'' \end{bmatrix}$$

onde: $\rho''= 1/\text{sen } 1''$ e o indíce c representa o valor calculado para as distâncias num ponto aproximado.

A matriz dos pesos P considerada diagonal, tem como elementos:

$p1=1/(0.012)^2$; $p2=1/(0.016)^2$; $p3=1/(0,038)^2$;

$p4=1/(0,014)^2$; $p5=1/(2,0)^2$

Os resultados a serem mostrados a seguir foram obtidos através de um programa computacional em linguagem " turbo pascal 6.0". Serão utilizados dois valores iniciais para testar os métodos. Primeiro o programa será executado com os valores (1065;825) e depois com os valores (825;1065), os resultados dos parâmetros encontrados estão nos quadros abaixo.

QUADRO 01 - COVERGÊNCIA DOS VALORES AJUSTADOS POR GAUSS-NEWTON

	x_0	x_1	x_2	x_3
xp	1065,00	1 065,255 4895	1 065,255 402	1 065,255 402
yp	825,00	825,185 719 1	825,185 719 41	825,185 719 1

Na 3°. iteração ocorreu a convergência usando o critério $||x_{i+1}-x_i||_\infty<1e-8$.

QUADRO 02 - COVERGÊNCIA DOS VALORES AJUSTADOS POR LEVENBERG-MARQUARDT

	x_0	x_1	x_2	x_3
xp	1065,00	1 065,255 286 7	1 065,255 402	1 065,255 402
yp	825,00	825,185 431 8	825,185 719 07	825,185 719 1
μ	0,001	0,0001	1e-05	1e-06

Na 3°. iteração ocorreu a convergência usando o critério $||x_{i+1}-x_i||_\infty<1e-8$.

QUADRO 03 - NÃO-COVERGÊNCIA DOS VALORES AJUSTADOS POR GAUSS-NEWTON

	x_0	x_1	x_2	x_{100}
xp	825,00	314,544 627 25	106,888 819 32	-13 474,495 020
yp	1065,00	1 022,170 538 9	-1 333,524 914 4	-15 897,367 574

Não ocorreu convergência com 100 iterações.

QUADRO 04 - COVERGÊNCIA DOS VALORES AJUSTADOS POR LEVENBERG-MARQUARDT

	x_0	x_1	x_5	x_8
xp	825,00	315,590 670 21	242,490 370 03	1 192,122 004 7
yp	1065,00	1 025,627 327	-144,829 613 47	144,334 959 54
μ	0,001	0,0001	0,1	0,1

	x_{10}	x_{11}	x_{12}	x_{13}
xp	988,3456	1 193,111 0811	1 085,355 378	1 065,478 926 2
yp	632,4211	846,533 281 64	803,024 112 82	825,769 004 8
μ	0,1	0,01	0,001	0,0001

	x_{14}	x_{15}	x_{16}
xp	1 065,255 767 8	1 065,255 402	1 065,255 402
yp	825,186 108 8	825,185 719 11	825,185 719 1
μ	1e-05	1e-06	1e-07

Na 16°. iteração ocorreu a convergência usando o critério $||x_{i+1}-x_i||_\infty<1e-8$ e $||\nabla f(x)||_\infty<1e-04$.

8. CONCLUSÕES E RECOMENDAÇÕES

Com base na teoria exposta e depois comprovada através dos vários testes realizados no capítulo 6 e no final do capítulo 7 é que chegamos as conclusões e recomendações a seguir.

8.1 CONCLUSÕES

Problema de mínimos quadrados linear:
método mais rápido : método de Cholesky;
método mais estável : método da decomposição de valor singular.
Problema de mínimos quadrados não-linear:
método mais eficiente : método de Levenberg-Marquardt.

Não basta apenas saber sobre certas características de determinado método com relação a outros métodos, deve-se também ter em mente o momento de aplicar um ou outro método. Nas recomendações a seguir foi colocado alguns dos procedimentos que podem ajudar na escolha da metodologia a ser adotada na solução de problemas de ajustamento.

8.2 RECOMENDAÇÕES

Quando se trata de resolver um problema de mínimos quadrados linear, e o sistema é bem condicionado, este deve ser resolvido pelo método mais rápido, o método de Cholesky.

Quando o sistema tem tendências de ser mal-condicionado, deve-se evitar de resolve-lo através das equações normais, pois estas podem deixá-lo ainda pior, nesse caso deve-se resolver o sistema pelo método QR, pois este resolve o problema usando as equações originais e é o segundo mais rápido.

Quando se deseja resolver o sistema de equações normais com mais de um vetor de termos independentes e o sistema é bem condicionado, pode-se utilizar o método de Gauss-Jordan, pois este permite a entrada com mais de um vetor de termos independentes.

Quando for necessária a inversão de uma matriz quadrada e deseja-se uma subroutina simples para fazer isso, pode ser usada a subroutina "versol". Essa subroutina mostrou através dos testes ser mais rápida que Gauss-Jordan na solução do problema de mínimos quadrados linear e com características de estabilidade semelhantes.

Para analisar se um sistema é ou não mal-condicionado através de variações na matriz dos coeficientes ou no termo independente deve-se tomar os cuidados de analisar a variação da solução e também a variação do resíduo, os testes 4 e 5 mostram esses resultados.

Quando o tempo para obter a solução não é importante, o método mais indicado é o da decomposição de valor singular, pois essa além de manter a estabilidade, fornece os valores singulares e consequentemente o número de condição da matriz dos coeficientes. O mais importante é que a análise pode ser feita através dos valores singulares. Se esses se encontrarem espaçados entre si de maneira aproximadamente iguais, o sistema tem indicativas de ser bem-condicionado, agora, se houver grandes discrepâncias entre os valores singulares, com certeza o sistema é mal-condicionado. Uma aplicação prática da SVD é que os valores singulares estão relacionados com os semi-eixos do elipsóide dos erros n-dimensional e a matriz V fornece as direções desses semi-eixos.

Um outro fator importante é que a decomposição de valor singular fornece a pseudo-inversa quando o sistema tem posto deficiente e com isso a solução de comprimento mínimo do problema de mínimos quadrados linear.

Sobre o problema de mínimos quadrados não-linear. Quando for trabalhar com o método de Gauss-Newton é recomendado resolver o problema de mínimos quadrados linear em cada iteração pelo método QR, para mais garantia de se estar numa direção de decrescimento da função, eq. (7.1). Se tal método não é disponível, recomendamos então a utilização de Cholesky. Quando a sequência convergir é util resolver mais uma vez pela SVD para uma análise mais detalhada do problema.

Quando não se tem certeza de um bom valor aproximado para iniciar a solução do problema, recomenda-se utilizar o método de Levenberg-Marquardt por causa de sua convergência global. Qualquer dos métodos anteriores podem ser usados na solução de cada iteração, a preferência fica com cholesky e QR.

Para não aceitar um ponto falso (PITFALLS) como solução, recomenda-se resolver o problema mais de uma vez com valores opostos ao obtido na primera solução e também com pontos mais afastados. Se a solução persistir a mesma, com certeza aquele ponto de mínimo é o ponto de ótimo para o problema. Se ocorrer mais de uma solução de mínimo deve-se verificar através de algum meio qual solução satisfaz o problema em questão.

Um outro fator favorável para a utilização do método de marquardt é que através dos pacotes de programas matemáticos hoje disponíveis no mercado e com linguagem computacional própria e subroutinas prontas, programar um algoritmo como o dado na secção (7.2.4) fica bastante simplificado.

Para um futuro trabalho, pretendemos aprofundar ainda mais nos métodos que se utilizam de região de confiança, trabalhar com métodos de

mínimização com restrição e também dar ao problema uma abordagem estatística.

REFERÊNCIAS BIBLIOGRÁFICAS

01 DALMOLIN, Q. Ajustamento de observações pelo processo iterativo . Curitiba, 1976. Dissertação (Mestrado em Geodésia) - Curso de Pós-Graduação em Ciências Geodésicas, UFPr.

02 DENIS, J. E ; SCHNABEL, R. B. Numerical methods for unconstrained optimization and nonlinear equations. New Jersey: Prentice Hall,Inc.,Eglewood Cliffs, 1983. 378p.

03 GASTINEL, N. Linear numerical analysis. New York: Academic Press, 1971. 350 p.

04 GEMAEL, C. Aplicações do cálculo matricial em geodésia. 2° parte: ajustamento de observações. Curso de Pós-Graduação em Ciências Geodésicas, UFPr, Curitiba. 1974.

05 _____ . Inversas generalizadas. Curso de Pós-Graduação em Ciências Geodésicas, UFPr, Curitiba, 1977. 30 p.

06 GILL, P. E ; MURRAY, W; WRIGHT, M.H. Numerical linear algebra and optimization. vol.1. California : Addison-Wesley Publishing Company, 1991. 426 p.

07 GOLUB, G.H.; REINSCH, C. Singular value decomposition and least squares solutions. Handbook series linear algebra, New York, 28 Numer Math 14.403, 403-420,1970.

08 HAMILTON,W.C. Statistics in physical science estimation, hypothesis testing, and least squares. New York: The Ronald Press Company, 1964. 225 p.

09 KRAKIWSKY, E.J. A syntesis of recent advances in the method of least squares. Canadá: Dep. of Surveying Engineering University of New Bruswick, Fredericton, 1975. 125 p.

10 LAWSON, C.L ; HANSON,R.J. Solving least squares problems. New Jersey: Prentice-Hall, Inc., Englewood Cliffs, N.J., 1974. 340 p.

11 LUENBERGER,D.G. Introdução to linear and nonlinear programming. California: Addison-Wesley Publishing company, 1973. 356 p.

12 LUGNANI, J.B. O problema dos sistemas de equações lineares mal condicionados e suas implicações em geodésia. Curitiba, 1975. Dissertação (Mestrado em Geodésia) - Curso de Pós-Graduação em Ciências Geodésicas, UFPr.

13 _____ . Introdução ao ajustamento. Curso de Pós-Graduação em Ciências Geodésicas, UFPr, Curitiba, 1983. 125 p.

14 MARSDEN, J. E ; TROMB, A. J. Vector calculus. San Francisco: W.H.Freeman and Company, 1976. 449 p.

15 MIKHAIL, E. M; ACKERMAM, F. Observations and least squares. Boston : Thomas y Crowell Company, Inc, 1976. 497 p.

16 _____ . ; GRACIE, G. Analysis and adjustment of survey mensurements. New York: Van Nostrand Reinhold Company offices, 1981. 340 p.

17 MODRO, N. Métodos para inversão de matrizes : aplicações às ciências geodésicas. Curitiba, 1981. Dissertação (Mestrado em Geodésia) - Curso de Pós-Graduação em Ciências Geodésicas, UFPr.

18 NOBLE, B.; DANIEL, J.W. Applied linear algebra. Prentice-Hall, inc, 1969. 477 p.

19 PRESS,W.H; FLANNERY,B.P; TEUKOLSKY,S.A; VETTERLING,W.T. Numerical recipes the art of scientific computing. Cambrige: Cambrige University Press, 1986. 818 p.

20 RAO, S. S. Optimization theory and applications. New York: John Wiley & Sons, 1978. 703 p.

21 STEINBRUCH, A; WINTERLE, P. Álgebra linear. 2. ed.. São Paulo: MC Graw-Hill, 1987. 583 p.

22 STRANG, G. Linear algebra and its application. 2. ed.. New York: Academic Press, Inc, 1976. 414 p.

23 TROMPSON, E. H. An introduction to the algebra of matrices with some applications. Canadá: The University of Toronto Press, 1969.