Marcia Mara Gelinski Kaiser Manfra

UMA CONTRIBUIÇÃO AO ESTUDO DA SOLUÇÃO NUMÉRICA DO PROBLEMA DE PROGRAMAÇÃO QUADRÁTICA

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia
de Produção e Sistemas
da Pontifícia Universidade Católica do Paraná
como requisito parcial para obtenção do título
de Mestre em Engenharia de Produção e Sistemas.

CURITIBA

2004

Proposta de um ....

Marcia Mara Gelinski Kaiser Manfra

Esta tese foi julgada adequada para a obtenção do título de **Mestre em Engenharia** na especialidade **Engenharia de Pordução e Sistemas**, área de concentração **Gerência de Produção e Logística**, e aprovada em sua forma final pelo curso de Pós-Graduação.

Curitiba, 20 de dezembro de 2004.

_____

Prof. Dr. Raimundo J. B. Sampaio, orientador

_____

Prof. Dr. ....
Coordenador do curso de Pós-Graduação em Engenharia de Pordução e Sistemas
da Pontifícia Universidade Católica do Paraná.

Banca Examinadora

_____ _____

Prof. Dr. Raimundo J. B. Sampaio, orientador Prof. Dr. ...., co-orientador

_____ _____

Prof. Dr. .... Prof. Dr. ....

Ao meu marido Angelo, eterno namorado e companheiro, pelo carinho compreensão, incentivo e por tudo que significa para mim. Aos meus pais, Angelo e Antonieta que sempre acreditaram em mim.

Agradecimentos

Ao professor Raimundo pela dedicação à orientação do meu trabalho;

À coordenação do curso de Mestrado pelo apoio para a realização desse trabalho;

Ao professor Claudio de Oliveira do Departamento de Ciência da Computação da PUCPR, pela colaboração em várias etapas da pesquisa, especialmente na utilização do *software* Latex;

Ao professor, Tosin do Laboratório de Física da PUCPR, pelo empréstimo de material de apoio;

Aos professores Penna e Maria Elena do Laboratório de Física da PUCPR, que tanto me incentivaram e apoiaram;

Aos professores Canova, Gomes, Suzana, Nelcy, Gil e Olimpio do Laboratório de Física da PUCPR que de uma maneira ou de outra colaboraram nas traduções, uso de *softwares* e também pelo incentivo;

Ao psicólogo Ismail Júnior pelas palavras de orientação nos momentos difíceis;

Aos colegas William Jacques e Edevilson Pereira pela colaboração e incentivo;

Aos colegas e amigos da PUCPR;

Aos meus pais, que por amor, não mediram esforços para que eu continuasse a caminhar para uma formação profissional;

Ao meu marido Angelo, pelo carinho, compreensão e incentivo;

Àqueles que de uma maneira especial, em diferentes momentos da minha vida me impulsionaram a "trilhar"o caminho acadêmico;

A Deus, por iluminar minha trajetória e por sustentar meu espírito na fé para superar os obstáculos encontrados.

Resumo

Este trabalho trata do problema de minimizar uma função quadrática definida positiva sobre a solução de um sistema de equações lineares, o problema denominado de programação quadrática definida positiva. É assumido que a matriz do sistema é retangular, de posto completo, e dessa forma o conjunto solução possui infinitos vetores. A principal preocupação foi estabelecer uma regra que a escolha de uma base para representar o espaço nulo da matriz de restrições exerce sobre o desempenho dos algoritmos de Gradiente Conjugado (CG) e Gradiente Conjugado Pré-condicionado (PGC) . Como é bem conhecido, escolhas de bases ortogonais para o espaço nulo da matriz não aumentam o número de condição da matriz reduzida enquanto a escolha de bases não ortogonais aumenta. Deste modo focalizou-se as experiências numéricas principalmente com relação as bases ortogonais. Desde que o problema de minimização com restrição inicial é substituído por um problema de minimização irrestrita sobre o espaço nulo das restrições, então os teoremas de convergência do CG e PCG ainda se aplicam, e assim concentrou-se em acompanhar como a seqüência de iterações se aproxima da solução. Executou-se o mesmo problema para diversas bases ortogonais permutando uma base ortogonal inicial obtida da decomposição QR para reforçar seu desempenho na seqüência gerada pelos algoritmos. Então, diversas conclusões foram obtidas e vários problemas abertos são apontados para pesquisas futuras.

Palavras-Chave: *Programação Quadrática, Pré-condicionamento, Base Ortogonal, Gradiente Conjugado.*

Abstract

This work deals with the problem of minimizing a definite positive quadratic function over a solution set of linear equations, the so called positive definite quadratic programming. It is supposed that the matrix of the system is retangular, full rank, and thus the solution set have infinite solution vectors. The main concern was to establish the role of the base representation to the null space of constraint matrix on the performance of the algorithms Conjugate Gradient (CG) and Preconditioned Conjugate Gradiente PCG. As it is well known, choosing ortogonal bases to the null space of matrix does not increase the condition number of the reduced matrix, while non ortogonal bases does. So we focus the numerical experiments on ortogonal bases. Since the initial constrained minimization problem is replaced to an unconstrained minimization problem over the nullity of constraints, then the convergence theorem for CG and PCG still applies, and so we concentrate following how the iterate approach to the solution. We run the same problem to several different ortogonal bases coming from permuting vectors in an initial ortogonal base obtained from QR decomposition to stress its rule in the generated sequence. Then, some conclusions are made and some open problems are pointed out.

Keywords: *Quadratic Programming, Preconditioning, Orthogonal Base , Conjugate Gradient.*

Sumário

Lista de Figuras

Lista de símbolos

Introdução

$A - V$	potenciais vetor e escalar	2
E	vetor campo elétrico	2

Revisão de literatura

$\aleph(A)$	Espaço nulo de A	6
$\Re^n$	Espaço dimensional n	9
$cond(A)$	número de condição de A	21
$F(x)$	Função objetivo	31
LU	Fatoração LU	22
P	Matriz de permutação	11
Q	Matriz ortogonal	9
QR	Decomposição QR	26
$R(A)$	Espaço coluna de A	6
$R(A^T)$	Espaço linha de A	6
V	Espaço vetorial real	5
Z	matriz cujas colunas geram $\aleph(A)$	44

Problema quadrático

$\nabla F(x)$	gradiente da função $F(x)$	39
$\nabla^2 F(x)$	Hessiana da função $F(x)$	39
A	Matriz de restrições	43
B	Matriz básica	47
b	vetor b	46
B^{-1}	Matriz inversa B	48
d	vetor direção	40
d_z	vetor direção	46
G	Matriz esparsa G	38
$g(\widehat{x})$	gradiente da função $F(x)$	46
h	vetor h	38
N	Matriz não básica	47
x	vetor x	38
x^*	vetor solução	46
x_B	Variáveis básicas	47

x_N	Variáveis não básicas	47

Métodos que usam direções conjugadas

α	tamanho do passo	53
β	parâmetro β	62
C^{-1}	Inversa da matriz C	65
CG	método de gradiente conjugado	59
L	fator de Cholesky	67
PCG	método de gradiente conjugado pré-condicionado	64

Experimentos numéricos

x_0	vetor inicial	71
x_k	vetor	72
Z^TGZ	Matriz Hessiana reduzida	71

Capítulo 1

Introdução

Este trabalho apresenta uma pesquisa sobre a influência da escolha de diferentes bases para o núcleo da matriz de restrições sobre a seqüência gerada pelos métodos de gradientes conjugados quando aplicados a um problema de minimização quadrática sujeita a restrições lineares de igualdade, ou seja

$$\underset{s/a \quad Ax=b}{\text{Minimizar}} \quad F(x) \tag{1.1}$$

onde F é definida positiva, isto é, $\nabla^2 F(x) = G$, onde G é uma matriz simétrica definida positiva. A matriz A é uma matriz $m \times n$, com $m < n$, $posto(A) = m$, onde a_i representa a i-ésima linha dessa matriz, e contém os coeficientes da i-ésima restrição linear:

$$a_i x = a_{i1} x_1 + \ldots + a_{in} x_n = b_i$$

Para GIL [Gil99] a pesquisa é um "processo formal e sistemático de desenvolvimento do método científico. O objetivo fundamental da pesquisa é descobrir respostas para problemas mediante o emprego de procedimentos científicos". Essa pesquisa é de caracter experimental, trata-se de demonstração analítica e experimentação numérica de comprovação, que significa verificação teórica e comprovação numérica. Segundo Gil [Gil91] "essencialmente a pesquisa experimental consiste em determinar um objeto de estudo, selecionar as variáveis que seriam capazes de influenciá-lo, definir as formas de controle e de observação dos efeitos que a variável produz no objeto".

O problema (1.1) é denominado de problema de programação quadrática, o qual aparece freqüentemente, ou como um problema isolado ou como sub-problema em métodos mais gerais para a solução de problemas de programação não linear (DENNIS

e SCHANBEL [DS96]).

Algumas situações onde aparecem sub-problemas a serem resolvidos a cada iteração pelos métodos de Gradiente Conjugado e Gradiente Conjugado Pré-condicionado, estão relatados abaixo,

BORGES, FALCÃO E COUTINHO [CLTB97] utilizaram o método de Gradiente Conjugado para solução do problema de fluxo de potência em redes de transmissão e distribuição de energia elétrica comparando-o com o método de Newton-Raphson, onde verificaram que o método de Gradiente Conjugado possui maior eficiência e destacam ainda que apesar da considerável parcela de tempo consumida no pré-condicionamento, o número de condicionamento da matriz de coeficientes passa a ser tão próximo à unidade que a redução no número de iterações necessárias à convergência desse método compensa o custo adicional e até acelera o método.

IGARASHI e HONNA [IH03] estudaram a convergência do método de Gradiente Conjugado Pré-condicionado para um mesmo problema de microondas formulado de maneiras diferentes. Utilizaram a formulação *Galerkin* para campos eletromagnéticos de alta freqüência em função do campo elétrico (E) e depois representaram o campo elétrico em função dos potenciais vetor e escalar (A-V). Observaram que a matriz elemento-finito do método (E) continha pequenos autovalores negativos que produz uma matriz pior condicionada mesmo após o pré-condicionamento. Por outro lado, no método (A-V) mostrou-se que os autovalores são compostos por zeros e uns normalizados após o pré-condicionamento, e que a redundância no problema (A-V) na sua formulação melhorou a convergência do método de Gradiente Conjugado. CHEN, FANG, TSANG e YUNG [RSCY00] apresentaram resultados da diferença de convergência dos métodos de Gradiente Conjugado e Gradiente Conjugado Pré-condicionado no estudo aplicado ao espalhamento de *wavelet* incidentes em estruturas metálicas gradeadas. Utilizando o método de Gradiente Conjugado Pré-condicionado *Multigrid* baseado em *wavelet* destacam o desempenho do método para ondas guiadas TE e TM (*transvers Eletric e transvers Magnetic*) incidentes em superfícies metálicas gradeadas, variando-se o período e a largura em uma faixa dessa estrutura. Para os diferentes métodos e diferentes erros residuais provou-se que o método de Gradiente Conjugado Pré-condicionado é mais eficiente que o simples método de Gradiente Conjugado, o número de iterações e o tempo de processamento computacional são menores.

Uma motivação para a realização dessa pesquisa é dada por COLEMAN e VERMA [CV01] que propõem o uso do método de Gradiente Conjugado Pré-condicionado

para solução de problema de minimização quadrática sujeita a restrições lineares de igualdade considerando a escolha de uma base fundamental ($Z_{n\times(n-m)}$) para o espaço nulo, $\aleph(A)$, da matriz de restrições,

$$Z = P\begin{pmatrix} -B^{-1}N \\ I_{n-m} \end{pmatrix} \tag{1.2}$$

onde P é uma matriz de permutação e B é não singular. A matriz A é particionada em $(B|N)$, partes básica e não básica, e I é a matriz Identidade.

COLEMAN [Col94] apresenta algumas técnicas básicas para a solução do problema (1.1), das quais a que motiva a realização desta pesquisa é a abordagem do Espaço Nulo, onde $Z_{n\times(n-m)}$ forma uma base para o espaço nulo da matriz A (matriz de restrições). A abordagem aqui utilizada é diferente daquela utilizada por GILL, *et al.* [PGS92] baseada em uma aplicação iterativa em sistemas indefinidos, bem como da abordagem utilizada por NASH e SOFER [NS93] que consiste em uma aproximação definida positiva do problema (1.1), e também diferente da técnica usada em [Col02] baseada em região de confiança.

Assumindo que x_0 é um ponto viável que satisfaz as restrições lineares, COLEMAN [Col94] aborda o método de Gradiente Conjugado Pré-condicionado para resolver o sistema

$$Z^T G Z d = -Z^T \nabla F(x_0) \tag{1.3}$$

onde G é uma matriz simétrica e definida positiva, $\nabla F(x_0)$ é o gradiente da função F em x_0, d é uma direção viável pertencente ao núcleo de A. Propõe ainda o uso da Fatoração de Cholesky para o pré-condicionamento de $Z^T G Z$.

O objetivo principal desse trabalho é estudar o efeito de diferentes escolhas de bases, $(Z_{n\times(n-m)})$, para o espaço nulo da matriz das restrições A, sobre o comportamento das seqüências geradas pelos métodos de Gradiente Conjugado e Gradiente Conjugado Pré-condicionado.

Os objetivos específicos estão relacionados com a escolha de bases ortogonais para representar o espaço nulo da matriz das restrições de modo que o número de condições do sistema (1.3) não seja aumentado pelo efeito da escolha da base. Diferentes escolhas de base serão realizadas permutando as colunas das bases obtidas.

Essas bases serão obtidas da decomposição QR da matriz A^T, (bases ortogonais) e

do particionamento da matriz A em partes [B|N] (bases não ortogonais), como já foi citado. As direções de busca usadas pelo algoritmo são as direções viáveis d, que são ortogonais às linhas de A^T, portanto $Ad = 0$, ou seja, dado $\Omega = \{x \in \Re^n : Ax = b\}$ e x_1 e $x_1 + \alpha d$ dois pontos quaisquer em Ω, então $Ad = 0$, pois se $Ax_1 = b = A(x_1 + \alpha d)$, então $\alpha Ad = 0$, $\forall \alpha \in \Re$. Portanto uma direção $d \in \Re^n, d \neq 0$, é viável em Ω se, e somente se, d está no núcleo de A, $d \in \aleph(A)$.

Este trabalho está organizado da seguinte forma, no Capítulo 2 desenvolve-se uma revisão de literatura sobre Tópicos de Álgebra Linear, onde encontram-se definições de espaços vetoriais, matriz definida, matriz de permutação, fatoração de matrizes, ortogonalização, norma de vetores e matrizes; procedendo-se também uma revisão de alguns Tópicos de Otimização, minimizações de funções, convexidade e algoritmo geral de otimização. O Capítulo 3 apresenta um estudo do problema quadrático, propriedades de funções quadráticas, a forma quadrática, condições necessárias e suficientes de otimalidade, e a abordagem adotada para a solução do problema de minimização quadrática. No Capítulo 4 desenvolveu-se exposição teórica dos métodos que usam direções conjugadas e suas propriedades, apresentando algoritmos de Gradiente Conjugado e Gradiente Conjugado Pré-condicionado. No Capítulo 5 descreve-se como os experimentos numéricos foram realizados e finalmente no Capítulo 6 faz-se as considerações finais sobre o trabalho realizado.

Capítulo 2

Revisão de Literatura

Este capítulo tem como objetivo apresentar uma revisão de literatura sobre alguns tópicos de Álgebra Linear e de Otimização.

2.1 Tópicos de Álgebra Linear

2.2 Espaços vetoriais

Um espaço vetorial real é um conjunto não vazio V munido de duas operações, adição e multiplicação por escalar, tais que:

1. $\forall\ u$ e $v \in V$, então $u + v \in V$, isto é, a operação de adição é fechada em V, e além disso:

   (a) $u + v = v + u$, $\forall\ u, v \in V$;

   (b) $u + (v + w) = (u + v) + w\ \forall\ u, v$ e $w \in V$;

   (c) Existe em V , um único vetor nulo, tal que $u + 0 = 0 + u$, $\forall\ u \in V$;

   (d) Para cada $u \in V$, existe um único vetor $-u \in V$, tal que $u + (-u) = 0$.

2. Se $u \in V$ e r um escalar, então $ru \in V$, a operação de multiplicação é fechada em V, isto é:

   (a) $r(u + v) = ru + rv$, $\forall$ escalar r e $\forall u, v \in V$;

   (b) $(r + s)u = rs + su$, $\forall$ escalar r e $\forall s, u \in V$;

(c) $r(su) = (rs)u$, $\forall$ escalar r e $\forall s, u \in V$;

(d) $1u = u$, $\forall\ u \in V$.

2.2.1 Subespaços de um espaço vetorial

Um subespaço W de um espaço vetorial V é um subconjunto não vazio contido em V que satisfaz as seguintes propriedades,

i. Se u e $v \in W$, $u + v \in W$;

ii. Seja r um escalar, e $u \in W$, $ru \in W$.

O item (i) acima declara que W é fechado em relação a operação de adição de vetores. O item (ii) estabelece que o conjunto W é fechado em relação a operação de multiplicação. Trivialmente W compartilha das propriedades de V , visto que $W \subset V$.

Todo subespaço W de V contém o vetor nulo. Todo espaço vetorial admite dois subespaços, o conjunto formado somente pelo vetor nulo e o próprio espaço vetorial.

Exemplo:

Seja a matriz $A_{m\times n}$, onde

i. O espaço coluna $R(A)$ é um subespaço de $\mathfrak{R}^m$, $R(A) \subset \mathfrak{R}^m$;

ii. O espaço linha $R(A^T)$ é um subespaço de $\mathfrak{R}^n$, $R(A^T) \subset \mathfrak{R}^n$;

iii. O espaço nulo $\aleph(A)$, é um subespaço de $\mathfrak{R}^n$, $\aleph(A) \subset \mathfrak{R}^n$;

iv. O espaço nulo $\aleph(A^T)$, é um subespaço de $\mathfrak{R}^m$, $\aleph(A^T) \subset \mathfrak{R}^m$.

2.2.2 Soma de subespaços

Se $S_1, \ldots, S_k$ são subespaços de $\mathfrak{R}^n$, então sua soma é um subespaço definido por

$$S = \{a_1 + a_2 + \ldots + a_k : a_i \in S_i,\ \ i = 1 : k\}$$

S é dito ser soma direta se cada $v \in S$ possui uma única representação

$$v = a_1 + \ldots + a_k$$

com $a_i \in S_i$. Neste caso escreve-se:

$$S = S_1 \oplus \ldots \oplus S_k.$$

2.2.3 Subespaços complementares

Sejam S_1 e S_2 subespaços do espaço vetorial V, então S_1 e S_2 são ditos complementares quando:

i. $V = S_1 \oplus S_2$;

ii. $S_1 \cap S_2 = 0$.

Neste caso, V é dito soma direta de S_1 e S_2 é representado por $V = S_1 \oplus S_2$.

2.2.4 Conjuntos geradores e dimensão

Seja S o conjunto de vetores de $\mathfrak{R}^n$, $S = \{v_1, v_2, \ldots, v_n\}$. O subspaço

$$[S] = \{v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n, v_i \in S, \alpha_i \in \mathfrak{R}\}$$

de todas as combinações lineares dos vetores de S é denominado subespaço gerado por S.

Se V é um espaço vetorial, tal que $V = [S]$, então V é um conjunto gerador de S.

A imagem (espaço range) de uma matrix $A \in \mathfrak{R}^{m \times n}$, $R(A)$ é o conjunto gerado pelas colunas de A, isto é,

$$R(A) = \{Ax : x \in \mathfrak{R}^n\} \subseteq \mathfrak{R}^m.$$

Similarmente, a imagem de A^T é o subespaço de $\mathfrak{R}^n$ definido por

$$R(A^T) = \{A^T y : y \in \mathfrak{R}^m\} \subseteq \mathfrak{R}^n.$$

Seja a matrix $A_{m \times n}$, as seguintes afirmações são verdadeiras:

i. $R(A) = [\{a_1, a_2, \ldots a_n\}]$;

ii. $R(A^T) = \left[\{a_1^T, a_2^T, \ldots a_n^T\}\right]$;

iii. $b \in R(A) \iff b = Ax$ para algum x;

iv. $a \in R(A^T) \iff a^T = x^T A$ para algum x^T.

Para uma matriz $A_{m\times n}$, o conjunto de todos m-vetores (x) tais que, $Ax = 0$, é denominado espaço nulo de A, ou seja,

$$\aleph(A) = \{x \in \Re^n : Ax = 0\} \subseteq \Re^n.$$

Seja a matriz $A_{m\times n}$, as seguintes afirmações são verdadeiras:

i. $R(A) \oplus R(A)^\perp = \Re^m$;

ii. $\forall x \in \Re^m$, x pode ser escrito de maneira única, como

$$x = x_1 + x_2, x_1 \in R(A) \text{ e } x_2 \in R(A)^\perp;$$

iii. Se $x = x_1 + x_2 \in \Re^m$ e $x = x_1' + x_2'$, então $x_1' = x_1$, $x_2' = x_2$;

iv. $\forall x \in R(A)$ e $\forall y \in R(A)^\perp, x^T y = 0$.

2.3 Produto interno

Seja V um espaço vetorial. Um produto interno em V é a função que associa a cada par de vetores u, v um número real, denotado por $\langle u, v\rangle$ ou $u^T v$, satisfazendo:

i. $\langle u, v\rangle = \langle v, u\rangle$, $\forall u, v \in V$;

ii. $\langle \alpha u + \beta v, w\rangle = \alpha \langle u, w\rangle + \beta \langle v, w\rangle$ e $\langle w, \alpha u + \beta v\rangle = \alpha \langle w, u\rangle + \beta \langle w, v\rangle$, $\forall u, v, w \in V$ e $\forall \alpha, \beta \in \Re^n$;

iii. $\langle u, u\rangle > 0$, se $u \neq 0$ e $\langle u, u\rangle = 0$, se e somente se, $u = 0$.

2.4 Ortogonalidade

2.4.1 Vetores ortogonais

Dados dois vetores x, y quaisquer do $\mathfrak{R}^n$, x e y são ditos serem ortogonais, quando $\langle x, y \rangle = 0$. Isto é,

$$x \perp y \iff x^T y = 0.$$

2.4.2 Conjuntos ortogonais

Seja $B = \{u_1, u_2, \ldots, u_n\}$ um conjunto de vetores do $\mathfrak{R}^n$. B é um conjunto ortogonal se $\forall\ i \neq j, \left\langle u_i, u_j \right\rangle = u_i^T u_j = 0$. Se em particular $\|u_i\| = 1, \forall\ i$, então B é dito ser um conjunto ortonormal. Em outras palavras:

$$\left\langle u_i, u_j \right\rangle = \begin{cases} 1 \text{ se } i = j \\ 0 \text{ se } i \neq j \end{cases}$$

Todo conjunto ortogonal é linearmente independente, e um conjunto de n vetores ortonormais de um espaço V n-dimensional é uma base ortonormal para V.

2.4.3 Complemento ortogonal de subespaços

O complemento ortogonal do subespaço $S \subseteq \mathfrak{R}^n$ é definido por:

$$S^{\perp} = \{y \in \mathfrak{R}^n : y^T x = 0, \forall x \in S\}$$

Dado qualquer subespaço $S \subseteq \mathfrak{R}^n, S \oplus S^{\perp} = \mathfrak{R}^n$, isto é,$\forall\ x \in \mathfrak{R}^n$, x é escrito de maneira única com $x = x_1 + x_2, x_1 \in S$ e $x_2 \in S^{\perp}$. Além disso, $\langle x_1, x_2 \rangle = x_1^T x_2 = 0$.

Uma coleção de subespaços $S_1, \ldots S_p$ de $\mathfrak{R}^n$ são mutuamente ortogonais se $x^T y = 0$, sempre que $x \in S_i$ e $y \in S_j$ para $i \neq j$.

2.4.4 Matriz ortogonal

Uma matriz $Q_{m \times n}$ qualquer é ortogonal se

$$Q^T Q = I_n$$

$$QQ^T = I_m$$

Para uma matriz ortogonal, as seguintes afirmações são equivalentes:

i. Q possui colunas ortonormais;

ii. Q possui linhas ortonormais;

iii. $Q^{-1} = Q^T$;

iv. $\|Qx\|_2 = \|x\|_2$ para todo $x \in \mathfrak{R}^n$.

2.4.5 Processo de ortogonalização de Gram-Schmidt

Seja W um subespaço não nulo de $\mathfrak{R}^n$ com base $S = \{u_1, u_2, \ldots, u_n\}$. Então, existe uma base ortogonal $T = \{w_1, w_2, \ldots, w_n\}$ para W obtida de S.

O processo de obtenção de W consiste de duas etapas. Na primeira obtém-se uma base ortogonal $T = \{v_1, v_2, \ldots, v_n\}$ para W, e numa segunda fase normaliza-se os vetores v_i da base T. Escolhe-se qualquer vetor em S, por exemplo u_1, e faz-se, $v_1 = u_1$. Toma-se $u_2 \in W$ e constrói-se v_2 ortogonal a v_1 do seguinte modo,

$$v_2 = u_2 - \frac{\langle u_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1$$

em seguida toma-se o vetor $u_3 \in W$ e constroi-se v_3 ortogonal a v_2 e $v_1$como,

$$v_3 = u_3 - \frac{\langle u_3, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 - \frac{\langle u_3, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1$$

e assim sucessivamente, de modo que no k-ésimo passo tem-se,

$$v_k = u_k - \sum_{i=1}^{k-1} \frac{\langle u_k, v_i \rangle}{\langle v_i, v_i \rangle} v_i$$

2.5 Matrizes

2.5.1 Posto de uma matriz

O posto linha de uma matriz A é o número de linhas não nulas dessa matriz, após a redução a forma escada. O posto linha de A é igual ao posto coluna de A, e, como são equivalentes, trabalha-se apenas com o posto de A.

Para uma matriz $A_{m\times n}$, que possui $posto(A) = r$, tem-se

i. $dimR(A) = r$,

ii. $dim\aleph(A) = n - r$,

iii. $dimR(A^T) = r$,

iv. $dim\aleph(A^T) = m - r$.

Donde então,

i. $dimR(A) + dim\aleph(A) = n$,

ii. $dim\aleph(A^T) + dimR(A^T) = m$

2.5.2 Matriz de permutação

Uma matriz de permutação (P) é justamente a matriz identidade (I) com suas linhas reordenadas. Se P é uma matriz de permutação, então pré-multiplicar A por uma matriz de permutação equivale a permutar as linhas da matriz A. Pós-multiplicar a matriz A por uma matriz de permutação equivale a permutar as colunas de A.

As matrizes de permutação são ortogonais, então

$$P^{-1} = P^T$$

Um produto de matrizes de permutação é também uma matriz de permutação.

Exemplo de matriz de permutação:

$$P = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Portanto, PA é a matriz A com as linhas 1 e 4 permutadas e AP é a matriz A com as colunas 1 e 4 permutadas.

2.5.3 Matriz esparsa

Matrizes esparsas são matrizes onde a maioria de seus elementos é nula.
Exemplo:

$$E = \begin{pmatrix} 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & -5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 \\ 8 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

2.5.4 Matriz definida

Dada uma matriz simétrica A de ordem n, esta é dita
Definida positiva se:

i. $x^T Ax > 0 \ \ \forall x \in \mathfrak{R}^n$ e $x \neq 0$;

ii. Todos autovalores de A são estritamente positivos ($\lambda_i > 0$);

Definida negativa se:

i. $x^T Ax < 0 \ \ \forall x \in \mathfrak{R}^n$ e $x \neq 0$;

ii. Todos autovalores de A são estritamente negativos ($\lambda_i < 0$);

Semidefinida positiva se:

i. $x^T Ax \geq 0 \ \ \forall x \in \mathfrak{R}^n$;

ii. Todos autovalores de A são não negativos ($\lambda_i \geq 0$);

Semidefinida negativa:

i. $x^T Ax \leq 0 \ \ \forall x \in \mathfrak{R}^n$;

ii. Todos autovalores de A são não positivos ($\lambda_i \leq 0$).

2.5.5 Matriz de rotação

Uma matriz ortogonal $Q_{2\times2}$ é uma matriz de rotação se ela possui a forma

$$Q = \begin{pmatrix} \cos(\theta) & sen(\theta) \\ -sen(\theta) & \cos(\theta) \end{pmatrix}$$

Se $y = Q^T x$, então y é obtido pela rotação de x na direção horária por um ângulo θ.

2.5.6 Matriz de reflexão

Uma matriz ortogonal $Q_{2\times2}$ é uma matriz de reflexão se, para algum, θ possui a forma

$$Q = \begin{pmatrix} \cos(\theta) & sen(\theta) \\ sen(\theta) & -\cos(\theta) \end{pmatrix}$$

Se $y = Q^T x = Qx$, então y é obtido pela reflexão do vetor x através da linha definida por

$$S = span\left\{ \begin{pmatrix} \cos(\frac{\theta}{2}) \\ sen(\frac{\theta}{2}) \end{pmatrix} \right\}$$

As reflexões e rotações são atraentes computacionalmente porque são facilmente construídas e também são usadas para introduzir zeros em um vetor por escolher apropriadamente ângulos de rotação e planos de reflexão.

Exemplo [GL96]:

Supondo $x = \begin{pmatrix} 1 & \sqrt{3} \end{pmatrix}^T$. Se

$$Q = \begin{pmatrix} \cos(-60^o) & sen(-60^o) \\ sen(-60^o) & -\cos(-60^o) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{-\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}$$

Então $Q^T x = \begin{pmatrix} 2 & 0 \end{pmatrix}^T$. Assim, fazendo uma rotação de -60^o , zera-se a primeira componente.

$$\text{Se} \quad Q = \begin{pmatrix} \cos(30^o) & sen(30^o) \\ sen(30^o) & -\cos(30^o) \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix}$$

Então $Q^T x = \begin{pmatrix} 2 & 0 \end{pmatrix}^T$. Assim, fazendo uma reflexão de um ângulo de 30^o do vetor

x , zera-se a segunda componente.

2.5.7 Matriz de Householder

Seja w um vetor não nulo, $w \in \mathbb{R}^n e H_w$ a matriz definida por

$$H_w = I - \left(\frac{2}{w^T w}\right) w w^T \tag{2.1}$$

é denominada de matriz de Householder, onde

H_w é uma matriz simétrica e ortogonal. De fato

i.

$$H_w^T = \left[I - \left(\frac{2}{w^T w}\right) w w^T\right]^T = \left[I - \left(\frac{2}{w^T w}\right) w w^T\right] = H_w$$

ii.

$$H_w^T H_w = \left[I - \left(\frac{2}{w^T w}\right) w w^T\right]^T \left[I - \left(\frac{2}{w^T w}\right) w w^T\right]$$

$$\begin{aligned} &= \left[I - \left(\frac{2}{w^T w}\right) w w^T\right]\left[I - \left(\frac{2}{w^T w}\right) w w^T\right] \\ &= I - \left(\frac{4}{w^T w}\right) w w^T + \left(\frac{2}{w^T w}\right)\left(\frac{2}{w^T w}\right) w w^T w w^T \\ &= I - \left(\frac{4}{w^T w}\right) w w^T + \left(\frac{4}{w^T w}\right) w w^T \\ &= I \end{aligned}$$

As reflexões de Householder são modificações de posto 1 da matriz identidade e podem ser usadas para especificar componentes nulas de um vetor. Ao multiplicar a expressão (2.1) por um vetor não nulo $x \in \Re^n$, de modo que Hx seja um múltiplo de e_1, tem-se

$$H_w x = \left(I - \frac{2 w w^T}{w^T w}\right) x = x - \frac{2 w^T x}{w^T w} w$$

e $H_w x \in span\{e_1\}$ implica $w \in span\{x, e_1\}$. Onde $w = x + \alpha e_1$ fornece que

$$w^T x = x^T x + \alpha x_1$$

e

$$w^T w = x^T x + 2\alpha x_1 + \alpha^2$$

portanto

$$H_w x = \left(1 - 2\frac{x^T x + \alpha x_1}{x^T x + 2\alpha x_1 + \alpha^2}\right) x - 2\alpha \frac{w^T x}{w^T w} e_1$$

Para que as componentes de x sejam zero, faz-se $\alpha = \pm \|x\|_2$, então

$$w = x \pm \|x\|_2 e_1 \Rightarrow H_w x = \left(I - 2\frac{ww^T}{w^T w}\right) x = \pm \|x\|_2 e_1$$

Exemplo [GL96]:
Seja $x = \begin{pmatrix} 3 & 1 & 5 & 1 \end{pmatrix}^T$ e $w = \begin{pmatrix} 9 & 1 & 5 & 1 \end{pmatrix}^T$ então

$$H_w = I - 2\frac{ww^T}{w^T w} = \frac{1}{54}\begin{pmatrix} -27 & -9 & -45 & -9 \\ -9 & 53 & -5 & -1 \\ -45 & -5 & 29 & -5 \\ -9 & -1 & -5 & 53 \end{pmatrix} \quad \text{logo}$$

$$H_w x = \begin{pmatrix} -6 & 0 & 0 & 0 \end{pmatrix}^T.$$

2.5.8 Rotação de Givens

Em cálculos onde é necessário selecionar um número maior de elementos zeros, rotações de Givens são mais adequadas. Estas matrizes são correções de posto 2 da matriz identidade do tipo

$$G(i,k,\theta) = \begin{pmatrix} & & i & & k & & & \\ 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \\ \vdots & \ddots & \vdots & & \vdots & & \vdots & \\ 0 & \cdots & c & \cdots & s & \cdots & 0 & i \\ \vdots & & \vdots & \ddots & \vdots & \cdots & \vdots & \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 & k \\ \vdots & & \vdots & & \vdots & \ddots & \vdots & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & \end{pmatrix}$$

onde $c = cos(\theta)$ e $s = sen(\theta)$ para algum θ. As rotações de Givens são ortogonais.

Pré-multiplicações de vetores por $G(i,k,\theta)^T$ equivale a uma rotação anti-horária de θ radianos em um plano de coordenada (i,k).

De fato, se $x \in \mathfrak{R}^n$ e $y = G(i,k,\theta)^T x$, então

$$y_j = \begin{cases} cx_i - sx_k & j = i \\ sx_i - cx_k & j = k \\ x_j & j \neq i,k \end{cases}$$

Vem desta formulação, que é possível forçar y_k ser nulo por,

$$c = \frac{x_i}{\sqrt{x_i^2 + x_k^2}} \qquad s = \frac{-x_k}{\sqrt{x_i^2 + x_k^2}}$$

Assim, o uso da rotação de Givens é uma maneira simples para introduzir zeros em uma posição específica de um vetor.

2.5.9 Matriz diagonalizável

Uma matriz B é dita semelhante a uma matriz A se existe uma matriz inversível X tal que:

$$B = X^{-1}AX$$

Se B é diagonal então A é semelhante a uma matriz diagonal, isto é A é diagonalizável, ou seja

$$X^{-1}AX = diag(\lambda_1, \ldots, \lambda_n)$$

A é diagonalizável, se e somente se, existem vetores $x_1, \ldots, x_n \in \mathfrak{R}^n$ e escalares $\lambda_1, \ldots, \lambda_n$ tais que

$$Ax_i = \lambda x_i, \quad i = 1 : n.$$

Isto é equivalente a existência de uma matriz não singular $X = [x_1, \ldots, x_n] \in \mathfrak{R}^n$ tal que $AX = XD$, onde $D = diag(\lambda_1, \ldots, \lambda_n)$.

2.5.10 Norma de vetores

Uma norma de um vetor x em $\Re^n$ é uma função $f : \Re^n \to \Re$, representada por $f(x) = ||x||$, onde:

i. $||x|| \geq 0 \qquad \forall\, x \in \Re^n$ e $||x|| = 0 \Longleftrightarrow x = 0$,

ii. $||\alpha x|| = |\alpha|\,||x|| \qquad \forall \alpha \in \Re$ e $x \in \Re^n$,

iii. $\left\|x + y\right\| \leq ||x|| + \left\|y\right\| \qquad \forall\, x, y \in \Re^n$.

A norma-p é definida por:

$$||x_p|| = (||x_1||^p + \ldots + ||x_n||^p)^{\frac{1}{p}} \quad 1 \leq p < \infty$$

Na prática, as normas mais usadas são respectivamente:
a norma-1:

$$||x||_1 = |x_1| + \ldots + |x_n| \ ,$$

a norma-2:

$$||x||_2 = \left(|x_1|^2 + \ldots + |x_n|^2\right)^{\frac{1}{2}} = \left(x^T x\right)^{\frac{1}{2}} ,$$

e a norma-∞ :

$$||x||_\infty = \lim_{p\to\infty} \ (||x_1||^p + \ldots + ||x_n||^p)^{\frac{1}{p}} = \max_i |x_i| \ \ ,$$

que são a norma p para $p = 1$, $p = 2$ e $p = \infty$.
Em $\Re^n$ duas normas quaisquer são equivalentes, isto é, existem constantes C_1, C_2 e C_3, tais que

$$C_1\,||x||_1 \leq C_2\,||x||_2 \leq C_3\,||x||_3$$

2.5.11 Erros absoluto e relativo

Seja $\widehat{x} \in \mathbb{R}^n$ uma aproximação de $x \in \mathbb{R}^n$. Então o erro absoluto que ocorre ao se representar x por $\widehat{x}$ é dado por:

$$\varepsilon_{abs} = \left\|\widehat{x} - x\right\|$$

e, se $x \neq 0$, o erro relativo correspondente será:

$$\varepsilon_{rel} = \frac{\left\|\widehat{x} - x\right\|}{||x||}.$$

O erro relativo na norma-∞ pode ser traduzido em uma relação sobre o número de dígitos significativos corretos em $\widehat{x}$. Em particular , se

$$\frac{\left\|\widehat{x} - x\right\|_\infty}{||x||_\infty} \approx 10^{-p}$$

então a maior componente de $\widehat{x}$ tem aproximadamente p dígitos significativos corretos.

2.5.12 Norma de matrizes

Seja ||.|| uma norma vetorial no espaço correspondente e uma matriz $A_{m\times n}$, a norma da matriz induzida pela norma de vetor, é definida por

$$||A||_{\mathbb{R}^n,\mathbb{R}^m} = \sup_{x\neq 0}\left\{\frac{||A(x)||_{\mathbb{R}^m}}{||x||_{\mathbb{R}^n}}\right\}$$

Usando-se as normas vetoriais definidas anteriormente, tem-se:
a norma-1:

$$||A||_1 = \max_{x\neq 0}\left\{\frac{||Ax||_1}{||x||_1}\right\}$$

a norma-2:

$$||A||_2 = \max_{x\neq 0}\left\{\frac{||Ax||_2}{||x||_2}\right\}$$

a norma-∞:

$$||A||_\infty = \max_{x\neq 0}\left\{\frac{||Ax||_\infty}{||x||_\infty}\right\}$$

A norma-2 de uma matriz A pode ser obtida extraindo a raiz quadrada do maior autovalor de A^TA,

$$||A||_2^2 = \lambda_{\max}(A^TA)$$

Quando A é simétrica, então $A^TA = A^2$ e a norma é o maior autovalor em módulo de A,

$$||A|| = \max|\lambda_i|$$

Seja x o autovetor de A^TA associado a seu maior autovalor $\lambda_{\max}$. Então,

$$||A||_2^2 = \max_{x\neq 0}\left\{\frac{||Ax||_2^2}{||x||_2^2}\right\} = \max_{x\neq 0}\left\{\frac{x^TA^TAx}{x^Tx}\right\} = \frac{x^T(\lambda_{\max})x}{x^Tx} = \lambda_{\max},$$

O quociente

$$\frac{x^TA^TAx}{x^Tx}$$

é conhecido como *quociente de Rayleigh.*

2.5.13 Número de condição de uma matriz

Dado o sistema linear $Ax = b$, onde A é inversível. Supondo que o lado direito é perturbado por $b + \delta b$ e que a solução exata do sistema é $x + \delta x$, tem-se:

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

onde "δ" é uma mudança pequena no vetor ou na matriz. Então,

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

desde que $x = A^{-1}b$,

$$\delta x = A^{-1}\delta b$$

Para medir δx recorre-se ao uso da norma de matriz induzida pela norma de vetor, onde

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

A perturbação na solução exata é limitada por $\left\|A^{-1}\right\|$ vezes a perturbação do lado direito $||\delta b||$.

Para determinar o efeito relativo dessas perturbações, nota-se que:

$$||b|| \leq ||A|| \, ||x||$$

E combinando as duas equações anteriores, tem-se:

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

Se a matriz A do sistema $Ax = b$ é perturbada por δA,atribui-se um procedimento similar para

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

A equação pode ser reescrita como

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

por esta razão

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

ou

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

Quando a mudança $||\delta A||$ é considerada relativa para $||A||$ tem-se

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

A mudança relativa na solução exata é limitada pelo fator $||A|| \left\|A^{-1}\right\|$ multiplicado pela perturbação relativa $\frac{||\delta b||}{||b||}$ ou $\frac{||\delta A||}{||A||}$.O número $||A|| \left\|A^{-1}\right\|$ é definido como número de condição de A e é representado por $cond(A)$.

Uma vez que $||I|| = 1$ em uma norma, e $I = AA^{-1}$, segue que

$$1 = \|I\| \leq \|A\| \left\|A^{-1}\right\|$$

assim

$$cond(A) \geq 1$$

O número de condição de uma matriz A indica o máximo efeito da perturbação sobre a solução exata de $Ax = b$. Se a matriz A é mal-condicionada o número de condição de A, $cond(A)$, é grande. Se a matriz A é bem-condicionada o número de condição de A, $cond(A)$, é pequeno.

2.6 Fatoração de matrizes

A fatoração ou decomposição de uma matriz é usualmente representada pelo produto ou soma de matrizes que possui uma forma especial.

Os métodos de fatoração tradicionais para sistemas lineares envolvem a conversão de um dado sistema quadrado para um sistema triangular que possuem alguma solução.

Considerando o sistema triangular inferior, $Lx = b$:

$$\begin{pmatrix} l_{11} & 0 \\ \\ l_{21} & l_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ \\ b_2 \end{pmatrix}$$

Se $l_{11}l_{22} \neq 0$, ou seja, o determinante da matriz triangular inferior L for diferente de zero, tem-se uma matriz não singular, e as incógnitas do sistema representado podem ser determinadas seqüencialmente, como segue

$$\begin{aligned} x_1 &= \frac{b_1}{l_{11}} \\ x_2 &= \frac{(b_2 - l_{21}x_1)}{l_{22}} \end{aligned}$$

Esta é uma versão 2×2 de um algoritmo conhecido como substituição para frente.

O procedimento geral é obtido resolvendo a i-ésima equação em $Lx = b$ por x_i:

$$x_i = \frac{\left(b_i - \sum_{j=1}^{i-1} l_{ij} x_j\right)}{l_{ii}}$$

Um algoritmo análogo para sistemas triangulares superiores, $Ux = b$ é denominado de substituição para trás. Pode-se obter x_i, da seguinte forma:

$$x_i = \frac{\left(b_i - \sum_{j=i+1}^{n} u_{ij} x_j\right)}{u_{ii}}$$

2.6.1 Fatoração LU

Supondo que uma matriz $A_{m \times n}$, pode ser fatorada como

$$A = LU$$

onde L é uma matriz triangular inferior com diagonal principal unitária e U é uma matriz triangular superior, esta fatoração pode ser usada para resolver sistemas do tipo $Ax = b$. Substituindo A por LU, obtém-se

$$(LU)x = b$$

ou ainda,

$$L(Ux) = b$$

Fazendo $Ux = y$, a equação matricial anterior fica

$$Ly = b$$

Como L é uma matriz triangular inferior, resolve-se diretamente para y . Uma vez determinado y, como U é triangular superior, resolve-se $Ux = y$ que fornecerá o vetor procurado x. Segue o algoritmo da Fatoração LU.

Algoritmo da Fatoração $A = LU$ [Cun00].

Dada a matriz $A = [a_{ij}]$

Para $i = 1 : n$,

para $j = i : n$,

$u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik}u_{kj}$

fim

para $j = i + 1 : n$,

$l_{ji} = \frac{\left(a_{ji} - \sum_{k=1}^{i-1} l_{jk}u_{ki}\right)}{u_{ii}}$

fim

Fim.

Isto é possível quando não é necessária nenhuma troca de linhas, o que ocorre quando existe um pivô nulo. Uma simples troca de linha define uma matriz de permutação P.

Quando existe a necessidade de permutar linhas ou colunas fazemos uso das matrizes de permutação e da estratégia de pivotamento parcial associada a fatoração LU.

A estratégia de pivotamento parcial consiste em escolher como pivô o maior elemento em módulo, da coluna onde os elementos serão eliminados, o que fará com que elimine o problema da existência de um pivô nulo.

O sistema $Ax = b$ será equivalente a

$$PAx = LUx = Pb.$$

Resolve-se então o sistema triangular $Ly = Pb$ e depois $Ux = y$ para x. Segue o algoritmo da fatoração LU com pivotamento parcial.

Algoritmo da fatoração $PA = LU$, com pivotamento parcial adaptado de [RdRL96].

Resolução de um sistema $Ax = b$, onde p representará as permutações realizadas durante a fatoração.

Cálculo dos fatores

Para $i = 1 : n$

$p(i) = i$

para $k = 1 : (n - 1)$

$pv = |a(k, k)|$

$r = k$

para $i = (k + 1) : n$

se $(|a(i, k)| > pv)$, faça:

$pv = |a(i, k)|$

$r = i$

fim
fim
se $pv = 0$, parar: a matriz é singular
fim
se $r \neq k$, faça:
$aux = p(k)$
$p(k) = p(r)$
$p(r) = aux$
para $j = 1 : n$
$aux = a(k, j)$
$a(k, j) = a(r, j)$
$a(r, j) = aux$
fim
fim
para $i = (k + 1) : n$
$l = \frac{a(i,k)}{a(k,k)}$
$a(i, k) = l$
para $j = (k + 1) : n$
$a(i, j) = a(i, j) - la(k, j)$
fim
fim
fim
Fim.
Resoluções dos sistemas triangulares
Resolução do sistema $c = Pb$
Para $i = 1 : n$
$r = p(i)$
$c(i) = b(r)$
Fim
Resolução do sistema $Ly = c$
Para $i = 1 : n$
$soma = 0$
para $j = 1 : (i - 1)$
$soma = soma + a(i, j)y(j)$

fim
$y(i) = c(i) - soma$
Fim
Resolução so sistema $Ux = y$
Para $i = n, (n-1) : 1$
$soma = 0$
para $j = (i+1) : 1$
$soma = soma + a(i,j)x(j)$
fim
$x(i) = \frac{(y(i) - soma)}{a(i,i)}$
Fim

2.6.2 Fatoração de Cholesky

Um caso especial da fatoração $A = LU$, ocorre quando A é uma matriz simétrica ($A = A^T$) , definida positiva ($x^T Ax > 0, \forall x \neq 0$). Nesse caso a matriz U da decomposição $A = LU$ é $U = DL^T$. Então tem-se a decomposição de A como:

$$A = LDL^T$$

onde L é uma matriz triangular inferior, e D é uma matriz diagonal com elementos estritamente positivos. Então pode-se escrever que,

$$A = LD^{\frac{1}{2}} D^{\frac{1}{2}} L^T = \tilde{L}\tilde{L}^T = R^T R$$

onde $\tilde{L}$ é uma matriz triangular inferior genérica, e R é uma matriz triangular superior genérica e R é chamado fator de Cholesky de A.

Dada uma matriz simétrica definida positiva A, o algoritmo que segue para a Decomposição de Cholesky calcula uma matriz triangular inferior $\tilde{L}$, tal que $A = \tilde{L}\tilde{L}^T$. Para todo $i \geq j, \tilde{L}\ (i,j)$ sobreescreve $A(i,j)$.

Algoritmo de Cholesky [GL96].

Algoritmo de Cholesky [GL96].
Para $k = 1 : n$
$a(k,k) = \sqrt{a(k,k)}$

$a(k+1:n,k) = \frac{a(k+1:n,k)}{a(k,k)}$

para $j = k+1:n$

$a(j:n,1) = a(j:n,1) - a(j:n,k)a(j,k)$

fim

Fim

2.6.3 Fatoração QR

Toda matriz $A_{m\times n}$, com $m > n$, pode ser fatorada como $A = QR$, onde $Q \in \Re^{m\times m}$ é ortogonal e $R \in \Re^{m\times n}$é uma matriz triangular superior. Se A possui posto completo, as n primeiras colunas de Q formam uma base ortonormal para $R(A)$ e as últimas colunas de Q formam uma base ortogonal para $R(A)^{\perp}$.

Teorema 2.6.1 – *Se $A = QR$ é uma fatoração QR de uma matrix $A \in \Re^{m\times n}$ de posto coluna completo, e $A = [a_1, .., a_n]$, $Q = [q_1, \ldots, q_m]$ onde a_i, $i = 1, \ldots, n$ e q_j, $j = 1, \ldots, m$, são respectivamente as colunas da matriz A e Q, então*

$$span\{a_1, .., a_k\} = span\{q_1, \ldots, q_k\} \qquad k = 1:n$$

Em particular, se $Q_1 = Q_{1:m\times 1:n}$ e $Q_2 = Q_{1:m\times n+1:m}$ então

$$R(A) = R(Q_1)$$

$$R(A)^{\perp} = R(Q_2)$$

e $A = Q_1R_1$ com $R_1 = R_{1:n\times 1:n}$. ■

PROVA: [GL96] Comparando a k-ésima coluna de $A = QR$ conclui-se que

$$a_k = \sum_{i=1}^{k} r_{ik}q_i \in span\{q_1, \ldots, q_k\}$$

Assim, $span\{a_1, .., a_k\} \subseteq span\{q_1, \ldots, q_k\}$. Contudo, desde que o posto de A seja igual a n , segue que o $span\{a_1, .., a_k\}$ possui dimensão k e deste modo deve ser igual ao $span\{q_1, \ldots, q_k\}$. O restante do teorema segue trivialmente.

2.6.4 Decomposição de valor singular

Qualquer matriz $A_{n\times n}$ pode ser escrita como

$$A = USV^T,$$

onde $U_{m\times m}$ é uma matriz ortogonal, $V_{n\times n}$ é uma matriz ortogonal e $S_{m\times n}$ é uma matriz diagonal,

$$S = diag\left(\sigma_1, \ldots \sigma_n\right),$$

com $p = \min(m, n)$ e $\sigma_i \geq 0$, $i = 1, \ldots, p$.

Os números não negativos $\{\sigma_i\}$ são denominados de valores singulares de A, e $A = USV^T$ é denominada de decomposição de valor singular (SVD).

Usualmente adota-se a convenção $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p$. Deste modo $\sigma_1(A)$ denota o maior valor singular de A.

Se a matriz A possui posto r e $r > 0$, então A possui exatamente r valores singulares estritamente positivos, de modo que $\sigma_r > 0$ e $\sigma_{r+1} = \ldots = \sigma_p = 0$.

No caso da matriz A possuir posto completo, todo valor singular é não nulo. Os valores singulares são raízes quadradas dos autovalores de A^TA (se $m \geq n$) ou de AA^T (se $m < n$). Se a matriz A é simétrica, os valores singulares são os valores absolutos dos autovalores.

2.7 Sistemas lineares

Dada uma matriz $A_{m\times n}$ e um vetor $b \in \Re^m$, considera-se o problema de determinar um vetor $x \in \Re^n$, tal que:

$$Ax = b$$

Para que o sistema possua solução o vetor b deve estar contido no subespaço gerado pelas colunas de A. Se $b \subseteq R(A)$ o sistema é dito ser compatível, isto é, b está no espaço gerado pelas colunas de A, b é uma combinação linear das colunas da matriz A.

Se $b \nsubseteq R(A)$, então o sistema é dito ser incompatível (ou inconsistente) e não existe combinação das colunas de A que gere b.

Se as colunas de uma matriz A são linearmente dependentes, e $b \subset R(A)$, então

existem infinitas soluções para $Ax = b$ e, dessa forma, a representação de b como uma combinação linear das colunas da matriz A não é única.

No caso de existirem infinitas soluções para o sistema linear $Ax = b$, se z e y são soluções desse sistema tem-se que,

$$Az = b \text{ e } Ay = b,$$

então

$$A(z - y) = Az - Ay = b - b = 0.$$

Seja $w = z - y$, a diferença das duas soluções so sistema linear, então w deve resolver o sistema homogêneo

$$Ax = 0,$$

Se w é alguma solução para $Ax = 0$ e y é alguma solução para o sistema $Ax = b$, então $z = y + w$ também resolve o sistema $Az = b$.

Pode-se dizer que toda solução para um sistema linear $Ax = b$ é a soma de uma solução particular, x_0, e uma solução do sistema homogêneo $Ax = 0$, ou seja,

$$x = x_{\text{particular}} + x_{\text{homogêneo}},$$

isto é, o conjunto solução S é então

$$S = x_0 + \aleph(A).$$

2.8 Autovetores e autovalores

Para qualquer matriz $A_{n\times n}$, existe no mínimo um número $\lambda \epsilon \Re$ e um vetor u, não nulo associado tal que

$$(A - \lambda I)u = 0, \tag{2.2}$$

ou, equivalentemente,

$$Au = \lambda u \tag{2.3}$$

Esta primeira equação (2.2) indica que subtraindo λ de cada elemento da diagonal da matriz A obtém-se uma matriz singular e a segunda equação equivalente (2.3) mostra que pré-multiplicar u por A não altera a direção de u. O valor λ é denominado autovalor de A, e o correspondente vetor u é denominado um autovetor de A.

Qualquer matriz $A_{n\times n}$ possui n autovalores $\{\lambda_1, \ldots \lambda_n\}$ não necessariamente distintos, os quais são as raízes n-ésimas da equação polinomial de grau n

$$\det(A - \lambda I) = 0.$$

A soma dos elementos da diagonal de qualquer matriz quadrada A (denominada de traço de A) é igual a soma dos autovalores, ou seja,

$$traço(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i(A).$$

O produto dos autovalores é igual ao determinante de A:

$$\Pi_{i=1}^{n} \lambda_i(A) = \det(A).$$

Multiplicando $Au = \lambda u$ por uma matriz não singular S, tem-se

$$SAu = S(\lambda u) = \lambda(Su).$$

Desde que $S^{-1}S = I$, segue que,

$$SAu = SAS^{-1}Su = SAS^{-1}(Su) = \lambda(Su),$$

o que mostra que λ é um autovalor e Su é um autovetor de SAS^{-1}.

Se a matriz A é singular, existe um vetor não nulo x tal que $Ax = 0$, o que mostra que uma matriz singular possui no mínimo um autovalor zero.

Caso a matriz A seja não singular, todos os autovalores são não nulos e os autovalores da A^{-1} são os recíprocos autovalores de A.

Se A é uma matriz simétrica, todos os autovalores de A são reais. Uma matriz simétrica sempre possui autovetores ortogonais que podem ser normalizados para formar uma base ortonormal para o $\mathfrak{R}^n$.

2.9 Projeções

Seja um subespaço $S \subseteq \Re^n$. P é uma projeção ortogonal sobre S, se

i. $R(P) = S$,

ii. $P^2 = P$,

iii. $P^T = P$.

Se $x \in \Re^n$, então $Px \in S$ e $(I - P)x \in S^\perp$.

Se P_1 e P_2 são projeções ortogonais, então para qualquer $z \in \Re^n$, com $z \neq 0$, tem-se que

$$\|(P_1 - P_2)\, z\|_2^2 = (P_1 z)^T (I - P_2)\, z + (P_2 z)^T (I - P_1) z$$

Se $R(P_1) = R(P_2) = S$, então o lado direito dessa equação é zero, mostrando que a projeção ortogonal de um vetor para um subespaço é único.

Se as colunas de $V = [v_1, \ldots v_k]$ são uma base ortogonal para o subespaço S, então $P = VV^T$ é a projeção ortogonal sobre S. Nota-se que se $v \in \Re^n$, então

$$P = \frac{vv^T}{v^T v}$$

é a projeção ortogonal sobre $S = span\,\{v\}$.

Seja V um espaço vetorial e V_0 um subespaço de V gerado por um conjunto ortogonal de vetores não nulos

$$S = \{v_1, v_2, \ldots, v_q\}\,.$$

Define-se projeção ortogonal P_0 sobre V_0 como segue:

$$P_0 v = \alpha_1 v_1 + \ldots + \alpha_q v_q,$$

onde

$$\alpha_i = \frac{\langle v_i, v \rangle}{\langle v_i, v_i \rangle}\,,$$

Então

i. $v - P_0 v$ é ortogonal a todo vetor $v_0 \in V_o$;

ii. $P_0(u + v) = P_0 u + P_0 v$, $\forall\ u, v \in V$;

iii. $P_0(\alpha v) = \alpha P_0 v$,$\forall$ escalar α, $\forall\ v \in V$.

Nota-se em (i) que $v - P_0 v$ é ortogonal a cada v_i:

$$\langle v_i, v - P_0 v\rangle = \langle v_i, v\rangle - \alpha_1 \langle v_i, v_1\rangle - \ldots - \alpha_q \left\langle v_i, v_q\right\rangle = \langle v_i, v\rangle - \alpha_i \langle v_i, v_i\rangle = 0.$$

Desde que cada $v_o \in V_o$ seja uma combinação linear de v_i, satisfazendo $\langle v_i, v - P_0 v\rangle = 0$, obtém-se $\langle v_0, v - P_0 v\rangle = 0$.

(ii.) e (iii.) seguem imediatamente da definição dos coeficientes α_i. Por exemplo, o coeficiente $\alpha_i \in P_0(\alpha v)$ é

$$\frac{\langle v_i, \alpha v\rangle}{\langle v_i, v_i\rangle} = \alpha \frac{\langle v_i, v\rangle}{\langle v_i, v_i\rangle},$$

que é exatamente α vezes o coeficiente $\alpha_i \in P_0 v$.

2.10 Otimização

Considerando o problema unidimensional de

$$\underset{x \in S}{\text{Minimizar}} \quad F(x) \tag{2.4}$$

Pode-se maximizar uma função F resolvendo simplesmente o problema de

$$\underset{x \in S}{\text{Minimizar}} \quad (-F(x))$$

e então multiplicar o valor ótimo por -1. Para um problema sem restrições o conjunto S seria o $\mathfrak{R}^n$. Para este trabalho considera-se apenas o caso de minimização.

Um ponto x^* minimiza F se

$$F(x^*) \leq F(x), \qquad \forall x \in S$$

O ponto x^* refere-se *ao minimizador global* de $F \in S$. Se x^* satisfaz

$$F(x^*) < F(x), \quad \forall x \in S : x \neq x^*$$

então x^* é *um minimizador global estrito*.

Nem todas as funções possuem minimizador global, e se possuem não existe uma garantia de que será um minimizador global estrito. A série de Taylor é muito usada para otimização, pois está baseada em informações a respeito da função em um dado ponto e essas informações são válidas dentro de uma vizinhança pequena deste ponto.

Se um ponto x^* satisfaz

$$F(x^*) \leq F(x), \quad \forall x \in S : \|x - x^*\| < \epsilon$$

onde ϵ é um número positivo pequeno, o ponto x^* é dito então *minimizador local*.

Se o ponto x^* satisfaz

$$F(x^*) < F(x), \quad \forall x \in S : x \neq x^*, \|x - x^*\| < \epsilon,$$

este ponto é dito *minimizador local estrito*.

2.10.1 Caracterização de um ponto de mínimo

Problemas de otimização em geral envolvem minimização de uma função sujeita a restrições, tal como

$$\underset{s/a \quad Ax=b}{\text{Minimizar}} \quad F(x) \tag{2.5}$$

onde um ponto $\widehat{x}$ que satisfaz todas as restrições para o problema (2.5) é dito ser viável.

Nota-se então que somente pontos viáveis podem ser ótimos, e que a otimalidade em um ponto x^* é definida pelo relacionamento com os pontos pertencentes a sua vizinhança.

Sendo x^* um ponto viável para o problema (2.5) e $V(x^*)$ o conjunto de pontos viáveis contidos em uma vizinhança de x^*, tem-se por definição que

Definição 2.10.1 – *O ponto x^* é um ponto de mínimo local se existe $\delta > 0$, tal que*

i. $F(x)$ *é definida sobre* $V(x^*)$*;*

ii. $F(x^*) \leqslant F(y)$, $\forall y \in V(x^*)$.

2.10.2 Convexidade

Considera-se neste trabalho o caso onde a função a ser minimizada é convexa e a região viável é um conjunto convexo.

Um conjunto S é convexo se, para quaisquer elementos $x, y \in S$

$$\alpha x + (1 - \alpha)y \in S, \quad \forall 0 \leq \alpha \leq 1$$

Ou seja, se $x, y \in S$, então o segmento de reta que une x e y também pertencem a S. Todo conjunto definido por um sistema linear restrito é um conjunto convexo.

Um função F é convexa sobre um conjunto convexo S se satisfaz

$$F(\alpha x + (1 - \alpha)y) \leq \alpha F(x) + (1 - \alpha)F(y), \quad \forall 0 \leq \alpha \leq 1 \quad e \quad x, y \in S$$

Ou ainda, se o segmento de reta que une os pontos $(x, F(x))$ e $(y, F(y))$ estão acima do gráfico da função. Ver Figura 2.1.

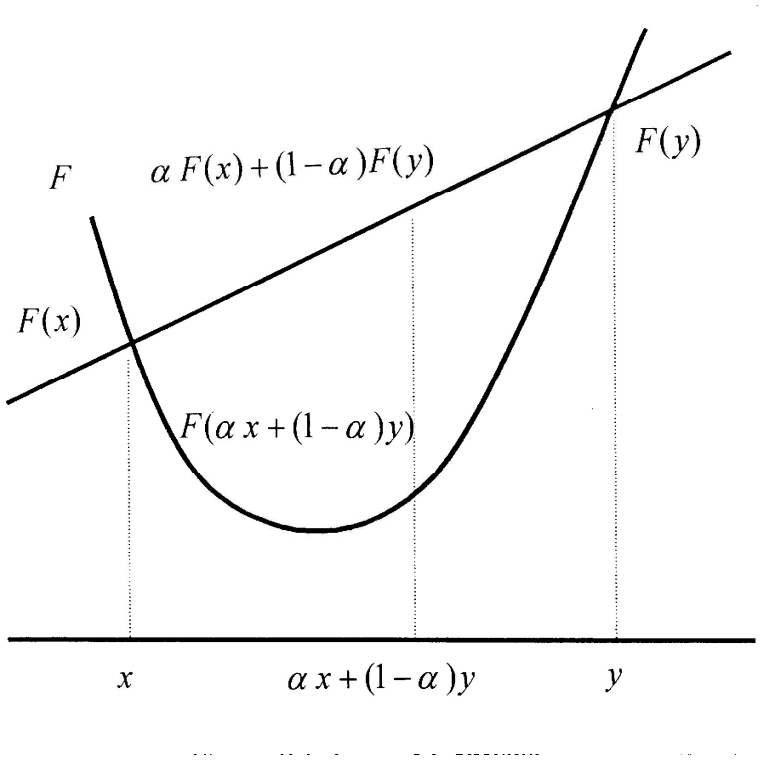

Figura 2.1: Função convexa [NS96]

Uma função F é estritamente convexa quando

$$F(\alpha x + (1-\alpha)y) < \alpha F(x) + (1-\alpha)F(y), \quad \forall x \neq y, \quad 0 \leq \alpha \leq 1 \quad e \quad x, y \in S$$

Define-se, portanto, um problema de programação convexa ao problema da forma:

$$\underset{x \in S}{\text{Minimizar}} \quad F(x)$$

onde S é um conjunto convexo e F é uma função convexa.

2.10.3 Série de Taylor

A Série de Taylor é de fundamental importância e freqüentemente usada em otimização devido ao fato de que se é possível conhecer o valor da função e de suas derivadas em um ponto, pode-se calcular aproximações para a função em todos os pontos pertencentes à vizinhança deste ponto.

Considerando o caso unidimensional, sendo a função F continuamente diferenciável e dado um ponto x_0, a aproximação de uma função F pela Série de Taylor de ordem n é:

$$F(x_0 + d) \cong F(x_0) + dF^{'}(x_0) + \frac{1}{2}d^2F^{''}(x_0) + \ldots + \frac{d^n}{n!}F^{(n)}x_0,$$

onde $F^n(x_0)$ representa a n-ésima derivada de F no ponto x_0, d é uma variável. A aproximação será melhor para valores pequenos de d.

No caso de várias variáveis, existe uma anologia ao caso anterior, pois trabalha-se com a notação de vetores e matrizes, portanto:

$$F(x_0 + d) \cong F(x_0) + d^T\nabla F(x_0) + \frac{1}{2}d^T\nabla^2 F(x_0)d + O(\| d \|_2^2)$$

onde x_0 e d são vetores e $\nabla F(x_0)$ refere-se ao gradiente da função no ponto x_0, a notação $\nabla^2 F(x_0)$ representa a Hessiana da F em x_0.

2.10.4 Algoritmos para otimização

Os algoritmos de otimização sugerem testes de parada para a procura de um ponto ótimo. Um exemplo para o caso unidimensional onde o problema é irrestrito, e se F é derivável, o teste de otimalidade é freqüentemente baseado na condição

$$F'(x) = 0$$

Se $F'(x_k) \neq 0$, então x_k não é ponto ótimo e o sinal do valor de $F'(x_k)$ indica se a função cresce ou decresce no ponto x_k.

Outros algoritmos possuem uma forma mais específica para o teste de otimalidade, tal como:

Dado um ponto x_0

Para k=0,1,. . . ,n

Se $F'(x_k) = 0$, parar

Determinar uma solução estimada melhor: $x_{k+1} = x_k + \alpha_k d_k$, onde d_k é uma direção de procura e α_k é o tamanho do passo.

Para um problema irrestrito, a direção d_k resolve o problema de

$$\underset{d}{\text{Minimizar}} \quad F(x_k + d)$$

se d_k for uma direção de descida para a função F no ponto x_k. Deste modo para passos pequenos ao longo de d_k o valor da função decresce:

$$F(x_k + \alpha d_k) < F(x_k), \qquad 0 < \alpha < \epsilon$$

Para funções lineares $F(x) = c^T x$, d_k será uma direção de descida se

$$c^T(x_k + \epsilon d_k) = c^T x_k + \epsilon c^T d_k < c^T x_k$$

ou ainda, se $c^T d_k < 0$

Supondo disponível d_k, determina-se o tamanho do passo α_k que minimize a função nesta direção, ou seja:

$$\underset{\alpha \geq 0}{\text{Minimizar}} \quad F(x_k + \alpha d_k)$$

A restrição $\alpha \geq 0$ é imposta para que d_k seja uma direção de descida.

O cálculo de αd_k é denominado *line search* (procura em uma linha) porque corresponde a uma procura ao longo da linha $x_k + \alpha d_k$. Ver Figura 2.2.

No caso de um problema restrito pode-se estabelecer uma importante condição para otimalidade. Dado o problema de

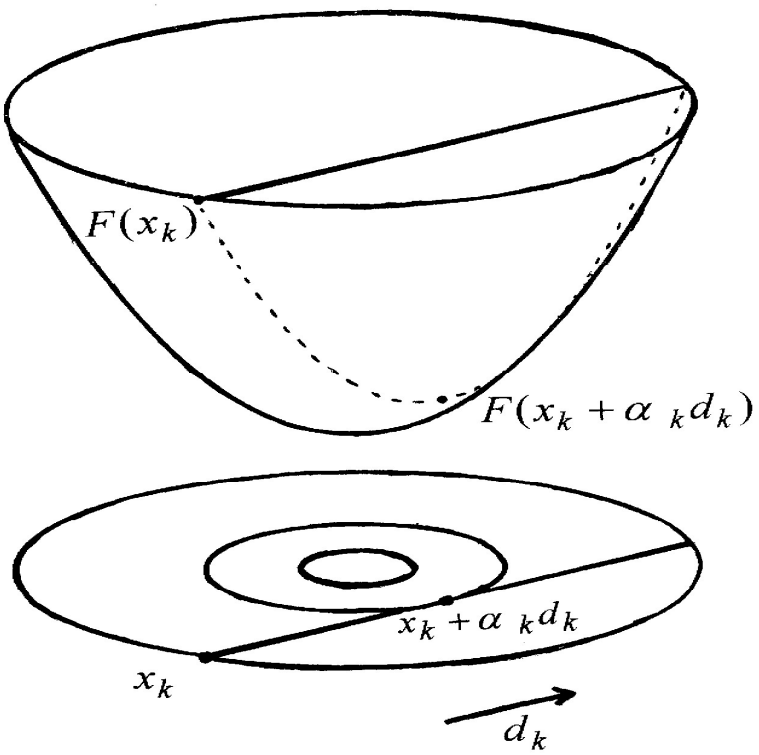

Figura 2.2: Procura ao longo de uma linha - *Line Search*. [NS96]

$$\underset{x \in S}{\text{Minimizar}} \quad F(x)$$

Define-se d uma direção de descida viável em um ponto $x_k \in S$ se, para algum $\epsilon > 0$

$$x_k + \alpha d \in S \text{ e } F(x_k + \alpha d) < F(x_k) \forall 0 < \alpha < \epsilon.$$

2.10.5 Convergência de seqüências

A maioria dos métodos de otimização são iterativos, segundo [WMW91] e [NS96], nestes métodos uma seqüência x_k é gerada para estimar o ponto ótimo x^*. Um método será praticável somente se a convergência ocorre com alguma rapidez em termos computacionais.

Assumindo que a seqüência x_k converge para x^*, que os elementos de x_k são distintos e que x_k é diferente de x^* para algum valor na iteração k; uma técnica para avaliar a

convergência seria comparar a sua melhora a cada passo com o passo anterior. A seqüência x_k converge com ordem r tal que

$$0 \leq \lim_{k \to \infty} \frac{||x_{k+1} - x^*||}{||x_k - x^*||^r} < \infty \tag{2.6}$$

onde r é o raio de convergência assintótica.

Se $r = 1$, a seqüência possui convergência linear, se $r = 2$, a seqüência possui uma convergência quadrática.

Se a seqüência x_k possui ordem de convergência r, a constante de erro assintótica é o valor γ que é definida por

$$\gamma = \lim_{k \to \infty} \frac{||x_{k+1} - x^*||}{||x_k - x^*||^r} \tag{2.7}$$

Observa-se que se $r = 1$, γ deve ser estritamente menor do que 1 para ocorrer a convergência.

Se a seqüência possuir convergência linear, o passo discreto decresce substancialmente em $||x_k - x^*||$ com valor da constante de erro assintótica. Se o limite em (2.7) é zero quando r é considerado unitário, a convergência é dita superlinear.

Capítulo 3

O Problema Quadrático

Este Capítulo apresenta um estudo do problema quadrático, propriedades de funções quadráticas, forma quadrática, condições necessárias e suficientes de otimalidade, e a abordagem adotada para a solução do problema de minimização quadrática.

3.1 Forma quadrática

Seja $F : \Re^n \rightarrow \Re$, uma função e G, uma matriz simétrica, então

$$F(x) = \frac{1}{2}x^T Gx + h^T x + c$$

é denominada de uma forma quadrática. A matriz G é denominada matriz associada à forma quadrática . Se a matriz G é semidefinida positiva então $F(x)$ é uma função convexa e se a matriz G é definida positiva então $F(x)$ é uma função estritamente convexa. Em geral, uma forma quadrática $F(x)$ assumirá valores positivos, negativos e zero para cada valor de x. Um exemplo simples em $\Re^2$ de uma forma quadrática que nunca será negativa é $(x_1 - x_2)^2$. Outro exemplo seria $x_1^2 + x_2^2$, que nunca assumirá valor negativo, e assumirá valor zero somente quando $x = 0$.Então

i. Uma forma quadrática $F(x)$ é dita ser definida positiva ($G > 0$) se $x^T Gx > 0, \forall x \neq 0$;

ii. Uma forma quadrática $F(x)$ é dita ser semidefinida positiva ($G \geq 0$), se $x^T Gx \geq 0$, $\forall x \neq 0$ e semidefinida negativa ($G \leq 0$), se $x^T Gx \leq 0, \forall x \neq 0$;

iii. Se F não é definida nem semidefinida, então é dita indefinida;

Definição 3.1.1 – *O gradiente da função $F(x)$ é dado pelas derivadas de primeira ordem de $F(x)$, isto é,*

$$\nabla F(x) = \begin{pmatrix} \frac{\partial F(x)}{\partial x_1} \\ \vdots \\ \frac{\partial F(x)}{\partial x_n} \end{pmatrix}$$

No caso quadrático,

$$\nabla F(x) = g(x) = Gx + h.$$

■

Definição 3.1.2 – *A Hessiana da função $F(x)$ é dada pelas derivadas de segunda ordem da função e é representada pela matriz simétrica,*

$$\nabla^2 F(x) = \begin{pmatrix} \frac{\partial^2 F(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 F(x)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 F(x)}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 F(x)}{\partial x_n^2} \end{pmatrix}$$

No caso quadrático,

$$\nabla^2 F(x) = G$$

■

3.1.1 Propriedades de funções quadráticas

Considerando a função quadrática $F(x)$ dada por:

$$F(x) = \frac{1}{2} x^T G x + h^T x + c$$

onde h é um vetor pertencente ao $\mathfrak{R}^n$ e $G_{n \times n}$ alguma matriz simétrica.

Expandindo essa função pela série de Taylor no ponto $\hat{x}$, tem-se que

$$F(\widehat{x} + \alpha d) \cong F(\widehat{x}) + \alpha d^T \nabla F(\hat{x}) + \frac{1}{2}\alpha^2 d^T \nabla^2 F(\hat{x}) d$$

$$F(\widehat{x} + \alpha d) \cong F(\widehat{x}) + \alpha d^T (G\widehat{x} + h) + \frac{1}{2}\alpha^2 d^T G d \tag{3.1}$$

onde $d \in \Re^n$, e α é um escalar qualquer.

Se a função F possui um ponto estacionário x^* então o gradiente de F em x^* é nulo, ou seja, $\nabla F(x^*) = Gx^* + h = 0$. Portanto, o ponto x^* deve ser solução do sistema de equações lineares

$$Gx^* + h = 0 \tag{3.2}$$

Se o sistema (3.2) é incompatível, o vetor h não pode ser expresso como uma combinação das colunas da matriz G, logo F não possui ponto estacionário e não é limitado superior nem inferiormente. Caso o sistema seja compatível, existe um ponto estacionário, o qual é único se a matriz G é não singular.

Se x^* é um ponto estacionário, segue de (3.1) e (3.2) que,

$$F(x^* + \alpha d) \cong F(x^*) + \frac{1}{2}\alpha^2 d^T G d. \tag{3.3}$$

Sejam λ_j, $j = 1 : n$ os autovalores da matriz A e u_j, $j = 1 : n$ os autovetores normalizados correspondentes, tem-se que

$$Gu_j = \lambda_j u_j, \quad j = 1 : n \, .$$

Os vetores u_j correspondentes a autovalores distintos são ortonormais. Quando d é igual a u_j, a equação (3.3) fica,

$$F(x^* + \alpha u_j) \cong F(x^*) + \frac{1}{2}\alpha^2 \lambda_j$$

Se λ_j é positivo, F cresce estritamente com o crescimento de $|\alpha|$.

Se λ_j é negativo, F decresce monotonamente com o crescimento de $|\alpha|$.

Se λ_j é nulo, os valores de F permanecem constantes quando movimenta ao longo de uma direção paralela a u_j como $Gu_j = 0$, além disso, F reduz-se para uma função linear ao longo de alguma direção, desde que os termos quadráticos da expressão (3.1) desapareçam.

Quando todos os autovalores da matriz G são positivos, x^* é o único mínimo global

de F. Neste caso, os contornos da função F são elipsóides onde os eixos principais estão na direção dos correspondentes autovetores. Se G é semidefinida positiva, um ponto estacionário (se existir) é um ponto de mínimo local. Se G é indefinida e não singular, x^* é um ponto de sela, e F não é limitado superior nem inferiormente.

3.2 Otimização irrestrita - condições necessárias e suficientes para otimalidade.

Considerando o problema de minimização irrestrita a seguir:

$$\underset{x\in\Re^n}{\text{Minimizar}}\ F(x) \tag{3.4}$$

onde F é uma função de $\Re^n \to \Re$, $F \in C^2(\Re^n)$.

As condições necessárias para x^* ser um ponto de mínimo local são:

1. $\nabla F(x^*) = 0$, ou seja, x^* é um ponto estacionário, e
2. $\nabla^2 F(x^*) \geq 0$.

Expandindo a função F em x^* pela série de Taylor até a segunda ordem, obtém-se:

$$F(x^* + \alpha d) \cong F(x^*) + \alpha d^T \nabla F(x^*) + \frac{1}{2}\alpha^2 d^T \nabla^2 F(x^* + \alpha\theta d)d, \qquad 0 \leq \theta \leq 1 \tag{3.5}$$

onde α é um escalar qualquer.

Assumindo que x^* é um mínimo local e supondo que ele não seja um ponto estacionário, se $\nabla F(x^*) \neq 0$, então deve existir um vetor d tal que,

$$d^T \nabla F(x^*) < 0. \tag{3.6}$$

Todo vetor d que satisfaz a expressão (3.6) é uma direção de descida para F em x^*. Dado alguma direção de descida d, existe um escalar $\bar{\alpha}$ positivo tal que para todo α positivo, $\alpha < \bar{\alpha}$, obtém-se

$$\alpha d^T \nabla F(x^*) + \frac{1}{2}\alpha^2 d^T \nabla^2 F(x^* + \alpha\theta d)d < 0.$$

Então toda vizinhança de x^* conterá pontos onde F tem valores estritamente menores do que em x^*, $F(x^* + \alpha d) < F(x^*)$, o que é um absurdo, logo a suposição de que $\nabla F(x^*) \neq 0$

é falsa, isto mostra que todo mínimo local deve ser ponto estacionário.

No caso unidimensional, o vetor gradiente pode anular-se no ponto que não é um mínimo local. Se o gradiente é zero no ponto $\hat{x}$, e este ponto não é ponto de mínimo nem ponto de máximo, então $\hat{x}$ é conhecido como um ponto de sela.

Da expressão (3.5), se $\nabla F(x^*) = 0$, então fica-se apenas com,

$$F(x^* + \alpha d) \cong F(x^*) + \frac{1}{2}\alpha^2 d^T \nabla^2 F(x^* + \alpha\theta d)d \tag{3.7}$$

Se $\nabla^2 F(x^*)$ é indefinida, por continuidade, F será indefinida para todos os pontos de alguma vizinhança de x^*, e escolhendo $|\alpha|$ em (3.5) suficientemente pequeno, $x^* + \alpha d$ está na vizinhança. Pela definição de matriz indefinida, d pode ser escolhido de forma que $d^T \nabla^2 F(x^* + \alpha\theta d)d < 0$. Conseqüentemente, toda vizinhança de x^* conterá pontos onde o valor da função F é estritamente menor do que em x^*, contradizendo assim a otimalidade de x^*.

Pode-se afirmar que para x^* ser um ponto de mínimo local estrito, as condições que seguem são suficientes.

1. $\nabla F(x^*) = 0$ e
2. $\nabla^2 F(x^*) > 0$.

Concluímos que se $\nabla^2 F(x^*)$ é definida positiva, por continuidade, F é definida positiva para todos os pontos dentro de alguma vizinhança de x^*. Se $|\alpha|$ é bastante pequeno, $x^* + \alpha d$ estará dentro desta vizinhança, por isso, para todo α e $d \neq 0$, tem-se que $d^T \nabla^2 F(x^* + \alpha\theta d)d > 0$. Da expressão (3.7) observa-se que $F(x^*)$ será estritamente menor quando o valor da função F para todos os pontos dentro de alguma vizinhança de x^*, e assim x^* é um ponto de mínimo local estrito, ou seja, $F(x^* + \alpha d) > F(x^*)$.

3.3 Otimização com restrições lineares - condições necessárias e suficientes

A forma geral de uma função linear é

$$l(x) = a^T x - b,$$

onde $a \in \Re^n$ e b é algum escalar.

Considera-se dois tipos de restrições lineares

1. $a^T x - b = 0$ (restrição de igualdade) ou ainda, $a^T x = b$
2. $a^T x - b \geq 0$ (restrição de desigualdade) ou ainda, $a^T x \geq b$

As restrições do tipo $a^T x - b \leq 0$ equivalem a $-a^T x + b \geq 0$.

Se x_i é uma variável real as restrições mencionadas tomam a forma:

1. $a^T x_i = b$, x_i é fixado em b.
2. $a^T x_i \geq b$, b é um limite inferior para x_i.
3. $a^T x_i \leq b$, b é um limite superior para x_i.

Os itens 2 e 3 acima são denominados simplesmente de limites inferior e superior, para a variável x_i.

Neste trabalho serão consideradas apenas restrições lineares de igualdade.

Considerando as condições de otimalidade para problemas que contém somente restrições lineares de igualdade do tipo,

$$\underset{s/a \quad Ax=b}{\text{Minimizar}} \; F(x) \tag{3.8}$$

onde F é uma função quadrática,

$$F(x) = \frac{1}{2} x^T G x + h^T x + c,$$

e A é uma matriz $m \times n$, com $m < n$, $posto(A) = m$, a_i^T representa a i-ésima linha dessa matriz, e contém os coeficientes da i-ésima restrição linear:

$$a_i^T x = a_{i1} x_1 + \ldots + a_{in} x_n = b_i$$

x^* será uma solução de (3.8) se $F(x^*) \leq F(x)$, para todo x viável em alguma vizinhança de x^* e se $Ax^* = b$.

Não existe ponto viável se as restrições são inconsistentes, assim assume-se que $b \in R(A)$. Se as m linhas da matriz A são linearmente independentes, as restrições diminuem em m os graus de liberdade da escolha do ponto x^*.

Sejam $\bar{x}$ e $\hat{x}$ pontos viáveis das restrições, por linearidade tem-se $A(\bar{x} - \hat{x}) = 0$, desde que $A\bar{x} = b$ e $A\hat{x} = b$. O passo d de algum ponto viável para outro ponto viável deve ser então ortogonal às linhas de A, ou seja, satisfazer:

$$Ad = 0 \tag{3.9}$$

Então se d é uma direção viável com respeito as restrições de igualdade, qualquer passo ao longo dessa direção não viola as restrições, desde que $A(\hat{x}+\alpha d) = A\hat{x} = b$. Como o espaço nulo de A tem dimensão $n - m$, precisa-se obter um conjunto de geradores para este espaço e denomina-se de Z a matriz formada por esses vetores, então $AZ = 0$, e toda direção viável pode ser escrita como uma combinação linear das colunas de Z. Se d satisfaz a equação (3.9) pode-se escrever $d = Zd_z$, para algum d_z. Examinando a expansão de F pela série de Taylor em x^* ao longo da direção d ($d = Zd_z$), tem-se

$$F(x^* + \alpha Zd_z) \cong F(x^*) + \alpha d_z^T Z^T \nabla F(x^*) + \frac{1}{2}\alpha^2 d_z^T Z^T \nabla^2 F(x^* + \alpha\theta d) Zd_z, \tag{3.10}$$

com $0 \leq \theta \leq 1$, e $\alpha \in \Re$.

Se $d_z^T Z^T \nabla F(x^*) < 0$, então toda vizinhança de x^* conterá pontos viáveis com valores da função estritamente menores do que em x^* . Assim, x^* não será um ponto de mínimo local. Se $d_z^T Z^T \nabla F(x^*)$ desaparecer para todo d_z,

$$Z^T \nabla F(x^*) = 0 \tag{3.11}$$

Isto é, o gradiente da função F em x^* é ortogonal ao subespaço gerado pelas colunas de Z, que é igual ao $\aleph(A)$. O vetor $Z^T \nabla F(x^*)$ é denominado gradiente reduzido de F em x^*.

Qualquer ponto no qual $Z^T \nabla F(x) = 0$ é denominado ponto estacionário. O gradiente de F em x^*, $\nabla F(x^*)$, deve ser ortogonal as colunas de Z, e como as colunas de A são também ortogonais às colunas de Z então $\nabla F(x^*)$ deve ser paralelo às linhas de A, logo:

$$\nabla F(x^*) = \sum_{i=1}^{m} a_i \lambda_i^* = A^T \lambda^* \tag{3.12}$$

onde λ^* é denominado vetor de multiplicador de Lagrange, a_i, $i = 1 : m$ são as linhas de A. Se as linhas de A são linearmente independentes, o vetor de multiplicador de lagrange é único. Devido ao fato de que todo vetor de dimensão n pode ser escrito como uma combinação linear das linhas de A mais uma combinação linear das colunas

de Z, então $\nabla F(x^*)$ pode ser escrito como:

$$\nabla F(x^*) = A^T\lambda^* + Zd_z \tag{3.13}$$

para algum vetor λ^* e d_z.

Pré-multiplicando $\nabla F(x^*)$ por Z^T e usando a expressão (3.13), tem-se que

$$Z^T Z d_z = 0, \tag{3.14}$$

$$\begin{aligned} Z^T\nabla F(x^*) &= Z^T A^T\lambda^* + Z^T Z d_z \\ &= (AZ)^T\lambda^* + Z^T Z d_z \\ &= Z^T Z d_z \end{aligned}$$

desde que $Z^T Z$ seja não singular, a equação (3.13) será verdadeira somente se $d_z = 0$.

Desde que $Z^T\nabla F(x^*) = 0$, a expansão da série de Taylor da expressão (3.10) fica:

$$F(x^* + \alpha Z d_z) \cong F(x^*) + \frac{1}{2}\alpha^2 d_z^T Z^T\nabla^2 F(x^* + \alpha\theta d) Z d_z \tag{3.15}$$

Por um argumento similar ao caso irrestrito, a expressão (3.15) mostra que se a matriz $Z^T\nabla^2 F(x^*)Z$ é indefinida, toda vizinhança de x^* conterá pontos viáveis com um valor estritamente menor de F. Portanto, uma condição necessária para x^* ser ótimo para o problema com restrição linear é que a matriz $Z^T\nabla^2 F(x^*)Z$ deve ser semi-definida positiva.

Deste modo, as condições necessárias para x^* ser um ponto de mínimo local são:

1. $Ax^* = b$;
2. $Z^T\nabla F(x^*) = 0$ ou, $\nabla F(x^*) = A^T\lambda^*$;
3. $Z^T\nabla^2 F(x^*)Z \geq 0$.

Se na condição (3) anterior, $Z^T\nabla^2 F(x^*)Z > 0$, as condições suficientes são:

1. $Ax^* = b$;
2. $Z^T\nabla F(x^*) = 0$ ou, $\nabla F(x^*) = A^T\lambda^*$;
3. $Z^T\nabla^2 F(x^*)Z > 0$.

3.4 Programação quadrática

A programação quadrática trata de um problema onde a função objetivo é uma quadrática, também denominado simplesmente de um problema quadrático. Considerando então o seguinte problema de programação quadrática

$$\underset{s/a \quad Ax=b}{\text{Minimizar}} \; F(x) = \frac{1}{2}x^T Gx + h^T x + c, \tag{3.16}$$

para uma matriz simétrica constante G e um vetor c. Sendo Z uma base para a matriz A de restrições, tem-se que $Z^T GZ$ é definida positiva logo a solução de (3.16) é única. Inicia-se a solução com um dado ponto viável $\widehat{x}$, onde o gradiente da função é

$$g(\widehat{x}) = G\widehat{x} + h$$

e o passo d, de $\widehat{x}$ para x^* é a solução do problema

$$\underset{s/a \quad Ad=0}{\text{Minimizar}} \; F(x) = \frac{1}{2}d^T Gd + g(\widehat{x})^T d \tag{3.17}$$

Como já foi visto anteriormente d pode ser escrito como Zd_z, logo pode-se obter a solução do problema (3.17) resolvendo o problema irrestrito:

$$\underset{s/a \quad d_z \in \Re^{n-m}}{\text{Minimizar}} \; F(x) = \frac{1}{2}d_z^T Z^T GZd_z + g(\widehat{x})^T Z^T d_z$$

definido pelo problema linear

$$Z^T GZd_z = -Z^T g(\widehat{x})$$

Assim d é dado por

$$d = -Z(Z^T GZ)^{-1} Z^T g(\widehat{x})$$

e a solução do problema é

$$x^* = \widehat{x} + d \tag{3.18}$$

As condições de otimalidade discutidas anteriormente mantém-se, então

$$g(x^*) = g(\widehat{x}) + Gd = A^T\lambda$$

onde λ é um vetor multiplicador de Lagrange. Observa-se que se a matriz Hessiana G é indefinida, não existe solução finita para o problema (3.16).

3.5 Abordagem adotada para a solução do problema quadrático

A abordagem adotada neste trabalho para resolver o problema

$$\underset{s/a \quad Ax=b}{\text{Minimizar}} \quad F(x) \tag{3.19}$$

onde $F : \Re^n \rightarrow \Re$ tem derivadas de segunda ordem contínuas, $A_{m\times n}$, com $m < n$, de posto m e b um vetor do $\Re^n$, exige que o vetor de busca unidimensional ao longo do qual o algoritmo usado atualiza a aproximação da solução, deve estar no espaço nulo de A, $\aleph(A)$. Como por hipótese a matriz A tem posto m, então A tem m colunas linearmente independentes e portanto $posto(\aleph(A)) = n - m$ o que assegura que o núcleo de A contém vetores não nulos.

Seja P uma matriz de permutação tal que a matriz AP pode ser particionada como,

$$AP = \begin{pmatrix} B & N \end{pmatrix},$$

onde B é uma matriz $m \times n$, não singular. Chamando de $x_B \in \Re^m$, os componentes do vetor $x \in \Re^n$ correspondentes às colunas da matriz B, e $x_N \in \Re^{n-m}$, as componentes de x correspondentes às colunas de N, então tem-se que

$$P^T x = \begin{pmatrix} x_B \\ x_N \end{pmatrix},$$

Usando-se o fato de que P é uma matriz ortogonal, tem-se que

$$b = Ax = APP^T x = \begin{pmatrix} B & N \end{pmatrix} \begin{pmatrix} x_B \\ x_N \end{pmatrix}$$

$$= Bx_B + Nx_N$$

como B é não singular, então

$$x_B = B^{-1}b - B^{-1}Nx_N$$

Supondo então que as m primeiras colunas da matriz A sejam linearmente independentes, e portanto pode-se particionar A como

$$A = \begin{pmatrix} B & N \end{pmatrix}$$

Então o sistema $Ax = b$ é equivalente a

$$Ax = b \Leftrightarrow Bx_B + Nx_N = b$$

e portanto qualquer solução x de $Ax = b$, tem a forma

$$x = \begin{pmatrix} B^{-1}b - B^{-1}Nx_N \\ x_N \end{pmatrix} = \begin{pmatrix} B^{-1} \\ 0 \end{pmatrix} b + \begin{pmatrix} -B^{-1}N \\ I_{n-m} \end{pmatrix} x_N$$

Lembrando que $\aleph(A) = \{x \in \Re^n : Ax = 0\}$, para ober-se uma base Z para o núcleo da matriz de restrições A, considerando o sistema $Ad = 0$, seleciona-se um conjunto de m variáveis cujas colunas são linearmente independentes, denotando-as de variáveis básicas e de B a matriz $m \times n$ definida por estas colunas, e as outras variáveis como variáveis não básicas e de N a matriz $m \times (n - m)$ definidas por essa outras colunas.

Em geral, a solução do sistema Ad=0 é obtido em termos de B e N, então

$$Ad = \begin{pmatrix} B & N \end{pmatrix} \begin{pmatrix} x_B \\ x_N \end{pmatrix} = Bx_B + Nx_N = 0$$

Pré-multiplicando a equação anterior por B^{-1} tem-se

$$x_B = -B^{-1}Nx_N$$

Assim o conjunto solução do sistema $Ad = 0$ é

$$x^* = \begin{pmatrix} x_B \\ x_N \end{pmatrix} = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix} x_N$$

e a matriz $Z_{n\times(n-m)}$

$$Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}$$

é uma base para o espaço nulo de A, $\aleph(A)$.

Uma solução viável para o problema $Ax = b$ é

$$\overline{x} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$$

Se x é algum ponto que satisfaz $Ax = b$, então pode ser escrito como

$$x = \overline{x} + d = \overline{x} + Zx_N = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix} + \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix} x_N$$

.

Outra maneira de obter uma base para o $\aleph(A)$, é pela decomposição QR de A^T, esta é então uma base ortogonal. Fazendo

$$A^T = QR,$$

onde Q é uma matriz ortogonal de dimensão $n \times n$, onde as m primeiras colunas de Q geram o espaço linha de A, $R(A^T)$ e as $n - m$ últimas colunas geram o espaço nulo de A, $\aleph(A)$.

Pode-se particionar Q da forma

$$Q = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix}$$

e

$$R = \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$$

onde Q_1 é uma matriz $n \times m$ cujas colunas geram $R(A^T$ e Q_2 é uma matriz $n \times (n - m)$ cujas colunas geram $\aleph(A)$ e R_1 é uma matriz triangular superior, tem-se que

$$AQ_1 = R_1^T$$

$$AQ_2 = 0$$

Assim

$$Z = Q_2$$

é uma base ortogonal para o espaço nulo de A, portanto $Z^T Z = I$.

Se fosse o caso do uso do método de Newton, é visto que o número de condição de $Z^T GZ$ depende do número de condição de Z e G:

$$cond(Z^T GZ) \leq cond(G).(cond(Z))^2 \tag{3.20}$$

A vantagem de usar a fatoração QR é que a escolha de Z ortogonal não aumenta o número de condição do sistema. A matriz Z determinada da decomposição QR de A^T possui número de condição, $cond(Z) = 1$ e portanto

$$cond(Z^T GZ) \leq cond(G) \tag{3.21}$$

o que implica que o condicionamento do problema original não é pior junto ao método numérico iterativo. Considera-se neste trabalho a matriz G esparsa simétrica e definida positiva.

Obtidas as bases para o $\aleph(A)$, propõe-se permutar as colunas das mesmas utilizando os métodos de Gradiente Conjugado e Gradiente Conjugado Pré-condicionado para encontrar a direção pertencente ao $\aleph(A)$ e por fim encontrar a solução dada por (3.18),

$$x^* = \widehat{x} + d$$

onde o objetivo principal é analisar o impacto dos métodos de gradiente sobre o comportamento da seqüência gerada pelo algoritmo. No próximo Capítulo descreve-se os métodos de Gradiente Conjugado.

Capítulo 4

Métodos que usam Direções Conjugadas

Neste Capítulo desenvolveu-se uma exposição teórica dos métodos que usam direções conjugadas e suas propriedades, apresentando algoritmos de Gradiente Conjugado e Gradiente Conjugado Pré-condicionado.

4.1 Método das direções conjugadas

Método de direções conjugadas é, em geral, usado para resolver problemas quadráticos do tipo

$$\text{Minimizar } \frac{1}{2}x^T Ax + b^T x + c$$

onde A é uma matriz simétrica definida positiva.

Definição 4.1.1 – *Dada uma matriz quadrada A, dois vetores d_1 e d_2 são ditos A-ortogonais, ou conjugados com respeito a matriz A, se $d_1^T A d_2 = 0$.*

Nesse sentido o conceito de conjugacidade é uma generalização do conceito de ortogonalidade, isto é, se a matriz $A = I$, então o conceito de conjugacidade se reduz ao conceito usual de ortogonalidade.

Teorema 4.1.1 – *Se A é uma matriz definida positiva e $d_0, d_1, \ldots, d_{n-1}$ é um conjunto de vetores não nulos conjugados com respeito a A, então este conjunto é linearmente independente.*

PROVA: Suponha que existam constantes $\alpha_i, i = 0, 1, \ldots, k$ *tais que,*

$$\alpha_0 d_0 + \ldots + \alpha_k d_k = 0.$$

Multiplicando essa equação por A e fazendo o produto interno com d_i*, obtém-se*

$$\alpha_0 d_0^T A d_0 + \ldots + \alpha_k d_k^T A d_k = 0, \text{ donde}$$

$$\alpha_i d_i^T A d_i = 0$$

Visto que A é definida positiva, então $d_i^T A d_i > 0$*, donde obtém-se* $\alpha_i = 0, \forall i = 0, 1, \ldots, n-1$.

■

Discute-se agora porque o conceito de conjugacidade é tão útil nos problemas quadráticos do tipo

$$Minimizar \frac{1}{2} x^T A x - b^T x + c, \tag{4.1}$$

onde A é uma matriz definida positiva. A condição de otimalidade para esse problema implica que minimizar esta quadrática é equivalente a resolver o sistema de equações lineares

$$Ax = b. \tag{4.2}$$

Resolver o sistema $Ax = b$ é encontrar x que anula o gradiente da função quadrática e como a função quadrática é definida positiva, então $g(x) = 0$ é condição necessária e suficiente para otimalidade.

Seja A uma matriz definida positiva, e sejam $d_0, d_1, \ldots, d_{n-1}$, n vetores não nulos conjugados com relação a matriz A. Pelo teorema anterior, esses n vetores são linearmente independentes no $\Re^n$, logo formam uma base para o $\Re^n$, o que implica que a solução x^* de (4.1) ou de (4.2) pode ser escrita como,

$$x^* = \alpha_0 d_0 + \ldots + \alpha_{n-1} d_{n-1}. \tag{4.3}$$

Pré-multiplicando a relação (4.3) por A e fazendo o produto interno com d_i, tem-se

$$d_i^T A x^* = \alpha_0 d_i^T A d_0 + \ldots + \alpha_1 d_i^T A d_1 + \ldots + \alpha_i d_i^T A d_i + \ldots + \alpha_{n-1} d_i^T A d_{n-1}$$

como $d_i^T A d_j = 0$, $\forall i \neq j$ obtém-se

$$d_i^T A x^* = \alpha_i d_i^T A d_i, \text{ logo}$$

$$\alpha_i = \frac{d_i^T A x^*}{d_i^T A d_i} \tag{4.4}$$

Devido ao fato de que $Ax^* = b$, α_i pode ser escrito definitivamente como

$$\alpha_i = \frac{d_i^T b}{d_i^T A d_i}. \tag{4.5}$$

Isto mostra que os $\alpha_i's$ e conseqüentemente a solução x^* podem ser determinados por simples cálculo de produto interno. Substituindo (4.5) em (4.3) tem-se

$$x^* = \frac{d_i^T b}{d_i^T A d_i} d_0 + \ldots + \frac{d_i^T b}{d_i^T A d_i} d_{n-1}, \text{ logo}$$

$$x^* = \sum_{i=0}^{n-1} \frac{d_i^T b}{d_i^T A d_i} d_i. \tag{4.6}$$

Existem duas idéias fundamentais por trás da solução (4.6). A primeira é a de selecionar um conjunto de vetores conjugados com relação a matriz A, $\{d_0, d_1, \ldots, d_{n-1}\}$, de modo que ao fazer o produto interno, todos os termos do lado direito de (4.3) desaparecem, exceto o i-ésimo. Esse mesmo resultado poderia ser obtido tomando-se o conjunto $\{d_0, d_1, \ldots, d_{n-1}\}$ ortogonal no sentido usual.

De fato

$$d_i^T x^* = \alpha_0 d_i^T d_0 + \ldots + \alpha_1 d_i^T d_1 + \ldots + \alpha_i d_i^T d_i + \ldots + \alpha_{n-1} d_i^T d_{n-1}$$

Devido a ortogonalidade, fica então que $d_i^T x^* = \alpha_i d_i^T d_i$, donde finalmente,

$$\alpha_i = \frac{d_i^T x^*}{d_i^T d_i}$$

então

$$x^* = \sum_{i=0}^{n-1} \frac{d_i^T x^*}{d_i^T d_i} d_i$$

que não é um resultado tão bom quanto o obtido quando os vetores são conjugados com relação a matriz A.

A segunda idéia fundamental é que usando o conceito de conjugacidade consegue-se expressar os $\alpha_i's$ em termos de vetores conhecidos.

Observando a relação (4.3) pode-se considerar que x^* é obtido após n passos onde em cada passo acrescenta-se ao anterior o passo $\alpha_i d_i$,

$$x^* = \frac{d_o^T b}{d_0^T A d_0} d_0 + \ldots + \frac{d_i^T b}{d_i^T A d_i} d_i + \ldots + \frac{d_{n-1}^T b}{d_{n-1}^T A d_{n-1}} d_{n-1}$$

Teorema 4.1.2 – *Teorema das direções conjugadas.*

Seja $\{d_0, d_1, \ldots, d_{n-1}\}$ um conjunto não nulo de vetores conjugados em relação a A, uma matriz definida positiva. Para qualquer $x_0 \in \mathfrak{R}^n$ a seqüência $\{x_k\}$ gerada por

$$x_{k+1} = x_k + \alpha_k d_k \qquad k \geq 0 \tag{4.7}$$

com

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T A d_k} \tag{4.8}$$

e

$$g_k = A x_k - b = \nabla F(x_k)$$

converge para uma solução única, x^* do problema (4.1) , e no máximo em n passos, isto é, $x_k = x^*$, $k \leq n$.

■

PROVA: [Lue73] Desde que $d_k's$ sejam linearmente independentes, pode-se escrever

$$x^* - x_0 = \alpha_0 d_0 + \alpha_1 d_1 + \ldots + \alpha_{n-1} d_{n-1} \tag{4.9}$$

para algum conjunto de $\alpha_i's$. Se multiplicar (4.9) por A e fazer o produto escalar com d_k,

tem-se,

$$\alpha_k = \frac{d_k^T A(x^* - x_0)}{d_k^T A d_k} \tag{4.10}$$

Seguindo o processo iterativo (4.7) de x_0 até x_k tem-se

$$x^k - x_0 = \alpha_0 d_0 + \alpha_1 d_1 + \ldots + \alpha_{k-1} d_{k-1} \tag{4.11}$$

e por conjugacidade, segue que

$$d_k^T A(x^k - x_0) = 0 \tag{4.12}$$

Substituindo (4.12) em (4.10) tem-se,

$$\alpha_k = \frac{d_k^T A(x^* - x_k)}{d_k^T A d_k} = -\frac{g_k^T d_k}{d_k^T A d_k} \tag{4.13}$$

que é idêntica a (4.8).

■

Se a matriz A é semidefinida positiva e diagonal então, as curvas de níveis da função $F(x)$ são elipses cujos eixos principais são paralelos aos eixos do sistema, como mostra a Figura 4.1.

Quando a matriz A não é diagonal, os eixos principais das curvas de níveis (elipses) não são paralelos aos eixos coordenados e, a estratégia de minimização sucessiva ao longo dessas direções conjugadas não mais conduz a solução em n iterações. O que é ilustrado na Figura 4.2.

Como toda quadrática onde a matriz A é simétrica pode ser diagonalizada pode-se recuperar o comportamento da função representada na Figura 4.1. Isto é realizado diagonalizando a Hessiana da quadrática que é a matriz constante A. Seja

$$\widehat{x} = V^{-1} x \tag{4.14}$$

Onde V é uma matriz $n \times n$ definida por

$$V = [d_0, d_1, \ldots, d_{n-1}],$$

onde $\{d_0, d_1, \ldots, d_{n-1}\}$ é o conjunto de direções conjugadas com respeito a matriz A.

A função quadrática F agora é

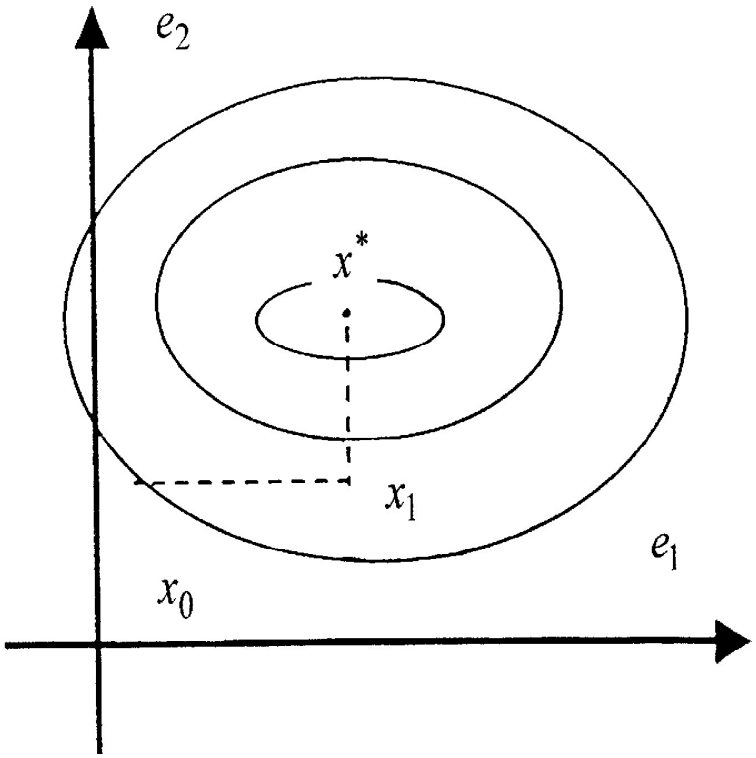

Figura 4.1: Minimização da quadrática ao longo de direções conjugadas sucessivas com uma Hessiana diagonal em n iterações [NW99].

$$\phi(\widehat{x}) = F(V\widehat{x}) = \frac{1}{2}\widehat{x}^T(V^TAV)\widehat{x} - (V^Tb)^T\widehat{x} + c$$

Como V é uma matriz não singular, V^TAV é definida positiva, quando A é definida positiva e diagonal pode-se minimizar essa função ϕ como no caso da Figura 4.1. Pela propriedade de conjugacidade, $d_i^TAd_j = 0$, para $i \neq j$, a matriz V^TAV é diagonal, pode-se então determinar o mínimo de ϕ em n minimizações unidimensionais ao longo dos eixos conjugados.

Cada valor de $\widehat{x}$ obtido pelas minimizações consecutivas ao longo das direções conjugadas tem sua correspondente no espaço das variáveis x por meio da equação (4.14).

Quando a matriz Hessiana da função é diagonal os vetores conjugados são exatamente os vetores canônicos do sistema e portanto, cada minimização ao longo das direções conjugadas fornece exatamente uma coordenada da solução. O Teorema do

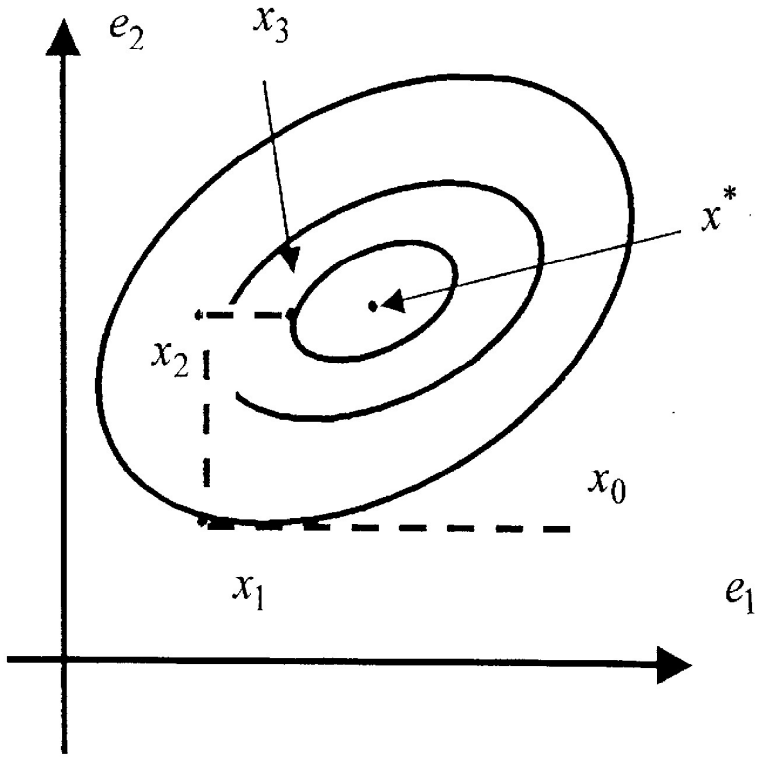

Figura 4.2: Minimizações sucessivas ao longo dos eixos coordenados não determinam a solução em n iterações, para uma quadrática convexa geral. [NW99].

Subespaço Gerado, que será apresentado a seguir, mostra que este resultado é verdadeiro, em geral, mesmo para o caso em que a Hessiana não é uma matriz diagonal.

4.1.1 Propriedades de descida do método de direções conjugadas

Seja V_k um subespaço de $\Re^n$ gerado por $\{d_0, d_1, \ldots, d_{k-1}\}$, e $x_0 \epsilon \Re^n$. O método de direções conjugadas na k-ésima iteração minimiza a função objetivo sobre a variedade linear $x_0 + V_k$, onde $V_k = [d_0, d_1, \ldots, d_{k-1}]$.

Teorema 4.1.3 – *Teorema do Subespaço Gerado.*

Seja $\{d_0, d_1, \ldots, d_{k-1}\}$ *, um conjunto de vetores não nulos conjugados em relação a matriz simétrica e definida positiva A em* $\Re^n$*. Então para qualquer* $x_0 \in \Re^n$ *a seqüência* $\{x_k\}$ *gerada por*

$$x_{k+1} = x_k + \alpha_k d_k \tag{4.15}$$

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T A d_k} \tag{4.16}$$

possui a propriedade de que x_k *minimiza* $F(x) = \frac{1}{2}x^T Ax - b^T x + c$ *sobre o hiperplano* $x = x_{k-1} + \alpha d_{k-1}$, $-\infty < \alpha < \infty$, *bem como sobre a variedade linear* $x_0 + V_k$.

■

A idéia fundamental da prova é que o gradiente da função $\nabla F(x)$ ou $g(x)$ é ortogonal à variedade linear V_k. Logo quando $V_k = \mathfrak{R}^n$, $\nabla F(x)$ é ortogonal a $\mathfrak{R}^n$ e portanto $\nabla F(x) = 0$, que é a condição de otimalidade. Então o método converge em no máximo n passos.

PROVA: [Lue73] Prova-se que g_k é ortogonal a V_k por indução. Por hipótese V_0 é vazio para $k = 0$, assume-se que g_k é ortogonal a V_k e mostra-se que g_{k+1} é ortogonal a V_{k+1}. Tem-se então que,

$$g_{k+1} = g_k + \alpha_k A d_k \tag{4.17}$$

multiplicando a equação (4.17) por d_k^T obtém-se que

$$d_k^T g_{k+1} = d_k^T g_k + \alpha_k d_k^T A d_k = 0$$

pela definição de α_k. E para $i < k$

$$d_i^T g_{k+1} = d_i^T g_k + \alpha_k d_i^T A d_k$$

Então:

1. $d_i^T g_k = 0$ por hipótese da indução (ortogonalidade).
2. $\alpha_k d_i^T A d_k = 0$ pelos vetores $d_i's$ serem conjugados com relação a matriz A.

Portanto g_{k+1} é ortogonal a V_{k+1}.

■

A Figura 4.3 mostra um esquema geométrico do funcionamento do Método de Direções Conjugadas.

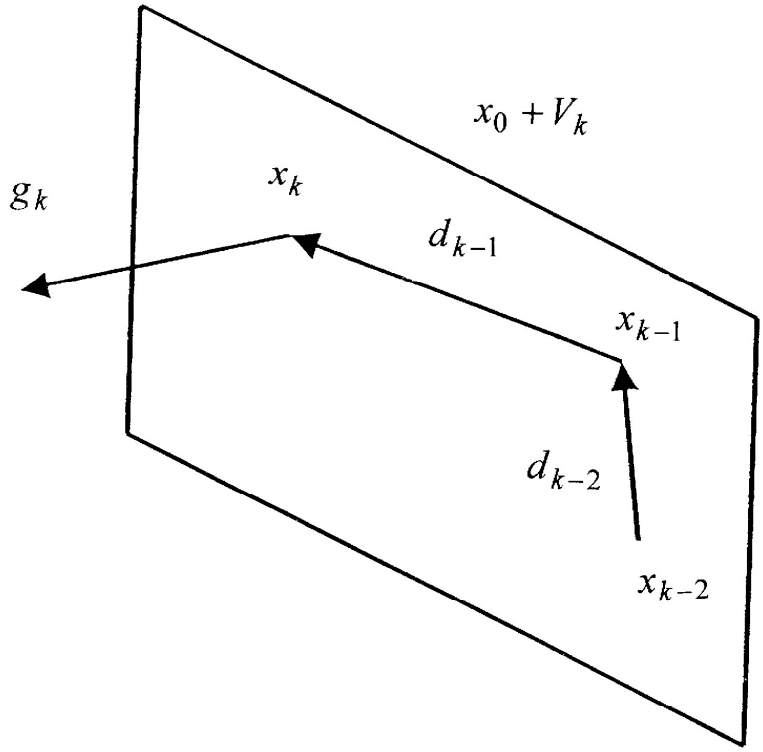

Figura 4.3: Método das direções conjugadas. [Lue73]

Corolário 4.1.4 – *No método de direções conjugadas, os gradientes* g_k, $k = 0 : n$ *satisfazem* $g_k^T d_i = 0$, *para* $i < k$.

Já a Figura 4.4 ilustra um esquema geométrico da interpretação do Teorema do Subespaço Gerado.

4.1.2 Método de gradiente conjugado

O método de Gradiente Conjugado, (*CG*), é um método iterativo, que quando aplicado a minimização de uma função quadrática ou para resolver sistemas lineares possui terminação finita, onde as direções são obtidas seqüencialmente a cada iteração conservando a propriedade de conjugacidade entre essas direções. Como cada iteração do método corresponde a minimizar uma quadrática sobre a variedade linear gerada pelas direções já disponíveis, então no n-ésimo passo o problema corresponde exatamente a minimizar a função quadrática sobre o $\Re^n$. E portanto $x_n = x^*$. O que há de notável

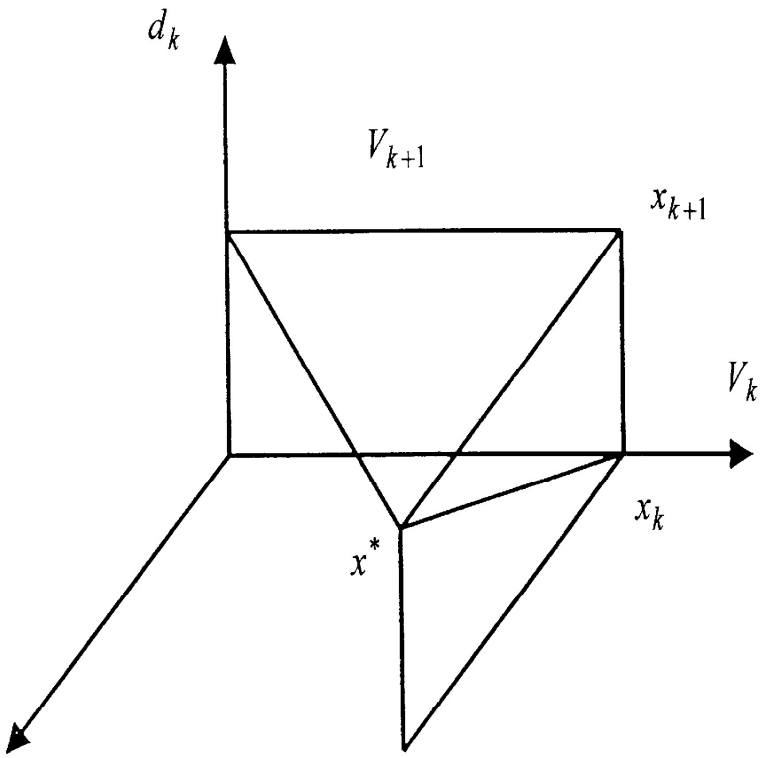

Figura 4.4: Interpretação do Teorema do Subespaço Gerado. [Lue73]

neste método de minimização é que, por realizar multiplicações de vetores e matrizes, requer pouco armazenamento e cada iteração tem um custo computacional pequeno.

Um caso particular das direções conjugadas pode ser obtido usando a direção $d_0 = -\nabla F(x_0)$ e daí em diante tomando como direção uma combinação da direção do gradiente no ponto obtido na iteração corrente e a direção usada na iteração anterior. Esta escolha tem pelo menos três vantagens no uso do método. A primeira vantagem é que enquanto não se estiver na solução ótima o gradiente é diferente de zero, o que torna o método bem definido. Segundo, a maneira de se obter gradientes conjugados é essencialmente simples. A terceira vantagem é que muitas vezes o método faz progressos uniformes na direção de ponto de mínimo da função quando a matriz tem uma distribuição favorável de autovalores.

O algoritmo de Gradiente Conjugado usa como a primeira direção de busca a mesma direção de máxima descida, isto é, $d_0 = -\nabla F(x_0)$. Depois todas as direções são sempre modificações do gradiente na iteração corrente de modo que as direções sejam sempre conjugadas com todas as direções anteriores.

Algoritmo Padrão de Gradiente Conjugado, adaptado de [Lue73]

Dado qualquer $x_0 \in \Re^n$, define-se $d_0 = -g_0 = b - Ax_0$

Enquanto $g_k \neq 0$

$$x_{k+1} = x_k + \alpha_k d_k \tag{4.18}$$

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T A d_k} \tag{4.19}$$

$$d_{k+1} = -g_{k+1} + \beta_k d_k \tag{4.20}$$

$$\beta_k = \frac{g_{k+1}^T A d_k}{d_k^T A d_k} \tag{4.21}$$

Fim

onde $g_k = Ax_k - b$.

O Teorema do Subespaço Gerado garante que o gradiente $\nabla F(x_k)$ ou g_k é ortogonal a $d_0, d_1, \ldots, d_{k-1}$. Escrevendo a equação (4.20) para k obtém-se:

$$d_k = -g_k + \beta_{k-1} d_{k-1} \tag{4.22}$$

Se pré-multiplicar a equação (4.22) por $-g_k^T A$ obtém-se

$$-g_k^T A d_k = g_k^T A g_k - \beta_{k-1} d_{k-1} A g_k^T \tag{4.23}$$

como no primeiro passo $d_0 = -g_0, d_1 = g_1$, para $k = 1$, quando $k > 1$ o segundo termo do lado direito da equação (4.23) desaparecerá devido a propriedade de que $g_{k-1} A g_k^T = 0$, logo

$$-g_k^T d_k = g_k^T g_k \tag{4.24}$$

isto nos permite escrever α_k da equação (4.19) como:

$$\alpha_k = \frac{g_k^T g_k}{d_k^T A d_k} \tag{4.25}$$

ou ainda

$$\alpha_k = \frac{\left\|g_k\right\|^2}{d_k^T A d_k} \tag{4.26}$$

lembrando que

$$\left\|g_k\right\|^2 = d_k^T g_k = (g_k + \beta_{k-1} d_{k-1})^T g_k \tag{4.27}$$

Usando novamente o Teorema do Subespaço Gerado, sabe-se que o gradiente $\nabla F(x_k)$ ou g_k é ortogonal a $d_0, d_1, \ldots, d_{k-1}$, por conjugacidade sabe-se que $d_i^T A d_j = 0$, para $i \neq j$, logo

$$d_k^T A d_{k+1} = 0, \tag{4.28}$$

substituindo a equação (4.20) em (4.28) tem-se:

$$d_k^T A(-g_{k+1} + \beta_k d_k) = 0 \tag{4.29}$$

$$-d_k^T A g_{k+1} + d_k^T A \beta_k d_k = 0 \tag{4.30}$$

$$\beta_k = \frac{d_k^T A g_{k+1}}{d_k^T A d_k} \tag{4.31}$$

Por definição de gradiente sabe-se que o gradiente g_{k+1} no ponto x_{k+1} é:

$$g_{k+1} = b - A x_{k+1} \tag{4.32}$$

E substituindo a equação (4.18) na equação (4.32) tem-se que:

$$\begin{aligned} g_{k+1} &= b - A(x_k + \alpha_k d_k) \\ &= b - A x_k - \alpha_k A d_k \end{aligned}$$

$$= g_k - \alpha_k A d_k \tag{4.33}$$

Como as direções são conjugadas com os gradientes em relação a matriz simétrica e definida positiva A, e usando o relacionamento:

$$\left\|g_{k+1}\right\|^2 = g_{k+1}^T g_{k+1} \tag{4.34}$$

$$= (g_k + \alpha_k A d_k)^T g_{k+1} \tag{4.35}$$

$$= (g_k^T g_{k+1} + \alpha_k A d_k)^T g_{k+1} \tag{4.36}$$

$$= g_k^T g_{k+1} + \beta_k g_k^T g_k \tag{4.37}$$

donde

$$\beta_k = \frac{\left\|g_{k+1}\right\|^2 - g_k^T g_{k+1}}{g_k^T g_k} \tag{4.38}$$

e como

$$g_k^T g_{k+1} = 0 \quad \text{para} \quad k = 1 : m-1 \tag{4.39}$$

portanto pode-se escrever β_k como:

$$\beta_k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \tag{4.40}$$

O Algoritmo Padrão do Método de Gradiente Conjugado com as modificações está a seguir:

Algoritmo de Gradiente Conjugado com modificações [NW99].

Dado qualquer $x_0 \in \Re^n$, define-se $d_0 = -g_0 \;\; g_0 = Ax_0 - b$ e

Enquanto $g_k \neq 0$

$$\alpha_k = \frac{g_k^T g_k}{d_k^T A d_k} \tag{4.41}$$

$$x_{k+1} = x_k + \alpha_k d_k \tag{4.42}$$

$$g_{k+1} = g_k + \alpha_k A d_k \tag{4.43}$$

$$\beta_k = \frac{(g_{k+1})^T g_{k+1}}{g_k^T g_k} \tag{4.44}$$

$$d_{k+1} = -g_{k+1} + \beta_k d_k \tag{4.45}$$

$$k = k + 1$$

Fim.

No caso da matriz A possuir r distintos autovalores, então o mínimo da solução é determinado em no máximo r iterações. Geralmente, se os autovalores da matriz A ocorrem em r distintos agrupamentos, o Método de Gradiente Conjugado dará uma boa solução do problema em r iterações. A convergência baseada no número de condição da matriz A, $cond(A)$, (calculada com relação a norma Euclidiana) é dada por

$$\|x_k - x^*\|_A \leq \left(\frac{\sqrt{cond(A)} - 1}{\sqrt{cond(A)} + 1} \right)^{2k} \|x_0 - x^*\|_A \,. \tag{4.46}$$

onde $cond(A)$ é:

$$cond(A) = \|A\|_2 \left\|A^{-1}\right\|_2 = \frac{\lambda_1}{\lambda_n}$$

de modo que, λ_1 são os maiores autovalores da matriz A e λ_n são os menores autovalores da matriz A.

4.1.3 Método do gradiente conjugado pré-condicionado

O método de Gradiente Conjugado funciona bem com matrizes bem condicionadas ou que possuem poucos autovalores distintos. Quando não é esse o caso, é possível acelerar a convergência deste método de Gradiente Conjugado usando um pré-condicionamento (PCG) do sistema procurando-se melhorar a distribuição dos autovalores da matriz A. Considerando o problema de resolver o sistema linear

$$Ax = b,$$

onde A é uma matriz não singular e fazendo a substituição de variável

$$y = Cx, \tag{4.47}$$

onde C é uma matriz definida positiva, o sistema equivalente resulta,

$$AC^{-1}y = b \tag{4.48}$$

Pré-multiplicando os dois membros da equação (4.48) por C^{-1} tem-se,

$$C^{-1}AC^{-1}y = C^{-1}b$$

A matriz C deve satisfazer as seguintes propriedades:

i. C deve ser não singular;

ii. $C^{-1}AC^{-1}$ deve ser bem condicionada;

iii. Os autovalores de $C^{-1}AC^{-1}$ devem ter uma distribuição favorável.

Na função quadrática $F(x) = \frac{1}{2}x^TAx - b^Tx + c$, fazendo a substituição de variável por (4.47), tem-se

$$F(y) = \frac{1}{2}y^T(C^{-T}AC^{-1})y - (C^{-1}b)^Ty + c. \tag{4.49}$$

Minimizar $F(y)$ equivale a resolver o sistema linear

$$C^{-T}AC^{-1}y = C^{-1}b$$

A convergência do algoritmo de Gradiente Conjugado para este problema, agora dependerá dos autovalores da matriz $C^{-T}AC^{-1}$.

Procura-se escolher uma matriz C de forma que o número de condição da matriz $C^{-T}AC^{-1}$ seja muito menor do que o número de condição da matriz A, ou seja, que a constante da equação (4.46) seja pequena. Também pode-se escolher C de maneira que os autovalores da matriz $C^{-T}AC^{-1}$ sejam agrupados, pois o número de iterações para determinar uma boa aproximação da solução não é maior do que o número de grupos de autovalores agrupados [NW99].

Definindo uma matriz precondicionadora $M = C^2 = C^TC$ e fazendo $My_k = g_k$, o algoritmo a seguir não usa a matriz C explicitamente, mas a matriz $M = C^TC$ a qual é simétrica e definida positiva por construção.

Algoritmo de Gradiente Conjugado Pré-condicionado [NW99].

Dado x_0, e a matriz $M = C^T C$ pré-condicionadora
Calcule $My_0 = g_0$ para y_0;
$d_0 = -g_0$, $k = 0$;
Enquanto $g_k \neq 0$

$$\alpha_k = \frac{g_k^T y_k}{d_k^T A d_k}; \tag{4.50}$$

$$x_{k+1} = x_k + \alpha_k d_k; \tag{4.51}$$

$$g_{k+1} = g_k + \alpha_k A d_k; \tag{4.52}$$

$$My_{k+1} = g_{k+1}; \tag{4.53}$$

$$\beta_k = \frac{g_{k+1}^T y_{k+1}}{g_k^T y_k}; \tag{4.54}$$

$$d_{k+1} = -y_{k+1} + \beta_{k+1} d_k; \tag{4.55}$$

$$k = k + 1; \tag{4.56}$$

Fim.

Outra versão do Algoritmo de Gradiente Conjugado Pré-condicionado é dado por [GL96]. Dada a matriz simétrica e definida positiva A, o vetor b, e uma matriz simétrica definida positiva pré-condicionadora M, $x_0 \in \Re$, tem-se

Algoritmo de Gradiente Conjugado Pré-condicionado, [GL96]

$k = 0$
$g_0 = b - Ax_0$
Enquanto $g_k \neq 0$
Resolver $Mz_k = g_k$
$k = k + 1$

Se $k = 1$

$d_1 = d_0$

senão

$\beta_k = \frac{g_{k+1}^T z_{k-1}}{g_{k+2}^T z_{k-2}}$

$d_k = z_{k-1} + \beta_k d_{k-1}$

fim

$\alpha_k = \frac{g_{k+1}^T z_{k-1}}{d_k^T A d_k}$

$x_k = x_{k-1} + \alpha_k d_k$

$g_k = g_{k-1} + \alpha_k A d_k$

fim

$x = x_k$

É importante observar que os gradientes e as direções satisfazem:

$$g_i^T M^{-1} g_j = 0, \qquad \forall i \neq j$$

$$d_j^T (C^{-1} A C^{-1}) d_i = 0, \qquad \forall i \neq j\,.$$

A diferença entre o Método de Gradiente Conjugado Pré-condicionado e não Pré-condicionado em termos de esforço computacional é a necessidade de resolver os sistemas do tipo $My_k = g_k$, onde M é uma matriz simétrica e é decomposta uma única vez, a partir da segunda vez o custo computacional é da ordem de n^2.

Enfatiza-se mais uma vez, que para a ocorrência do sucesso dessa abordagem, a escolha da matriz C deve ser de modo que $C^{-1}AC^{-1}$, ou seja bem condicionada, ou tenha autovalores agrupados. Existem várias técnicas pré-condicionadoras, conforme [NW99], mas a técnica que parece ser a de maior sucesso é a de fatoração incompleta de Cholesky.

Conforme [NW99] a idéia fundamental para a fatoração incompleta de Cholesky é seguir o procedimento de Cholesky, mas em vez de computar o fator exato (L) de Cholesky que satisfaça $A = LL^T$, computa-se uma aproximação do fator, $\tilde{L}$ que conserva a esparsidade da matriz A. Usualmente, se A é não densa requer-se que $\tilde{L}$ seja também não densa ou não mais denso que os fatores L da decomposição de Cholesky da matriz A. Tem-se que $A \cong \tilde{L}\tilde{L}^T$, fazendo a escolha de C, $C = \tilde{L}^T$, obtém-se $M = \tilde{L}\tilde{L}^T$ e então:

$$C^{-T} A C^{-1} = \tilde{L}^{-1} A \tilde{L}^{-T} \cong I$$

desta forma a distribuição dos autovalores da matriz $C^{-T}AC^{-1}$ é favorável.

É importante observar que se for feita a decomposição de Cholesky de $A = LL^T$, a fatoração QR de C, $C = QV^T$, então

$$C^{-1}AC^{-1} = C^{-T}AC^{-1} = \left(VQ^T\right)^{-1} A \left(QV^T\right)^{-1} = Q\left(V^{-1}AR^{-T}\right)Q^T \tag{4.57}$$

donde então,

$$C^{-1}AC^{-1} = Q\left(V^{-1}LL^TV^{-T}\right)Q^T \cong I \tag{4.58}$$

Quanto melhor V aproxima L melhor o problema é condicionado. Para manter a esparsidade na decomposição de Cholesky da matriz A deve-se fazer $l_{ij} = 0$ quando $a_{ij} = 0$, e portanto tem-se apenas a decomposição $\tilde{L}\tilde{L}^T \cong A$.

O seguinte algoritmo produz uma decomposição incompleta de Cholesky.

Fatoração Incompleta de Cholesky adaptado de [RdRL96].

Para $k = 1 : n$
  $soma = 0$
  para $j = 1 : k - 1$
    $soma = soma + a(k, j) * a(k, j)$
  fim
  $r = a(k,k) - soma$
    se $r \leq 0$
    Pare, a matriz não é definida positiva
    fim
  $a(k,k) = \sqrt{r}$
    para $i = k + 1 : n$
      $soma = 0$
        para $j = 1 : k - 1$
          $soma = soma + a(i, j) * a(k, j)$
        fim
          if $a(i,k) \neq 0$
            $a(i,k) = (a(i,k) - soma)/a(k,k)$
          fim
    fim

Fim.

Se em alguma etapa $r \leqslant 0$, o algoritmo falha e o processo é interrompido e conseqüentemente a matriz A não é definida positiva. Se uma matriz A é definida positiva então a decomposição de Cholesky de A retorna uma matriz triangular superior tal que $R^T R = A$.

No próximo Capítulo apresenta-se como foram desenvolvidos os experimentos numéricos.

Capítulo 5

Experimentos Numéricos

Neste Capítulo faz-se a descrição de como foram elaborados os experimentos numéricos.

5.1 O problema de minimizar uma função quadrática

Supondo que a solução do problema de

$$\underset{\substack{s/a \\ Ax=b}}{\text{Minimizar}} \quad F(x) = \frac{1}{2}x^T Gx + h^T x + q, \tag{5.1}$$

é aproximado pelo modelo quadrático

$$\underset{d\in\Re^{n-m}}{\text{Minimizar}}\, q(d) = f(x_k) + \frac{1}{2}\nabla f(x_k)^T Zd + \frac{1}{2}d^T Z^T \nabla^2 f(x_k) Zd. \tag{5.2}$$

o que permite sempre obter soluções viáveis. Sejam $\bar{x}$ e $\hat{x}$ pontos viáveis das restrições, por linearidade tem-se $A(\bar{x} - \hat{x}) = 0$, desde que $A\bar{x} = b$ e $A\hat{x} = b$. O passo d de algum ponto viável para outro ponto viável deve ser então ortogonal às linhas de A, ou seja, satisfazer:

$$Ad = 0 \tag{5.3}$$

Então se d é uma direção viável com respeito as restrições de igualdade, qualquer passo ao longo dessa direção não viola as restrições, desde que $A(\hat{x} + \alpha d) = A\hat{x} = b$. O espaço nulo de A possui dimensão $n - m$, precisa-se então obter uma base para este espaço, sendo denominada de Z, então $AZ = 0$, e toda direção viável pode ser escrita

como uma combinação linear das colunas de Z, ou seja, pode-se escrever $d = Zd_z$, para algum d_z.

O objetivo desse trabalho é obter diferentes bases para $\aleph(A)$ e analisar o impacto sobre o comportamento da seqüência gerada pelo algoritmo de Gradiente Conjugado Pré-condicionado. Essas bases provém da decomposição QR da matriz A^T e do particionamento da matriz A em matrizes B e N (básicas e não básicas) e das permutações das linhas de A obtendo novas bases por decomposição QR, ou mais precisamente usando o comando $null(A)$ do *software* Matlab da empresa *Mathworks* e do particionamento da matriz A em $(B|N)$, que pode ser obtido pelo comando null(A,'r') do Matlab.

Para minimizar (5.2) é necessário resolver o sistema:

$$Z^T \nabla^2 f(x_k) Z d_z = -Z^T \nabla f(x_k). \tag{5.4}$$

Substituindo $\nabla^2 f(x_k)$, no sistema (5.4), por uma aproximação definida positiva G, o sistema a ser resolvido fica então:

$$Z^T G Z d_z = -Z^T \nabla f(x_k). \tag{5.5}$$

Como as colunas da matriz Z são linearmente independentes e G é definida positiva, então a matriz $Z^T GZ$ é definida positiva. Desse modo o sistema (5.5) possui solução única e será resolvido pelos métodos de Gradiente Conjugado e Gradiente Conjugado Pré-condicionado. Resolver o problema sem pré-condicionador significa obter o produto $Z^T GZ$, com o uso de pré-condicionadores será necessário pré-condicionar a matriz $Z^T GZ$.

5.2 Resolução do problema pelo método de gradiente conjugado

Utilizou-se matrizes G esparsas com aproximadamente 75 por cento de zeros e A com elementos entre 10^{-6} e 10^{6}.

Supondo disponível um ponto inicial viável x_0 que satisfaça as restrições lineares de igualdade,

$$Ax_0 = b \tag{5.6}$$

uma base para o núcleo da matriz das restrições, Z, que é obtido pelas últimas $n - m$

colunas da matriz ortogonal Q da decomposição QR da matriz A^T, e do particionamento da matriz A em $(B|N)$, ou ainda mais precisamente pelo comando $(null(A,'r'))$ do *software* Matlab, ou seja,

$$[qr] = QR(A^T)$$

onde Z

$$Z = Q(:, [m+1:n])$$

e

$$Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}$$

Então calcula-se o gradiente da função a ser minimizada no ponto x_0,

$$\nabla F(x_0) = Gx_0 + h$$

bem como a Hessiana projetada,

$$B = Z^T GZ$$

e o cálculo do Gradiente projetado.

$$Z^T \nabla F(x_0)$$

Resolve-se o sistema (5.5) usando método de Gradiente Conjugado adaptado para obter uma direção viável ($d_z = x_k$), a qual deve pertencer ao núcleo de A, $\aleph(A)$. Para tanto, inicializa-se x_k com um vetor nulo, a tolerância de 10^{-6}, o gradiente $g_k k = \nabla f(x_0) + Z^T GZxk$. Calcula-se o resíduo que é a norma de g_k a direção de descida, d_k pelo gradiente negativo, ou seja, $d_k = -(Z^T GZx_k + \nabla f(x_0))$, onde o método fará o cálculo do tamanho do passo α_k, do parâmetro β_k, atualizando d_k e x_k. Ao final das iterações atualiza-se o ponto pela soma do ponto inicial x_0 com a direção x_k, (como visto anteriormente em (3.18)), obtida no método de Gradiente Conjugado multiplicado por Z.

$$x = x_0 + Zx_k$$

Permutam-se de forma aleatória as colunas de Z (ortogonal) obtendo-se novas bases e repete-se o processo.

Finalmente verifica-se se as seguintes condições são satisfeitas:

1. $Ax^* = b$.
2. $Z^T \nabla f(x^*) = 0$

Ao final o programa apresenta uma tabela, com as informações das iterações, k, os respectivos resíduos e as taxas de convergência, $\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|}$.

5.3 Resolução do problema pelo método de gradiente conjugado pré-condicionado

Considerando o problema de resolver o sistema

$$Z^T G Z d = -Z^T \nabla f(x_k). \tag{5.7}$$

onde $Z^T GZ$ é uma matriz simétrica definida positiva. Fazendo a matriz M pré-condicionadora tal que $M = C^T C$ tem-se

$$C^T (Z^T GZ) C \cong I. \tag{5.8}$$

Computa-se a decomposição de Cholesky incompleta da matriz $Z^T GZ$, obtendo $Z^T GZ = LL^T$, então

$$Z^T GZ = LL^T. \tag{5.9}$$

Fazendo

$$C = L^{-1} \tag{5.10}$$

Define-se a matriz pré-condicionadora M, da seguinte forma:

$$M = C^T C = L^{-T} L^{-1} \tag{5.11}$$

Da mesma forma que a anterior, utilizou-se matrizes G esparsas com aproximadamente 75 por cento de zeros e A com elementos entre 10^{-6} e 10^6.

Supondo disponível um ponto inicial viável x_0 que satisfaça as restrições lineares de igualdade,

$$Ax_0 = b \tag{5.12}$$

uma base para o núcleo da matriz das restrições, Z que é obtido pelas últimas $n - m$ colunas da matriz ortogonal Q da decomposição QR da matriz A^T, e do particionamento

da matriz A em $(B|N)$, ou ainda mais precisamente pelo comando $(null(A,'r'))$ do *software* Matlab, ou seja,

$$[qr] = QR(A^T$$

onde Z

$$Z = Q(:, [m+1:n])$$

e

$$Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}$$

Então calcula-se o gradiente da função a ser minimizada no ponto x_0,

$$\nabla F(x_0) = Gx_0 + h$$

bem como a Hessiana projetada,

$$B = Z^T GZ$$

e o cálculo do Gradiente projetado.

$$Z^T \nabla F(x_0)$$

Resolve-se o sistema (5.7) usando método de Gradiente Conjugado Pré-condicionado adaptado para obter uma direção viável ($d_z = x_k$), a qual deve pertencer ao núcleo de A, $\aleph(A)$. Para tanto, faz-se a decomposição de Cholesky Incompleta da Hessiana projetada. Calcula-se a matriz pré-condicionadora

$$MP = L^{-T}L^{-1}$$

Inicializa-se x_k com um vetor nulo, a tolerância de 10^{-6}, o gradiente $g_k = \nabla f(x_0) + Z^T GZxk$. Calcula-se o resíduo que é a norma de g_k a direção de descida, d_k pelo gradiente negativo, ou seja, $d_k = -(Z^T GZx_k + \nabla f(x_0))$. Resolve-se o sistema $Mz_k = r_k$, de forma que:

$$z_k = (MP)^{-1} r_k \tag{5.13}$$

Onde então o método fará o cálculo do tamanho do passo α_k, do parâmetro β_k, atualizando d_k e x_k. Ao final das iterações atualiza-se o ponto pela soma do ponto inicial x_0 com a direção x_k (como visto anteriormente em (3.18)), obtida no método de Gradiente

Conjugado multiplicado por Z.

$$x = x_0 + Zx_k$$

Permutam-se de forma aleatória as colunas de Z (ortogonal) obtendo-se novas bases e repete-se o processo.

Finalmente verifica-se se as seguintes condições são satisfeitas:

1. $Ax^* = b$.
2. $Z^T \nabla f(x^*) = 0$

Ao final o programa apresenta uma tabela, com as informações das iterações, k, os respectivos resíduos e as taxas de convergência, $\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|}$.

5.4 Algoritmo para solução do problema quadrático

1) Calcule uma base, Z, para o núcleo de A;
2) Calcule o gradiente da quadrática no ponto inicial viável, $\nabla F(x_0) = Gx_0 + h$;
3) Calcule a Hessiana reduzida, $B = Z^T GZ$;
4) Calcule o gradiente reduzido, $Z^T \nabla F(x_0)$;
5) Use os algoritmos *CG* e *PCG* para obter $d = Zd_z$;
6) Calcule a solução $x^* = x_0 + d$;
7) Verifique as condições: $Ax^* = b$ e $Z^T \nabla f(x^*) = 0$;
8) Permute as colunas de Z e repita o processo.

5.5 Relação das matrizes e das permutações utilizadas nos experimentos numéricos

Para as bases, Z, ortogonais realizou-se dez permutações aleatórias das colunas da base. O número de permutações possíveis é $n!$ (n=colunas). As matrizes A relatadas a seguir foram construídas de forma que possuam elementos entre 10^{-6} e 10^{6}.

5.5.1 Matriz A com tamanho 2x7

Para os dois métodos, *CG* e *PCG*, com relação as bases ortogonais, verificou-se que o produto de matriz A pela base Z ficou próximo de zero, na ordem de 10^{-15}. A condição

de otimalidade, $Ax = b$, foi verificada de modo que $Ax - b$ ficou também próximo de zero, na ordem de 10^{-9}. As ordens de permutações das colunas realizadas foram:

1. $1 - 2 - 4 - 3 - 5$
2. $5 - 1 - 2 - 4 - 3$
3. $4 - 1 - 5 - 2 - 3$
4. $4 - 2 - 1 - 3 - 5$
5. $5 - 4 - 1 - 2 - 3$
6. $4 - 2 - 1 - 5 - 3$
7. $2 - 3 - 4 - 1 - 5$
8. $3 - 5 - 4 - 1 - 2$
9. $4 - 5 - 1 - 2 - 3$
10. $2 - 4 - 3 - 1 - 4$

O número de condição das matrizes Z, G e Z^TGZ foram:
$cond(Z) = 1.00000000000000$;
$cond(G) = 1.344719538528939.10^3$;
$cond(Z^TGZ) = 1.102584944439224.10^3$.
O número de iterações para CG foi de 6 e para PCG foi de 7.

Para a base não ortogonal, verificou-se que o produto de matriz A pela base Z ficou próximo de zero, na ordem de 10^{-10}. A condição de otimalidade, $Ax = b$, foi verificada de modo que $Ax - b$ ficou também próximo de zero, na ordem de 10^{-9}.

O número de condição das matrizes Z, G e Z^TGZ foram:
$cond(Z) = 1.873703898745359e.10^{12}$;
$cond(G) = 1.344719538528939.10^3$;
$cond(Z^TGZ) = 4.793699559451190.10^{24}$.
O número de iterações para CG foi de 14 e para PCG foi de 22.

Para ambos os métodos e todas as bases usadas os valores da função mantiveram-se os mesmos, sendo no ponto x_0 o valor $3.36692721000000.10^4$ e no ponto solução, $0.03259053954595.10^4$.

5.5.2 Matriz A com tamanho 5x10

Para as bases ortogonais, verificou-se que o produto de matriz A pela base Z ficou próximo de zero, na ordem de 10^{-9}, para ambos os métodos. A condição de otimalidade, $Ax = b$, foi verificada de modo que $Ax - b$ ficou também próximo de zero, na ordem de 10^{-8}, para o método de CG e 10^{-7} para o método de PCG. Para os dois métodos as ordens de permutações das colunas realizadas foram:

1. $1 - 3 - 2 - 5 - 4$
2. $1 - 5 - 4 - 2 - 3$
3. $1 - 2 - 5 - 3 - 4$
4. $5 - 2 - 3 - 4 - 1$
5. $5 - 4 - 3 - 1 - 2$
6. $1 - 5 - 3 - 4 - 2$
7. $5 - 1 - 2 - 3 - 4$
8. $5 - 4 - 2 - 1 - 3$
9. $5 - 4 - 1 - 2 - 3$
10. $3 - 5 - 1 - 2 - 4$

O número de condição das matrizes Z, G e Z^TGZ foram:
$cond(Z) = 1.00000000000000$;
$cond(G) = 1.685646961620519.10^6$;
$cond(Z^TGZ) = 1.623526583064040.10^6$.
O número de iterações para CG foi de 5 e para PCG foi de 6.

Para a base não ortogonal, verificou-se que o produto de matriz A pela base Z ficou próximo de zero, na ordem de 10^{-9}. A condição de otimalidade, $Ax = b$, foi verificada de modo que $Ax - b$ ficou também próximo de zero, na ordem de 10^{-10} para o método de CG e 10^{-7} para PCG.

O número de condição das matrizes Z, G e Z^TGZ foram:
$cond(Z) = 1.752541264399723.10^6$;
$cond(G) = 1.685646961620519.10^6$;

$cond(Z^TGZ) = 3.604361461803635.10^{12}$.

O número de iterações para *CG* foi de 16 e para *PCG* foi de 21.

Para ambos os métodos e todas as bases usadas os valores da função mantiveram-se os mesmos, sendo no ponto x_0 o valor $5.15575867311631.10^5$ e no ponto solução, $0.03943356934031.10^5$.

5.5.3 Matriz A com tamanho 10x20

Para as bases ortogonais, verificou-se que o produto de matriz A pela base Z ficou próximo de zero, na ordem de 10^{-9}. A condição de otimalidade, $Ax = b$, foi verificada de modo que $Ax - b$ ficou também próximo de zero, na ordem de 10^{-8}, para o método de *CG* e entre 10^{-7} a 10^{-8} para o método de *PCG* . Para os dois métodos as ordens de permutações das colunas realizadas foram:

1. $8 - 6 - 4 - 1 - 2 - 3 - 9 - 10 - 5 - 7$
2. $10 - 9 - 1 - 5 - 8 - 7 - 2 - 3 - 4 - 6$
3. $2 - 7 - 5 - 4 - 8 - 6 - 3 - 10 - 1 - 9$
4. $2 - 7 - 3 - 1 - 4 - 9 - 8 - 10 - 5 - 6$
5. $1 - 8 - 7 - 5 - 2 - 4 - 10 - 9 - 3 - 6$
6. $4 - 6 - 5 - 1 - 9 - 7 - 8 - 10 - 2 - 3$
7. $1 - 5 - 9 - 8 - 6 - 2 - 7 - 10 - 3 - 4$
8. $7 - 10 - 3 - 2 - 1 - 5 - 6 - 8 - 9 - 4$
9. $3 - 5 - 6 - 10 - 4 - 9 - 2 - 8 - 7 - 1$
10. $5 - 8 - 9 - 1 - 2 - 10 - 6 - 4 - 7 - 3$

O número de condição das matrizes Z, G e Z^TGZ foram:
$cond(Z) = 1.00000000000000$;
$cond(G) = 2.026135983191580.10^6$;
$cond(Z^TGZ) = 1.57071321694447$.
O número de iterações para *CG* foi de 7 e para *PCG* foi de 27.
Para a base não ortogonal não houve convergência.

Para ambos os métodos e todas as bases usadas os valores da função mantiveram-se os mesmos, sendo no ponto x_0 o valor $2.49663368431454.10^8$ e no ponto solução, $2.28475770214419.10^8$.

5.5.4 Matriz A com tamanho 20x50

Para as bases ortogonais, verificou-se que o produto de matriz A pela base Z ficou próximo de zero, na ordem de 10^{-9}. A condição de otimalidade, $Ax = b$, foi verificada de modo que $Ax - b$ ficou também próximo de zero, na ordem de 10^{-6}, para o método de CG e entre 10^{-6} a 10^{-7} para o método de PCG . Para os dois métodos as ordens de permutações das colunas realizadas foram:

1. 7 – 3 – 8 – 25 – 10 – 6 – 24 – 29 – 11 – 27 – 21 – 2 – 30 – 22 – 18 – 1 – 5 – 13 – 12 – 28 – 26 – 4 – 20 – 15 – 23 – 14 – 17 – 16 – 19 – 9

2. 7 – 20 – 10 – 26 – 3 – 2 – 27 – 24 – 6 – 4 – 29 – 14 – 21 – 25 – 11 – 5 – 19 – 12 – 18 – 1 – 15 – 13 – 17 – 30 – 28 – 9 – 23 – 22 – 8 – 16

3. 20 – 14 – 17 – 3 – 5 – 15 – 28 – 18 – 1 – 30 – 7 – 27 – 13 – 19 – 24 – 16 – 22 – 21 – 9 – 25 – 2 – 6 – 12 – 10 – 11 – 29 – 23 – 4 – 26 – 8

4. 7 – 22 – 30 – 12 – 5 – 29 – 17 – 4 – 9 – 11 – 1 – 26 – 20 – 15 – 25 – 10 – 19 – 6 – 21 – 2 – 13 – 16 – 23 – 24 – 18 – 28 – 27 – 8 – 14 – 3

5. 4 – 28 – 5 – 20 – 17 – 30 – 29 – 6 – 27 – 12 – 9 – 1 – 3 – 13 – 16 – 11 – 15 – 26 – 24 – 14 – 19 – 2 – 8 – 25 – 10 – 22 – 18 – 7 – 23 – 21

6. 5 – 2 – 9 – 28 – 20 – 21 – 3 – 6 – 19 – 13 – 25 – 7 – 26 – 24 – 15 – 17 – 30 – 12 – 29 – 8 – 4 – 27 – 23 – 16 – 22 – 11 – 14 – 10 – 18 – 1

7. 2 – 20 – 15 – 22 – 14 – 26 – 28 – 24 – 18 – 4 – 6 – 19 – 21 – 30 – 23 – 1 – 25 – 16 – 12 – 8 – 13 – 17 – 29 – 5 – 27 – 10 – 3 – 9 – 7 – 11

8. 1 – 7 – 3 – 12 – 8 – 23 – 4 – 25 – 6 – 13 – 11 – 29 – 26 – 2 – 28 – 10 – 19 – 9 – 5 – 21 – 15 – 18 – 27 – 20 – 24 – 16 – 17 – 22 – 14 – 30

9. 28 – 25 – 15 – 3 – 23 – 27 – 17 – 18 – 24 – 11 – 30 – 9 – 29 – 22 – 5 – 7 – 2 – 12 – 26 – 14 – 6 – 13 – 21 – 4 – 20 – 10 – 16 – 19 – 1 – 8

10. 20 – 16 – 24 – 27 – 10 – 13 – 8 – 25 – 5 – 28 – 26 – 12 – 9 – 18 – 19 – 30 – 29 – 17 – 11 – 15 – 22 – 1 – 6 – 14 – 3 – 4 – 7 – 23 – 21 – 2

O número de condição das matrizes Z, G e Z^TGZ foram:
$cond(Z) = 1.00000000000000$;
$cond(G) = 1.432201801491824.10^3$;
$cond(Z^TGZ) = 8.244408255539313.10^2$.
O número de iterações para CG foi de 13 e para PCG foi de 60.
Para a base não ortogonal não houve convergência.
Para ambos os métodos e todas as bases usadas os valores da função mantiveram-se os mesmos, sendo no ponto x_0 o valor $1.038240638723 15.10^7$ e no ponto solução, $0.009397501492 32.10^7$.

5.5.5 Matriz A com tamanho 50x100

Para as bases ortogonais, verificou-se que o produto de matriz A pela base Z ficou próximo de zero, na ordem de 10^{-9}. A condição de otimalidade, $Ax = b$, foi verificada de modo que $Ax - b$ ficou também próximo de zero, na ordem de 10^{-6}, para ambos os métodos. Para os dois métodos as ordens de permutações das colunas realizadas foram:

1. 37 – 18 – 34 – 13 – 27 – 6 – 35 – 46 – 47 – 3 – 31 – 29 – 36 – 5 – 26 – 12 – 41 – 45 – 28 – 42 – 49 – 40 – 1 – 15 – 48 – 43 – 11 – 38 – 25 – 32 – 9 – 7 – 19 – 33 – 17 – 30 – 16 – 4 – 23 – 21 – 44 – 2 – 8 – 39 – 20 – 14 – 10 – 50 – 22 – 24.

2. 27 – 49 – 29 – 36 – 38 – 48 – 18 – 23 – 35 – 19 – 30 – 42 – 4 – 11 – 39 – 16 – 45 – 6 – 28 – 41 – 22 – 25 – 37 – 26 – 31 – 8 – 5 – 3 – 40 – 34 – 32 – 1 – 2 – 13 – 15 – 33 – 44 – 9 – 14 – 46 – 24 – 20 – 21 – 12 – 50 – 17 – 10 – 43 – 47 – 7.

3. 11 – 15 – 3 – 19 – 2 – 22 – 27 – 32 – 5 – 4 – 42 – 28 – 45 – 23 – 39 – 24 – 36 – 25 – 7 – 20 – 50 – 44 – 35 – 30 – 41 – 31 – 37 – 13 – 8 – 38 – 18 – 14 – 29 – 16 – 46 – 33 – 26 – 10 – 21 – 9 – 6 – 43 – 47 – 49 – 1 – 34 – 48 – 40 – 17 – 12.

4. 28 – 31 – 43 – 29 – 12 – 18 – 48 – 39 – 45 – 27 – 3 – 24 – 37 – 21 – 23 – 22 – 32 – 42 – 33 – 41 – 19 – 2 – 35 – 36 – 4 – 5 – 46 – 47 – 44 – 7 – 30 – 10 – 11 – 9 – 15 – 40 – 14 – 6 – 20 – 34 – 49 – 17 – 38 – 8 – 16 – 13 – 50 – 25 – 26 – 1.

5. 28 – 50 – 11 – 45 – 5 – 21 – 4 – 48 – 32 – 24 – 36 – 20 – 7 – 37 – 14 – 46 – 39 – 33 – 35 – 25 – 47 – 34 – 18 – 8 – 41 – 13 – 6 – 9 – 15 – 17 – 3 – 16 – 49 – 22 – 26 – 23 – 44 – 27 – 43 – 31 – 42 – 30 – 2 – 40 – 19 – 10 – 38 – 12 – 1 – 29.

6. 38 – 9 – 11 – 29 – 50 – 36 – 2 – 4 – 49 – 41 – 1 – 18 – 7 – 31 – 35 – 39 – 34 – 30 – 44 – 43 – 26 – 17 – 24 – 19 – 12 – 6 – 16 – 10 – 23 – 13 – 22 – 40 – 3 – 21 – 15 – 33 – 25 – 48 – 46 – 8 – 27 – 37 – 20 – 5 – 32 – 47 – 42 – 28 – 14 – 45.

7. 26 – 36 – 35 – 9 – 28 – 30 – 31 – 2 – 40 – 22 – 24 – 1 – 6 – 16 – 7 – 14 – 42 – 25 – 46 – 48 – 49 – 38 – 33 – 20 – 3 – 8 – 11 – 29 – 13 – 19 – 32 – 27 – 44 – 23 – 34 – 15 – 41 – 45 – 21 – 37 – 12 – 10 – 4 – 18 – 43 – 17 – 39 – 5 – 50 – 47.

8. 9 – 29 – 50 – 23 – 8 – 40 – 19 – 49 – 14 – 39 – 38 – 18 – 21 – 20 – 33 – 15 – 42 – 11 – 25 – 32 – 12 – 7 – 48 – 22 – 47 – 6 – 45 – 36 – 5 – 24 – 3 – 34 – 43 – 10 – 30 – 16 – 26 – 28 – 37 – 1 – 13 – 4 – 31 – 44 – 27 – 46 – 35 – 41 – 17 – 2.

9. 48 – 50 – 8 – 3 – 49 – 10 – 16 – 9 – 14 – 22 – 37 – 20 – 43 – 39 – 32 – 4 – 25 – 45 – 11 – 7 – 47 – 12 – 35 – 27 – 44 – 40 – 46 – 21 – 42 – 13 – 38 – 18 – 41 – 5 – 36 – 30 – 31 – 24 – 19 – 15 – 29 – 23 – 6 – 2 – 17 – 1 – 33 – 28 – 26 – 34.

10. 27 – 16 – 10 – 37 – 17 – 15 – 35 – 44 – 48 – 13 – 25 – 47 – 18 – 26 – 34 – 14 – 11 – 32 – 43 – 36 – 45 – 23 – 4 – 39 – 28 – 42 – 40 – 21 – 24 – 50 – 2 – 46 – 31 – 49 – 30 – 8 – 1 – 5 – 6 – 33 – 22 – 19 – 29 – 41 – 12 – 3 – 38 – 9 – 20 – 7.

O número de condição das matrizes Z, G e Z^TGZ foram:
$cond(Z) = 1.00000000000000$;
$cond(G) = 1.929887307895277.10^6$;
$cond(Z^TGZ) = 6.169111208978302.10^5$.
O número de iterações para *CG* foi de 18 e para *PCG* foi de 230.
Para a base não ortogonal não houve convergência.

Para ambos os métodos e todas as bases usadas os valores da função mantiveram-se os mesmos, sendo no ponto x_0 o valor $1.02390875887600.10^{10}$ e no ponto solução, $0.00002577636305.10^{10}$.

Todos os resultados dos experimentos numéricos encontram-se em disquete que acompanha esta dissertação.

Os significados das expressões abaixo são:
k = número de iterações;

$valorfx_0$ = valor da função no ponto inicial;
$valorfx_k$ = valor da função no ponto solução;
x_0 = ponto inicial;
x = solução x^*;
$\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}$ = taxa de convergência;
$ZTgrad$ = condição de otimalidade ($Z^T\nabla F(x_k)$);
$Ax-b$ = factibilidade, condição de otimalidade $Ax=b$;
$cond(G)$ = número de condição da matriz G;
$cond(Z)$ = número de condição da matriz Z;
$cond(Z^TGZ)$ = número de condição da matriz Z^TGZ.

Capítulo 6

Conclusões

Durante a experimentação numérica os seguintes fatos relevantes foram observados relativamente ao comportamento das seqüências geradas pelo algoritmos.

- Quando a matriz Z cujas colunas formam uma base para o espaço nulo da matriz das restrições A, é ortogonal, o número de condições do sistema

$$Z^T G Z d_z = -Z^T \nabla F(x_k)$$

  não piora, e na prática, na maioria dos casos, melhora;

- Quando a base para o espaço nulo da matriz das restrições A não é ortogonal o número de condições do sistema sempre aumenta;

- Quanto menor for o número de condição da Hessiana da função objetivo, G, melhor é a performance dos algoritmos. Esse fato é mais evidente quando a escolha da base recai sobre uma base ortogonal;

- Quando é usado o algoritmo *PCG* para obtenção do ponto de mínimo, mantida a mesma base, há, em geral, um aumento no número de iterações para a obtenção de uma aproximação da solução com o mesmo critério de tolerância, devido ao fato de estarmos usando uma decomposição incompleta de Cholesky;

- Os métodos do tipo gradientes conjugados pré-condicionados têm sucesso em geral quando

$$\tilde{L}^{-1} Z^T G Z \tilde{L}^{-T} \approxeq I;$$

- Nos casos onde a dimensão do espaço nulo da matriz das restrições é grande a aritmética de ponto flutuante utilizada tem grande importância na capacidade do algoritmo em gerar direções conjugadas. De fato, nesses casos, mesmo usando o algoritmo *PCG*, há uma evidente deterioração na performance do algoritmo, devido a qualidade das direções conjugadas obtidas;

- O uso de diferentes bases ortogonais parecem ter pouca importância no desempenho dos algoritmos de *CG* e *PCG* com a abordagem utilizada nesse trabalho. Sugere-se para trabalhos futuros o uso de diferentes abordagens de pré-condicionamento, [NW99], bem como na escolha de bases para a nulidade da matriz das restrições.

Referências Bibliográficas

[CLTB97] C. L. T. Borges, D. M. Falcão, A. L. G. A. Coutinho: *Utilização de método tipo gradiente conjugado na aceleração do fluxo de potência em computação vetorial*. XIV SNPTEE Seminário Nacional de Produção e Transmissão de Energia Elétrica, Belém -PA, 1997.

[Col94] Coleman, T. F.: *Linearly constrained optimization and projected preconditioned conjugate gradients*. in Proceedings of the Fifth SIAM Conference on Applied Linear Algebra: Philadelphia, páginas 118–122, 1994.

[Col02] Coleman, T. F.: *A new trust region algorithm for equality constrained optimization*. Computation Optimization and Applications, (21):177–199, 2002.

[Cun00] Cunha, C.: *Métodos numéricos para as engenharias e ciências aplicadas*. Editora da Unicamp, São Paulo - São Paulo, 2 edição, 2000.

[CV01] Coleman, T. F. e A. Verma: *A preconditioned conjugate gradient approach to linear equality constrained minimization*. Computer Science Department and Cornell Theory Center, Cornell University, Ithaca, New York, páginas 61–72, 2001.

[DS96] Dennis, J. E. e R. B. Schnabel: *Nunerical methods for unconstrained optimization and nonlinear equations*. SIAM, Philadelphia, 1996.

[Gil91] Gil, A. C.: *Como elaborar projetos de pesquisa*. Atlas, São Paulo, 1991.

[Gil99] Gil, A. C.: *Métodos e técnicas de pesquisa social*. Atlas, São Paulo, 1999.

[GL96] Golub, G. H. e C. F. Van Loan: *Matrix computation*. The John Hopkins University Press Ltd., London, 3 edição, 1996.

[IH03] Igarashi, H. e T. Honma: *Convergence of preconditioned conjugate gradient method applied to driven microwave problems*. IEEE Transactions on Magnetics, 39(3), 2003.

[Lue73] Luenberger, D. G.: *Introduction to linear and nonlinear programming*. Addison-Wesley Publishing Company, Massachusetts, 1973.

[NS93] Nash, S. G. e A. Sofer: *Preconditioning of reduced matrices*. Dept. of Operations Research and Engineering, George Mason University, Tech. Rep., 1993.

[NS96] Nash, S. G. e A. Sofer: *Linear and nonlinear programming*. McGraw-Hill, Singapore, 1996.

[NW99] Nocedal, J. e S. J. Wright: *Numerical optimization*. Springer, New York, 1999.

[PGS92] P. Gill, W. Murray, D. Ponceleon e M. Saunders: *Preconditiones for indefinite systems arising in optimization*. SIAM J. Matrix Anal. Appl., 13:292–311, 1992.

[RdRL96] Ruggiero, M. A. G. e V. L. da R. Lopes: *Cálculo numérico: aspectos teóricos e computacionais*. Makron Books Ltda, 2 edição, 1996.

[RSCY00] R. S. Chen, D. G. Fang, K. F. Tsang e E. K. N. Yung: *Analysis of millimeter wave scattering by an electrically large metallic grating using wavelet-based algebratic multigrid preconditioned CG method*. International Journal of Infrared and Millimeter Waves, 21(9):1541–1560, 2000.

[WMW91] Wright, M. H., W. Murray, e P. E. Walter: *Numerical linear algebra and optimization*. Addison-Wesley Publishing Company, 1991.

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo