ANA PAULA DELOWSKI

MÉTODO DAS DIREÇÕES CONJUGADAS NO NÚCLEO DAS RESTRIÇÕES PARA MINIMIZAÇÃO DE UMA FUNÇÃO QUADRÁTICA SUJEITA A RESTRIÇÕES LINEARES DE IGUALDADE

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.

Área de concentração: Gerência de Produção e Logística
Orientador: Prof. Dr. Raimundo José Borges de Sampaio

Curitiba-PR

ANA PAULA DELOWSKI

MÉTODO DAS DIREÇÕES CONJUGADAS NO NÚCLEO DAS RESTRIÇÕES PARA MINIMIZAÇÃO DE UMA FUNÇÃO QUADRÁTICA SUJEITA A RESTRIÇÕES LINEARES DE IGUALDADE

Curitiba-PR

TERMO DE APROVAÇÃO

ANA PAULA DELOWSKI

MÉTODO DAS DIREÇÕES CONJUGADAS NO NÚCLEO DAS RESTRIÇÕES PARA MINIMIZAÇÃO DE UMA FUNÇÃO QUADRÁTICA SUJEITA A RESTRIÇÕES LINEARES DE IGUALDADE

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Curso de Mestrado em Engenharia de Produção e Sistemas, Programa de Pós-Graduação em Engenharia de Produção e Sistemas, do Centro de Ciências Exatas e de Tecnologia da Pontifícia Universidade Católica do Paraná, pela seguinte banca examinadora:

Presidente da Banca

Prof. Dr. Raimundo José Borges de Sampaio (PPGEPS/PUCPR - Orientador)

Prof. Dr. Marco Antonio Cândido Barbosa (PUCPR - Membro Titular)

Prof. Dr. Ricardo Ferrari Pacheco. (PUCGO - Membro Titular)

A minha família

Agradecimentos

A Deus por mais essa oportunidade, pois, sem o Pai Maior, nada fazemos como nada somos.

Ao Prof. Raimundo J. B. de Sampaio, pela paciência e dedicação na orientação deste trabalho. O verdadeiro mestre é aquele que desprende palavras duras quando necessário e incentiva o crescimento do aluno de forma incondicional. Suas lições serão sempre lembradas.

Aos meus genitores Luiz E. Delowski e Gabriela R. D. Delowski, que me conceberam, os quais com humildade, mas com sabedoria, ensinaram-me os caminhos de uma vida vitoriosa estribada na honestidade.

Ao meu irmão Marcos Paulo Delowski, pela ajuda e compreensão.

Ao meu querido Rafael de Paula Ciniello, insubistituível amigo, amante e companheiro, cujas qualidades primam pela perseverança no trabalho e pela certeza de um sol mais brilhante no amanhã.

Às colegas Viviane C. Bini e Elaine C. P. da Silva, pela colaboração e incentivo.

E, a todos aqueles que, direta ou indiretamente, positiva ou negativamente, me ajudaram a crescer como pessoa e como profissional.

Sumário

Lista de Tabelas

Lista de Siglas

GS	Gram-Schmidt Clássico
CG	Gradiente Conjugado
$\mathbb{R}$	Conjunto dos números reais
QP	Programação Quadrática
$\nabla f(x)$	Gradiente da função f no ponto x
$\nabla^2 f(x)$	Hessiana da função f no ponto x
$Z^T \nabla f(x_k) Z$	Hessiana reduzida no ponto x_k
A	Matriz das restrições
b	Vetor b em $\mathbb{R}^m$
$\aleph(A)$	Espaço nulo de A
$\mathbb{R}^n$	Espaço dimensional n
$cond(A)$	Número de condição de A
$f(x)$	Função objetivo
$\Re(A)$	Espaço coluna de A
$\Re(A^T)$	Espaço linha de A
V	Espaço vetorial real
LU	Fatoração LU
Q	Matriz ortogonal
QR	Decomposição QR
Z	Matriz cujas colunas geram $\aleph(A)$
B	Matriz básica
N	Matriz não básica
B^{-1}	Matriz inversa B
d	Vetor direção em $\mathbb{R}^n$

dz	Vetor direção em $\mathbb{R}^{n-m}$
x^*	solução ótima
x_B	Variáveis básicas
x_N	Variáveis não básicas
α_k	Tamanho do passo na iteração k
$Z^T G Z$	Matriz hessiana reduzida
$Z^T g$	Vetor gradiente reduzido
$\mathrm{Im}\,(A)$	Conjunto imagem de A
I_{n-m}	Matriz Identidade de ordem $n-m$

Resumo

Este trabalho trata do problema de minimizar uma função quadrática sujeita a restrições lineares de igualdade que aparece em geral como um subproblema nos problemas de programação não linear com restrições. Uma grande variedade de algoritmos de otimização com restrições lineares e não lineares usam resolver a cada iteração um subproblema da forma [3], [24].

$$\begin{array}{ll} \underset{x}{Minimizar} & \varphi(x) = \frac{1}{2}x^T Gx - h^T x \\ sujeito\ a & Ax = b. \end{array}$$

Onde o vetor h_k representa em geral o gradiente da função objetivo ou o gradiente do Lagrangeano, a matriz simétrica G_k é a hessiana da função na k-ésima iteração e a solução x_k representa uma direção de busca. Assumiremos nesse trabalho que A é $m \times n$, com $n > m$ e que A tem posto linha completo. Também assumiremos por conveniência que G é definida positiva no espaço nulo das restrições, garantindo com isso que o problema acima tenha solução única.

Ordinariamente o problema de programação quadrática é resolvido calculando-se uma base Z para o espaço nulo de A, e usando-se essa base para eliminar as restrições, e então utilizando-se um método de minimização irrestrita para resolver o problema reduzido [4].

Neste trabalho apresentaremos um algoritmo baseado no método do gradiente conjugado, que não utiliza eliminar as restrições. Existem pelo menos duas boas razões para isso. As dificuldades inerentes ao precondicionamento quando a base escolhida não é ortogonal, e o problema de perda de esparsidade da matriz hessiana reduzida $Z^T GZ$, que é usualmente densa, mesmo quando a G é esparsa.

Palavras-chave: Conjugacidade, Hessiana Reduzida, Sistema de Newton.

Abstract

This work deals with problem to minimize a quadratic function subject to linear equality constraints which appears in general like a subproblem on the problems of nonlinear quadratic programming with constrained. A large variety of algorithms for linearly and nonlinearly constrained optimization use the conjugate gradient method to solve the each iteration subproblems of the form [3], [24].

$$\begin{array}{ll} \underset{x}{Minimizar} & \varphi(x) = \frac{1}{2}x^T Gx - h^T x \\ sujeito\ a & Ax = b. \end{array}$$

Where o vector h_k usually represents the gradient of the objective function or the gradient of the Lagrangian, the symmetric matrix G_k represents the Hessian of the Lagrangian and a solution x_k represents a search direction. We will assume here that A is an $m \times n$ matrix, with $n > m$, and that A has full row rank. We also assume for convenience that G is positive definite in the null space of the constraints, as this guarantees that the previous problem has a unique solution.

The problem of quadratic program can be solved by computing a basis Z for the null space of A, using this basis to eliminate the constraints, and then applying the conjugate gradient method to resolve the reduced problem [4].

In this work we present an algorithm based in conjugate gradient method without to eliminate the constraints. There are at least two good reasons for this. The difficulties inherent to the preconditioning when the base chosen is not orthogonal, and the problem loss of sparsity of the reduced Hessian matrix $Z^T GZ$, that is usually dense, even when G is sparse.

Key words: Conjugacy, Reduced Hessian, Newton System.

1 Introdução

1.1 Desafio

Uma grande variedade de algoritmos de otimização com restrições lineares e não lineares utilizam o método de gradiente conjugado para resolver os subproblemas da forma

$$\begin{array}{ll} \underset{x}{Minimizar} & \varphi(x) = \frac{1}{2}x^T G x - h^T x \\ sujeito\ a & Ax = b. \end{array}$$

Na otimização não linear, o vetor h representa o gradiente da função objetivo ou o gradiente do Lagrangeano, a matriz simétrica $G_{n\times n}$ é a hessiana da função ou do Lagrangeano e a solução x representa uma direção de busca. Assumiremos aqui que A é $m \times n$, com $n > m$, que A tem posto linha completo, e que as restrições $Ax = b$ constituem m equações linearmente independentes. Também assumimos por conveniência que G é definida positiva no espaço nulo das restrições, e com isso garantimos que o problema acima tenha solução única.

O problema de programação quadrática pode ser resolvido calculando-se uma base Z para o espaço nulo de A, e usando-se essa base para eliminar as restrições, e então aplicando-se o método do gradiente conjugado para o problema reduzido.

Neste trabalho estudamos como aplicar o método do gradiente conjugado, sem eliminar as restrições. Existem duas razões para isto. Vários algoritmos de otimização requerem a solução de duas formas distintas de sistemas lineares de equações em todas as iterações; uma para calcular os multiplicadores de lagrange e a factibilidade, e outra para calcular a base do espaço nulo Z, que é usado para encontrar a solução do problema acima. O uso da base Z para obtenção do sistema reduzido pode nos conduzir a dificuldades com relação ao

precondicionamento da matriz hessiana reduzida Z^TGZ que é usualmente densa, mesmo quando G é esparsa.

Assim o objetivo deste trabalho é apresentar o desenvolvimento de métodos baseados em idéias de conjugacidade que contornem principalmente a questão da esparsidade, além de permitir ao algoritmo parar em alguma solução satisfatória antes de esgotar a dimensão do espaço de busca.

O problema de programação quadrática desempenha um papel fundamental na questão geral de otimização porque desempenha um duplo papel. Primeiro porque muitos problemas de otimização são formulados diretamente como um problema de programação quadrática , incluindo-se aí o problema de quadrados mínimos linear e não linear sujeito a restrições lineares, que é provavelmente um dos mais freqüentes problemas de otimização na estatística, onde aparece com o nome de problema de regressão. A segunda razão importante é porque o problema de programação quadrática aparece como um subproblema a ser resolvido a cada iteração no problema geral de programação não linear sujeita a restrições não lineares.

Nas aplicações de engenharia de produção esse problema aparece em diversas áreas, ou diretamente [1],[19] ou combinado com outras técnicas de otimização [21],[10]. Por exemplo, no problema de programação de produção, através da relaxação do lagrangeano [15]; no problema de layout, combinado com programação inteira [22], etc.

1.2 Motivação

Existem várias motivações para o estudo de programação quadrática, tanto do ponto de vista das aplicações em engenharia de produção quanto do ponto de vista de seu aparecimento em um subproblema. Nesse último caso, trata-se de fato de um problema que deve ser resolvido como um subproblema a cada iteração do problema geral de programação não linear. O desafio aqui é obter um algoritmo que produza uma solução de boa qualidade com baixo custo computacional.

1.3 Proposta e Contribuição

Neste trabalho é apresentado um algoritmo novo para resolver o problema de programação quadrática que não envolve nem o cálculo da hessiana reduzida, nem o cálculo do sistema Karush-Kuhn-Tucker. O método trabalha diretamente no espaço nulo da matriz das restrições, gerando um conjunto G-conjugado, onde G é a hessiana da função

a ser minimizada. A cada iteração é feito a checagem do critério de otimalidade, e nos termos do teorema (36) o algoritmo converge em no máximo $n - m$ iterações, onde n e m são as dimensões da matriz das restrições.

1.4 Estrutura do Trabalho

Este trabalho está organizado em 6 capítulos, sendo que o primeiro consiste na introdução. O segundo capítulo apresenta uma revisão dos conceitos básicos de álgebra linear. O terceiro capítulo apresenta revisão de programação não linear e de otimização. No quarto capítulo são descritos os algoritmos propostos. O quinto capítulo apresenta os resultados dos experimentos numéricos, e no sexto capítulo são apresentadas as conclusões.

2 Tópicos de Álgebra Linear

2.1 Espaços vetoriais

Definição 1 *Um espaço vetorial é um conjunto não vazio V munidos das operações de adição e multiplicação por escalar, tais que:*

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

   - $u + v = v + u$, $\forall\ u, v \in V$;
   - $u + (v + w) = (u + v) + w$, $\forall\ u, v, w \in V$;
   - *Existe em V, um único vetor nulo, tal que $u + \mathbf{0} = u$ e $\mathbf{0} + u = u, \forall\ u \in V$;*
   - *Para cada $u \in V$, existe um único vetor $-u \in V$, tal que $u + (-u) = \mathbf{0}$ e $(-u) + u = \mathbf{0}$.*

2. *Se $u \in V$ e α é um escalar, então, $\alpha u \in V$, isto é, a operação de multiplicação é fechada em V, e satisfaz:*

   - $\alpha(\beta u) = (\alpha\beta)u$, $\forall\ \alpha, \beta \in \mathbb{R}$, $\forall\ u \in V$;
   - $(\alpha + \beta)u = \alpha u + \beta u$, $\forall \alpha, \beta \in \mathbb{R}$, $\forall\ u \in V$;
   - $\alpha(u + v) = \alpha u + \alpha v$, $\forall \alpha \in \mathbb{R}$, $\forall\ u, v \in V$;
   - $1u = u$, $\forall\ u \in V$.

2.2 Subespaços vetoriais

Definição 2 *Suponha que V seja espaço vetorial real, e que V_0 é um subconjunto de V (isto é, cada elemento de V_0 é também um elemento de V). Suponha também que as operações dos elementos de V_0 são as memas operações que em V. Então V_0 é dito ser um subespaço vetorial de V.*

Teorema 1 *Suponha que V é um espaço vetorial e que V_0 é um subconjunto de V. Então, V_0 é um subespaço de V se, e somente se, as seguintes condições se cumprem:*

- *V_0 é não vazio.*
- *V_0 é fechado para a operação de multiplicação por escalar, no sentido que $\alpha v_0 \in V_0$, $\forall v_o \in V_0$, $\forall \alpha \in \mathbb{R}$.*
- *V_0 é fechado para a operação de adição, no sentido que $v_0 + v_0' \in V_0$, $\forall\ v_0, v_0' \in V_0$.*

2.2.1 Os Subesbaços Fundamentais de uma Matriz

Seja $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$e $b \in \mathbb{R}^m$. Então, o sistema linear $Ax = b$ pode ser resolvido se, e somente se, o vetor b puder ser expresso como uma combinação das colunas de A. Esse conjunto que consiste em todas as combinações das colunas de A, denotado por $\Re(A)$ é um subespaço de $\mathbb{R}^m$. Quando $b = 0$ sempre existe a solução particular $x = 0$, mas podem existir infinitas outras soluções. O conjunto de soluções para $Ax = 0$ é ele mesmo um espaço vetorial - o espaço nulo de A. O espaço nulo de uma matriz consiste em todos os vetores x tais que $Ax = 0$, é denotado por $\aleph(A)$ e é um subespaço de $\mathbb{R}^n$. O espaço das linhas de A é gerado pelas linhas de A, que é o espaço das colunas de A^T, denotado por $\Re(A^T)$. O espaço nulo de A^T contém todos os vetores y tais que $A^T y = 0$, e é escrito como $\aleph(A^T)$.

Resumindo, os quatro subespaços fundamentais de uma matriz A são:

1. O espaço das colunas de A é denotado por $\Re(A)$;
2. O espaço nulo de A é denotado por $\aleph(A)$;
3. O espaço das linhas de A é denotado por $\Re(A^T)$;
4. O espaço nulo de A^T é denotado por $\aleph(A^T)$.

2.3 Base

Antes de definirmos o que é uma base faz-se necessária a compreensão de conjuntos geradores e independência linear.

2.3.1 Conjunto Geradores

Para possibitar uma melhor compreensão sobre conjunto gerador faz-se necessário antes, citar o conceito de combinação linear.

Combinação Linear

Uma combinação linear dos vetores v_1,$v_2, ..., v_n$ é uma expressão da forma

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n,$$

onde os α_i's são escalares.

Vetores Linearmente Dependentes e Vetores Linearmente Independentes

Um vetor v é dito linearmente dependente dos vetores $v_1, v_2, ..., v_n$ se e somente se v pode ser escrito como alguma combinação linear de $v_1, v_2, ..., v_n$; caso contrário, v é dito ser linearmente independente dos vetores $v_1, v_2, ..., v_n$.

Com base nesses conceitos de combinação linear e vetores linearmente independentes, pode-se definir o que é um conjunto gerador.

Definição 3 *Seja S um conjunto de vetores $v_1, v_2, ..., v_n$ em V. S é dito gerador de algum subespaço V_0 de V se e somente se S é um subconjunto de V_0 e todo vetor v_0 em V_0 é linearmente independente dos vetores do conjunto S.*

2.3.2 Independência Linear

Definição 4 *Seja $L = \{v_1, v_2, ..., v_k\}$ um conjunto não vazio de vetores.*

Suponha que

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = 0$$

implica que $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$. *L é então dito ser linearmente independente. Assim, um conjunto que não é linearmente independente é dito ser linearmente dependente; equivalentemente, L é linearmente dependente se e somente se existem escalares* $\alpha_1, \alpha_2, \cdots, \alpha_k$, *não todos nulos, com*

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = 0.$$

O seguinte teorema formaliza que de fato as duas diferentes maneiras de definir o que é um conjunto linearmente independente são equivalentes. Dele também extraímos outros resultados úteis.

Teorema 2

1. *Suponha que*

   $$L = \{v_1, v_2, ..., v_k\}$$

   com $k \geq 2$ *e todo vetor* $v_i \neq \mathbf{0}$. *Então, L é linearmente independente se e somente se ao menos um dos* v_j *é linearmente dependente dos vetores restantes* v_i $(i \neq j)$; *em particular, L é linearmente dependente se e somente se ao menos um dos vetores* v_j *é linearmente dependente dos seus vetores precedentes* $v_1, v_2, ..., v_{j-1}$.

2. *Qualquer conjunto contendo o vetor* $\mathbf{0}$ *é linearmente dependente.*

3. $\{v\}$ *é linearmente dependente se e somente se* $v \neq \mathbf{0}$.

4. *Suponha que v seja linearmente dependente de um conjunto*

   $$L = \{v_1, v_2, ..., v_k\},$$

   e que v_j *seja linearmente dependente sobre os outros vetores de L, isto é,*

   $$L_j^{'} = \{v_1, v_2, ..., v_{j-1}, v_{j+1}, ..., v_k\}.$$

   Então, v é linearmente dependente sobre $L_j^{'}$.

5. *Todo subconjunto de um conjunto linearmente independente é linearmente independente.*

6. *Suponha que L é um conjunto finito devetores e que algum subconjunto L_0 de L é linearmente dependente. Então L é linearmente dependente.*

A definição de base para um espaço vetorial V é dada a seguir.

Definição 5 *Uma base para um espaço vetorial V é um conjunto gerador linearmente independente.*

O seguinte teorema, refere-se a unicidade da representação do vetor quando esse pertencer a uma base.

Teorema 3 *Seja $B = \{v_1, ..., v_r\}$ uma base. Então, a representação de todo $v \in B$ é única: se $v = \alpha_1 v_1 + \cdots + \alpha_r v_r$ e também $v = \alpha_1^{'} v_1 + \cdots + \alpha_r^{'} v_r$, então $\alpha_i = \alpha_i^{'}$ para $1 \leq i \leq r$.*

2.3.3 Dimensão

Definição 6 *O número de vetores em uma base para um espaço vetorial é conhecida como dimensão do espaço. Se a dimensão de V é p, V é dito ser de dimensão p. Quando um espaço vetorial tem uma base consistindo de algum número finito de vetores, o espaço é dito ser de dimensão finita. A dimensão de um espaço $\{0\}$ é dito ser 0.*

Corolário 4 *O espaço vetorial real $\mathbb{R}^n$ é de dimensão n.*

Teorema 5 *Seja V um espaço vetorial de dimensão p. Então:*

- *Todo conjunto contendo estritamente mais que p vetores é linearmente dependente.*
- *Se $D = \{v_1, v_2, ..., v_r\}$ é linearmente independente e r é menor que p, então existe vetores $v_{r+1}, ..., v_p$ de modo que $\{v_1, v_2, ..., v_p\}$ é uma base para V.*
- *Se um conjunto de exatamente p vetores ou é um conjunto gerador para V ou é linearmente independente, então são ambos e são uma base para V.*

Teorema 6 *Um espaço vetorial tem dimensão finita k se, e somente se, k é o número máximo de vetores em um conjunto linearmente independente.*

2.4 Produto interno

Definição 7 *Seja V um espaço vetorial. Um produto interno em V é a função que associa a cada par ordenado de vetores u, v em V um número real, denotado por $\langle u, v\rangle$, satisfazendo:*

- $\langle u, v\rangle = \langle v, u\rangle, \forall u, v \in V$;
- $\langle \alpha u + \beta v, w\rangle = \alpha \langle u, w\rangle + \beta \langle v, w\rangle$ *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* $\forall \alpha, \beta \in \mathbb{R}$;
- $\langle u, u\rangle > 0$, *se* $u \neq 0$ *e* $\langle u, u\rangle = 0$, *se e somente se,* $u = 0$.
- *O ângulo entre dois vetores não nulos u e v, é definido por:*

$$\cos\theta = \frac{\langle u, v\rangle}{\langle u, u\rangle^{1/2} \langle v, v\rangle^{1/2}}.$$

2.5 Norma de vetores

Definição 8 *Uma norma vetorial de v é um número real não negativo, denotado por $\|v\|$, satisfazendo as seguintes propriedades:*

- $\|v\| > 0$ *para* $v \neq 0$, *e* $\|v\| = 0$ *exatamente quando* $v = 0$.
- $\|\alpha v\| = |\alpha|.\|v\|, \forall\ v \in \mathbb{R}^n, \alpha \in \mathbb{R}$.
- $\|u + v\| \leq \|u\| + \|v\|$, *para todo vetor u e v em V (Desigualdade Triangular).*

A norma-p de um vetor-n v é denotado por $||v||_p$, e é definido como

$$||v||_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{1/p}.$$

Os três valores mais comuns para p são $p = 1, 2$ e ∞, que correspondem as seguintes normas:

1. *norma-1:* $||v||_1 = |v_1| + ... + |v_n|$,
2. *norma-2:* $||v||_2 = (|v_1|^2 + ... + |v_n|^2)^{\frac{1}{2}} = (v^T v)^{\frac{1}{2}}$,

3. *norma-∞:* $||v||_\infty = \lim_{p\to\infty}(||v_1||^p + ... + ||v_n||^p)^{\frac{1}{p}} = \max_i |v_i|$.

Existem várias desigualdades úteis que relaciona o produto interno de dois vetores com suas respectivas normas. Considerado os vetores x, y e $y - x$ como um triângulo definido em $\mathbb{R}^n$, a fórmula do cosseno pode ser utilizada para relacionar o ângulo entre x e y e o comprimento dos vetores x, y e $y - x$:

$$||y - x||_2^2 = ||y||_2^2 + ||x||_2^2 - 2||y||_2||x||_2 \cos\theta.$$

Expandindo a expressão $||y - x||_2^2$ como $(y - x)^T(y - x)$, obtém-se

$$\cos\theta = \frac{y^T x}{||x||_2||y||_2}.$$

Visto que *cos* θ situa-se entre -1 e $+1$, segue que

$$|y^T x| \leq ||x||_2||y||_2,$$

que é conhecida como *Desigualdade de Schwartz*.

Teorema 7 *(Normas associadas do produto interno). Seja $\langle u, v\rangle$ um produto interno em V, e defina $\|v\| = \langle v, v\rangle^{1/2}$ .Então $\|\cdot\|$ é uma norma em V (dita ser a norma induzida pelo produto interno).*

2.6 Norma de matrizes

Definição 9 *A norma de uma matriz A, denotada por $||A||$, é um escalar não negativo que satisfaz as seguintes propriedades:*

- $||A|| \geq 0$ *para todo A;* $||A|| = 0$ *se e somente se A é a matrix nula;*
- $||\alpha A|| = |\alpha|\ ||A||$*;*
- $||A + B|| \leq ||A|| + ||B||$*;*
- $||AB|| \leq ||A||\ ||B||$.

Uma norma matricial pode ser convenientemente definida nos termos de uma norma vetorial. Dado uma norma vetorial $||\cdot||$ e uma matriz A, considere $||Ax||$ para todos vetores, tal que $||x|| = 1$. Uma norma matricial *induzida por*, ou *subordinada a*, uma norma vetorial, é dada por:

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

As normas de matrizes correspondentes a uma matriz $A_{m\times n}$, são:

- $||A||_1 = \max\limits_{1\leq\, j\, \leq n} \left(\sum\limits_{i=1}^{m} |a_{ij}|\right)$, máximo das somas dos valores absolutos de colunas ;
- $||A||_2 = \left(\alpha_{\max}(A^T A)\right)^{1/2}$, a raiz quadrada do maior autovalor de $A^T A$;
- $||A||_\infty = \max\limits_{1\leq\, i\, \leq m} \left(\sum\limits_{j=1}^{n} |a_{ij}|\right)$, máximo das somas dos valores absolutos de linhas.

A matriz norma-2 é também conhecida como *norma spectral*.

2.6.1 Norma de Frobenius

Uma norma matricial importante que não é subordinada a uma norma vetorial é a *norma de Frobenius*, denotada por $\|\cdot\|_F$. Esta norma surge considerando um matriz $A_{m\times n}$ com um vetor com mn elementos, e então calculando a *norma Euclidiana* daquele vetor:

$$||A||_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{\frac{1}{2}}.$$

Uma norma vetorial $||\cdot||$ e uma norma matricial $||\cdot||'$ são ditas ser *compatíveis* se, para todo A e x,

$$||Ax|| \leq ||A||'||x||. \tag{2.1}$$

Por definição, (2.1) sempre verifica-se para uma norma vetorial e a norma matricial subordinada, com possível igualdade para algum vetor (ou vetores) x. A norma vetorial Euclidiana e a norma matricial de Frobenius são também compatíveis.

2.7 Ortogonalidade

Definição 10 *Seja V um espaço vetorial munido de um produto interno, $\langle\cdot,\cdot\rangle$, e seja $\|\cdot\|$ sua norma induzida delo produto interno, ver "Norma induzida pelo produto interno"(pág. 10).*

i. Dois vetores u e v são ditos ortogonais se, e somente se, $\langle u, v\rangle = 0$.

ii. Um conjunto de vetores é dito ser ortogonal se, e somente se, todo par de vetores do conjunto são ortogonais: $\langle u, v\rangle = 0$, para todo $u \neq v$ naquele conjunto.

iii. Se um vetor não nulo u é usado para produzir $v = u/\|u\|$, tal que $\|v\| = 1$, então u foi normalizado para produzir o vetor normalizado v.

iv. Um conjunto de vetores é dito ser ortonormal se, e somente se, o conjunto é ortogonal e $\|v\| = 1$ para todo v no conjunto.

Pelas propriedades de produto interno, $\langle 0, v\rangle = \langle 0v, v\rangle = 0\langle v, v\rangle = 0$. Isto é, em qualquer espaço vetorial com um produto interno, o vetor $\mathbf{0}$ é ortogonal a todos os vetores.

2.7.1 Projeções Ortogonais

Teorema 8 *Sejam V um espaço vetorial munido de um produto interno e V_0 o subespaço de V gerado por um conjunto ortogonal*

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

de vetores não nulos. Define-se projeção ortogonal P_0 sobre V_0 como segue: para qualquer v em V, estabeleça:

$$P_0 v = \alpha_1 v_1 + \cdots \alpha_q v_q, \quad onde \quad \alpha_i = \frac{\langle v_i, v\rangle}{\langle v_i, v_i\rangle}.$$

Então:

- *$v - P_0 v$ é ortogonal para todos os vetores v_0 em V_0;*
- *$P_0(u + v) = P_0 u + P_0 v, \forall u, v \in V$;*
- *$P_0(\alpha v) = \alpha P_0 v$, $\forall \alpha \in \mathbb{R}$ e $\forall v \in V$.*

2.7.2 Base Ortogonal e Ortonormal

Teorema 9 *Seja $B = \{v_1, v_2, ..., v_q\}$ uma base ortogonal (ou ortonormal). Então a representação de algum vetor v com respeito a base ortogonal B pode imediatamente ser escrita abaixo:*

$$v = \alpha_1 v_1 + ... + \alpha_q v_q, \quad onde \quad \alpha_i = \frac{\langle v_i, v\rangle}{\langle v_i, v_i\rangle}.$$

Deparando com um caso especial, expressando algum vetor p como uma combinação liner de $\mathbf{e}_1, ..., \mathbf{e}_p$: estes vetores formam uma base ortonormal para $\mathbb{R}^p$, assim os coeficientes na representação de v são

$$\alpha_i = \frac{\langle \mathbf{e}_i, v \rangle}{\langle \mathbf{e}_i, \mathbf{e}_i \rangle} = \langle v \rangle_i .$$

2.7.3 Subespaços Ortogonais

Definição 11 *Dois subespaços V e W do mesmo espaço $\mathbb{R}^n$ são ortogonais se todo $v \in V$ é ortogonal a todo vetor $w \in W$:*

$$v^T w = 0, \quad \forall v, w.$$

Processo de ortogonalização de Gram-Schmidt

O processo de Gram-Schmidt produz uma base ortogonal a um conjunto gerador e detecta se o conjunto original é linearmente independente.

Teorema 10 *Seja $S = \{v_1, .., v_q\}$ o gerador do espaço vetorial V munido de um produto interno. O processo de Gram-Schmidt é como segue. Seja V_i o subespaço de V gerado por $\{v_1, ..., v_i\}$ e seja P_i a projeção ortogonal sob V_i; se $V_i = \{0\}$, seja $P_i v = 0$ para todo v.*

1. *Define-se $u_1 = v_1$.*
2. *Para $2 \leq i \leq q$, define-se $u_i = v_i - P_{i-1} v_i$.*

A seguintes propriedades, mantém-se para os vetores produzidos por esse processo:

- *$B = \{u_1, ..., u_q\}$ é um conjunto ortogonal gerador V.*
- *Para $1 \leq i \leq q$, $B_i = \{u_1, ..., u_i\}$ é um conjunto ortogonal gerador V_i, o subespaço gerado por $\{v_1, ..., v_i\}$.*
- *$u_i = 0$ se e somente se v_i é linearmente dependente nos vetores $v_1, ..., v_{i-1}$.*
- *Uma base ortogonal para V pode ser obtida de um conjunto gerador ortogonal B omitindo $u_i = 0$, se esse existir.*
- *Se S é uma base para V, então B é uma base ortogonal para V.*

O processo tradicional de Gram-Schmidt

Dado um conjunto gerador $v_1, ..., v_q$:

1. Define-se $u_1 = v_1$.
2. Para $2 \leq i \leq q$, define-se

$$u_i = v_i - \alpha_{1i}u_1 - ... - \alpha_{i-1,i}u_{i-1},$$

onde $\alpha_{ji} = \frac{\langle u_j, v_i \rangle}{\langle u_j, u_j \rangle}$ se $u_j \neq 0$ e $\alpha_{ji} = 0$ se $u_j = 0$.

2.8 Matrizes Definidas

2.8.1 Matrizes Simétricas

Definição 12 *$A \in \mathbb{R}^{n \times n}$ é simétrica se $A = A^T$.*

2.8.2 Matriz Definida Positiva

Definição 13 *Seja uma matriz simétrica $A \in \mathbb{R}^{n \times n}$. Diz que A é definida positiva se $x^T Ax > 0$ para todo vetor não-nulo $x \in \mathbb{R}^n$.*

Teorema 11 *Cada uma das cinco propriedades seguintes é uma condição necessária e suficiente para que uma matriz simétrica $A \in \mathbb{R}^{n \times n}$ seja definida positiva:*

1. *$x^T Ax > 0$ para todo vetor x não nulo;*
2. *Todos os autovalores de A satisfazem $\lambda_i > 0$;*
3. *Todas as submatrizes triangulares superiores esquerdas A_k possuem determinantes positivos;*
4. *A possui um conjunto de n pivôs $d_i > 0$;*
5. *Existe uma matriz R, com colunas linearmente independentes, tal que $A = R^T R$.*

Demonstração. Suponha que x_i é um autovetor unitário associado ao autovalor λ_i. Então,

$$Ax_i = \lambda_i x_i \tag{2.2}$$

de modo que

$$x_i^T A x_i = x_i^T \lambda_i x_i = \lambda_i, \tag{2.3}$$

pois $x_i^T x_i = 1$. Pela definição $x^T A x > 0$, então, em particular, $x_i^T A x_i > 0$ e, assim, tem-se $\lambda_i > 0$.

Agora suponha todos os $\lambda_i > 0$. Isso deve ser provado para todo vetor x, não somente para os autovalores de A. Como matrizes simétricas simétricas possuem um conjunto completo de autovalores ortonormais, é possível escrever qualquer vetor x como uma combinação $c_1 x_1 + \cdots + c_n x_n$. Então,

$$Ax = c_1 A x_1 + \cdots + c_n A x_n = c_1 \lambda_1 x_1 + \cdots + c_n \lambda_n x_n. \tag{2.4}$$

Devido à ortogonalidade e à normalização

$$\begin{aligned} x_i^T x_i &= 1, \\ x^T A x &= (c_1 x_1^T + \cdots + c_n x_n^T)(c_1 \lambda_1 x_1 + \cdots + c_n \lambda_n x_n) \\ &= c_1^2 \lambda_1 + \cdots + c_n^2 \lambda_n. \end{aligned} \tag{2.5}$$

Se cada $\lambda_i > 0$, então a expressão (2.5) mostra que $x^T A x > 0$. Portanto, a condição (2.3) implica na condição (2.2). As condições (2.4) e (2.5) e suas equivalências com a condição (2.2) serão provadas em três passos. Se (2.2) é verdadeiro então (2.4) também é: Primeiro, o determinante de qualquer matriz é o produto dos seus autovalores. Sabe-se que esses autovalores são positivos

$$det(A) = \lambda_1 . \lambda_2 ... \lambda_n > 0. \tag{2.6}$$

Para provar o mesmo resultado para todas as submatrizes A_k, verifica-se que, se A é definida positiva, então todas as $A_k^{'}s$ também são. Verificam-se os vetores cujas últimas $n - k$ componentes são nulas:

$$x^T A x = \begin{bmatrix} x_k^T & 0 \end{bmatrix} \begin{bmatrix} A_k & * \\ * & * \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = x_k^T A_k x_k. \tag{2.7}$$

Se $x^T A x > 0$, $\forall x \neq 0$, então, em particular, $x_k^T A_k x_k$, $\forall x_k \neq 0$. Portanto, a condição (2.2) se verifica para A_k, e a submatriz permite os mesmos argumentos utilizados para A. Seus autovalores (que não são os mesmos λ_i) devem ser positivos, e seu determinante é o produto deles, de modo que os determinantes superiores esquerdos são positivos. Se (2.4)

é verdadeira, então (2.5) também é : Existe uma relação direta entre os números $det(A)$ e os pivôs. O k-ésimo pivô d_k é exatamente a razão entre $det(A_k)$ e $det(A_{k-1})$. Portanto, se os determinantes são todos positivos, então os pivôs também são e nenhuma permutação de linha é necessária para matrizes definidas positivas. Para verificar a condição (2.6), deve-se reconhecer $x^T R^T Rx$ como a raiz quadrada de $\|Rx\|^2$. Isto não pode ser negativo e também não pode ser nulo (a não ser que $x = 0$), porque R possui colunas linearmente independentes: se x é não nulo então Rx é não nulo. Portanto, $x^T Ax = \|Rx\|^2$ é positivo e $A = R^T R$ é definida positiva. ■

2.8.3 Matriz Semidefinida Positiva

Definição 14 *Seja $A \in \mathbb{R}^{n\times n}$ uma matriz simétrica. A é dita ser semidefinida positiva se $x^T Ax \geq 0$ para todo vetor $x \in \mathbb{R}^n$.*

Teorema 12 *Cada uma das cinco propriedades seguintes é uma condição necessária e suficiente para que A seja semidefinida positiva:*

1. *$x^T Ax \geq 0$ para todo vetor x;*
2. *Todos os autovalores de A satisfazem $\lambda_i \geq 0$;*
3. *Todas as submatrizes principais possuem determinantes não negativos;*
4. *A possui um conjunto de n pivôs $d_i \geq 0$;*
5. *Existe uma matriz R, possivelmente possuindo colunas linearmente dependentes, tal que $A = R^T R$.*

2.8.4 Matriz Definida Negativa

Definição 15 *Seja $A \in \mathbb{R}^{n\times n}$ uma matriz simétrica. A é dita ser definida negativa se $x^T Ax < 0$ para todo vetor não nulo $x \in \mathbb{R}^n$.*

Teorema 13 *Cada uma das quatro propriedades seguintes é uma condição necessária e suficiente para que A seja definida negativa:*

1. *$x^T Ax < 0$ para todo vetor x;*
2. *Todos os autovalores de A satisfazem $\lambda_i < 0$;*

3. *Todas as submatrizes triangulares superiores esquerdas A_k possuem determinantes negativos;*

4. *A possui um conjunto de n pivôs $d_i < 0$.*

2.8.5 Matriz Semidefinida Negativa

Definição 16 *Seja $A \in \mathbb{R}^{n\times n}$ uma matriz simétrica. A é dita ser semidefinida negativa se $x^T Ax \leq 0$ para todo vetor $x \in \mathbb{R}^n$.*

Teorema 14 *Cada uma das quatro propriedades seguintes é uma condição necessária e suficiente para que A seja semidefinida negativa:*

1. *$x^T Ax \leq 0$ para todo vetor x;*

2. *Todos os autovalores de A satisfazem $\lambda_i \leq 0$;*

3. *Todas as submatrizes principais possuem determinantes negativos;*

4. *A possui um conjunto de n pivôs $d_i \leq 0$.*

2.8.6 Matrizes Indefinidas

As matrizes que não se enquadram nas definições anteriores são consideradas indefinidas.

2.9 Matriz de Projeção

Seja A uma matriz $m \times n$ e $b \in \mathbb{R}^m$ um vetor fora do espaço das colunas de A. A construção de uma linha perpendicular do ponto b até o espaço das colunas de A pode ser expressa em termos matriciais por $p = A(A^T A)^{-1} A^T b$. A matriz nesta fórmula é uma matriz de projeção e será denotada por

$$P : P = A(A^T A)^{-1} A^T$$

Esta matriz projeta qualquer vetor b no espaço das colunas de A. Em outras palavras, $p = Pb$ é a componente de b no espaço das colunas de A, e a diferença $b - Pb$ é a componente no complemento ortogonal. Tem-se então uma fórmula matricial para decompor um vetor

em duas componentes perpendiculares. Pb entá no espaço das colunas $\Re(A)$, e a outra componente $(I-P)b$ está no espaço nulo esquerdo $\aleph(A^T)$, que é ortogonal ao espaço das colunas de A.

A matriz projeção P possui duas propriedades básicas:

1. $P^2 = P$;
2. $P^T = P$.

Reciprocamente, qualquer matriz simétrica com $P^2 = P$ representa uma projeção.

2.10 Matriz de Permutação

Definição 17 *Uma matriz de permutação $P_{n\times n}$ é qualquer matriz $n \times n$ que resulta da permutação das linhas da matriz identidade I_n.*

Pré-multiplicar uma matriz A por uma matriz de permutação P produz o mesmo resultado que permutar as linhas de A, exatamente da mesma maneira que as linhas de I_n foram permutadas para produzir P.

Teorema 15 *Seja P uma matriz de permutação. Então:*

1. *Para cada A, PA pode ser obtida a partir da matriz A permutando suas linhas exatamente como as linhas de I foram permutadas para obter P;*
2. *P é não singular, e $P^{-1} = P^T$. Isto é, $PP^T = P^TP = I$.*

Demonstração.

1. Segue da definição de multiplicação de matrizes e das matrizes de permutação.
2. Particione P em suas linhas $r_1, ..., r_n$, as quais são simplesmente as linhas e_i^T de I em alguma ordem. Então P^T possui as r_i^T como suas colunas. A definição de multiplicação de matrizes implica que a entrada (i, j) de PP^T é simplesmente $r_i r_j^T$ (a entrada na matriz 1×1), e isto é 1 se $i = j$ e 0 se $i \neq j$; ou seja, $PP^T = I$. Um argumento similar em termos das colunas de P mostra que também $P^TP = I$.

■

2.11 Propriedades das Matrizes A^TA e AA^T

A matriz A^TA é simétrica, sua transposta é $(A^TA)^T = A^TA^{TT}$, que é A^TA novamente. Cada entrada (i, j) é um produto interno da coluna i de A com a coluna j de A. Isso concorda com a entrada (j, i), que é a coluna j vezes a coluna i. Se A tem posto completo, então A^TA tem o mesmo espaço nulo de A, $\aleph(A^TA) \subset \aleph(A)$. Certamente se $Ax = 0$, então $A^TAx = 0$. Os vetores x no espaço nulo de A estão também no espaço nulo de A^TA. Suponha que $A^TAx = 0$ e tome o produto interno com x

$$x^TA^TAx = 0$$

ou

$$||Ax||^2 = 0$$

ou

$$Ax = 0.$$

Portanto, x está no espaço nulo de A, $\aleph(A) \subset \aleph(A^TA)$ e os dois subespaços são idênticos. Em particular, se A possui colunas independentes (e somente $x = 0$ está em seu espaço nulo) então o mesmo é verdade para A^TA: se A possui colunas linearmente independentes, então A^TA é quadrada, simétrica, inversível e definida positiva.

2.12 Fatoração de matrizes

Para alguns problemas de álgebra linear computacional, é útil representar uma matriz como o produto ou soma de matrizes de forma especial. Cada representação é chamada de fatorização ou decomposição da matriz original.

2.12.1 Fatoração LU

Para se ver como uma fatoração pode ser usada para resolver $Ax = b$, assumindo que a matriz A não singular tenha sido fatorada na forma $A = BC$, onde sistemas envolvendo matrizes não singulares B e C são de fácil resolução. A propriedade associativa de multiplicação de matriz significa que a $Ax = b$ pode ser escrita como

$$Ax = BCx = B(Cx) = b.$$

O sistema composto $BCx = b$ é resolvido primeiro tratando o vetor Cx como uma incógnita, por exemplo y, e resolvendo o sistema $By = b$, para y. Uma vez conhecido y, a solução x do sistema original é obtida resolvendo $Cx = y$. De fato, usando uma fatoração de matrizes para resolver $Ax = b$ substitui o sistema original por uma seqüência de sistemas simples.

A fatoração associada com a eliminação Gaussiana é a fatoração LU, e tem a forma

$$A = LU,$$

onde L é triangular inferior unitária e U é triangular superior.

2.12.2 Fatoração de Cholesky

Uma fatoração famosa e extremamente útil é da forma $A = LDL^T$, que é definida quando A é simétrica e definida positiva. L é uma matriz triangular inferior.

Matrizes simétricas e definidas positivas têm algumas propriedades interessantes, algumas que seguem imediatamente da definição. Se A é simétrica e definida positiva, então:

(i) $a_{ii} > 0$ para todo i, isto é, todos os elementos da diagonal são positivos;

(ii) o maior elemento (em magnitude) na matriz inteira pode ocorrer na diagonal;

(iii) todos os autovalores de A são reais e estritamente positivos;

(iv) toda matriz selecionada simetricamente das linhas e colunas de A são definidas positivas. Em particular, as submatrizes $1 \times 1, 2 \times 2, ..., n \times n$ começando no canto superior esquerdo são definidas positivas;

(v) se a eliminação Gaussiana sem trocas é executado em A, a matriz restante é definida positiva para todo passo.

As propriedades (i) e (v) implicam que a fatoração $A = LDL^T$ sempre existe quando A é simétrica e definida positiva, e que todos os elementos da diagonal de D são estritamente positivos. Com isso em mente, pode-se escrever:

$$A = LDL^T = LD^{\frac{1}{2}}D^{\frac{1}{2}}L^T = \overline{LL}^T = R^T R,$$

onde $\overline{L} = LD^{\frac{1}{2}}$, $R = \overline{L}^T$ e R é uma matriz triangular superior não singular. A fatoração $A = R^T R$ é conhecida como *fatoração de Cholesky* e a matriz R é chamada de fator de Cholesky.

A fatoração de Cholesky pode ser calculada utilizando eliminação, com a importante diferença que as matrizes eliminadas podem ser aplicadas de ambos os lados esquerdo e direito para reter simetria nos fatores. Este processo pode ser entendido facilmente, representando A na forma particionada,

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

com o bloco $(1,1)$ obtido como o escalar a_{11} (que, de acordo com a propriedade (i), é estritamente positivo):

$$A = \begin{pmatrix} a_{11} & A_{21}^T \\ A_{21} & A_{22} \end{pmatrix},$$

onde A_{21}^T é o vetor linha $(a_{21}, ..., a_{n1})$ e A_{22} é $(n-1) \times (n-1)$. O primeiro passo da eliminação Gaussiana produz a seguinte expressão para a matriz reduzida particionada:

$$\begin{aligned} M_1 A &= \begin{pmatrix} 1 & 0 \\ -(1/a_{11})A_{21} & I_{n-1} \end{pmatrix} \begin{pmatrix} a_{11} & A_{21}^T \\ A_{21} & A_{22} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & 0 \\ 0 & A_{22} - (1/a_{11})A_{21}A_{21}^T \end{pmatrix} \\ &= A^{(2)}. \end{aligned}$$

A primeira coluna de $A^{(2)}$ é agora um múltiplo de e_1. Segue por simetria que, a primeira linha pode ser reduzida aplicando M_1^T à direita:

$$M_1 A M_1^T = \begin{pmatrix} a_{11} & 0 \\ 0 & A_{22} - (1/a_{11})A_{21}A_{21}^T \end{pmatrix}.$$

A propriedade (v) das matrizes definidas positivas implica que a matriz restante é definida positiva, e que o próximo passo da eliminação pode ser executado sem trocas. Continuando deste modo, a propriedade (v) assegura que todo passo subseqüente da eliminação de esquerda e direita é completada com sucesso. O resultado final é

$$MAM^T = D,$$

onde $M = M_{n-1} \cdots M_1$ e D é a matriz diagonal estritamente positiva. Fazendo $L = M^{-1}$, segue que $A = LDL^T = R^T R$, a forma desejada.

Os fatores de Cholesky podem ser calculados diretamente, pelas linhas e colunas. Para a representação $R^T R$, segue que:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} r_{11} & & & \\ r_{12} & r_{22} & & \\ \vdots & & \ddots & \\ r_{1n} & r_{2n} & \cdots & r_{nn} \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix}$$

Equacionando o elemento (1,1) de A satisfaz $a_{11} = r_{11}^2$. Segue da propriedade (i) de matrizes definidas positivas que $a_{11} > 0$, o elemento (1,1) de R é definido por

$$r_{11} = \sqrt{a_{11}}.$$

(A raiz quadrada positiva é escolhida por convenção.) Movendo para o outro lado da primeira linha de A, vê-se que $r_{11}r_{12} = a_{12}$, $r_{11}r_{13} = a_{13}$, ..., $r_{11}r_{1n} = a_{1n}$, que resulta

$$r_{1j} = \frac{a_{1j}}{r_{11}}, \qquad j = 2, ..., n,$$

e completa a primeira linha de R.

Continuando para o elemento-(2,2),

$$r_{12}^2 + r_{22}^2 = a_{22}, \quad \text{tal que} \quad r_{22}^2 = a_{22} - r_{12}^2.$$

A quantidade $a_{22} - r_{12}^2$ é o elemento diagonal levado da matriz restante depois do passo da eliminação Gaussiana e é positiva devido a propriedade (v). O elemento r_{22} é portanto bem definido. O elemento $(2, j)$ de A igual o produto interno da segunda e j-ésima coluna de R, tal que

$$r_{12}r_{1j} + r_{22}r_{2j} = a_{2j}, \qquad j = 3, ..., n.$$

Visto que os valores de r_{12}, r_{1j} e r_{22} são conhecidos, o j-ésimo elemento na segunda linha de R é dado por

$$r_{2j} = \frac{a_{2j} - r_{12}r_{1j}}{r_{22}}, \qquad j = 3, ..., n.$$

Continua-se deste modo até que todos os elementos sejam calculados. O k-ésimo elemento da diagonal de R é definido pela equação

$$r_{kk}^2 = d_k = a_{kk} - r_{1k}^2 - r_{2k}^2 - ... - r_{k-1,k}^2.$$

Porque d_k é o elemento diagonal levado da matriz restante para o passo k, a propriedade (v) de matrizes definidas positivas garante que d_k seja estritamente positivo para $k = 1, ..., n$. A k-ésima linha de R é determinada da relação

$$r_{1k}r_{1j} + r_{2k}r_{2j} + ... + r_{kk}r_{kj} = a_{kj}, \quad j = k+1, ..., n.$$

Uma característica importante da fatoração de Cholesky pode ser vista particionando A e o fator de Cholesky R como

$$A = \begin{pmatrix} A_{11} & A_{21}^T \\ A_{21} & A_{22} \end{pmatrix} = R^T R = \begin{pmatrix} R_{11}^T & \\ R_{12}^T & R_{22}^T \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ & R_{22} \end{pmatrix},$$

onde A_{11} e A_{22} são quadradas, as dimensões de R_{11} são as mesmas de A_{11}, e pela propriedade (iv) A_{11} é definida positiva. Equacionando os elementos, segue

$$\begin{pmatrix} A_{11} & A_{21}^T \\ A_{21} & A_{22} \end{pmatrix} = R^T R = \begin{pmatrix} R_{11}^T R_{11} & R_{11}^T R_{12} \\ R_{12}^T R_{11} & R_{12}^T R_{12} + R_{22}^T R_{22} \end{pmatrix},$$

que mostra que R_{22} é o fator de Cholesky.

Dado o fator de Cholesky R, a solução x de $Ax = R^T Rx = b$ pode ser calculada resolvendo dois sistemas triangulares (transposto):

$$R^T y = b \quad \text{e} \quad Rx = y.$$

O fator de Cholesky R requer $\frac{1}{6}n^3$ multiplicações (e adições), e n raízes quadradas.

Uma importante benefício da definição positiva na fatoração de Cholesky é que nenhuma troca é necessária para a estabilidade numérica. Esta propriedade verifica-se devido a seguinte relação entre os elementos de A e R:

$$r_{1k}^2 + r_{2k}^2 + \ ... + \ r_{kk}^2 = a_{kk}, \quad k = 1, ..., n.$$

Todos os termos da soma são não negativos, o que implica que esta expressão supri por um limite a *priori* na magnitude dos elementos de R:

$$|r_{ik}| \leq \sqrt{a_{kk}}.$$

Esta relação mostra que todo elemento da k-ésima linha de R é limitado na magnitude por uma raiz quadrada do k-ésimo elemento diagonal da matriz original, sendo que o crescimento não pode ocorrer em alguma matriz reduzida particionada.

2.12.3 Fatoração QR

Suponha que $v_1, v_2, ..., v_q$ são matrizes colunas $p \times 1$ e considere a implementação do processo de Gram-Schmidt. Da definição matrizes coluna u_j, segue

$$v_j = u_1\alpha_{1j} + u_2\alpha_{2j} + ... + u_{j-1}\alpha_{j-1,j} + u_j.$$

Na notação de matriz particionada, isso é justamente

$$\begin{bmatrix} v_1 & \cdots & v_q \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_q \end{bmatrix} \begin{bmatrix} 1 & \alpha_{12} & \cdots & \alpha_{1q} \\ 0 & 1 & \cdots & \alpha_{2q} \\ & & \cdots & \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Seja Q_0 denotar a matriz $p \times q$ $\begin{bmatrix} u_1 & \cdots & u_q \end{bmatrix}$ e seja R_0 denotar a matriz triangular superior unitária $q \times q$; visto que $A = \begin{bmatrix} v_1 & \cdots & v_q \end{bmatrix}$ pode ser reescrita como

$$A = Q_0R_0. \tag{2.8}$$

Os u_i em Q_0 são construídos pelo processo de Gram-Schmidt, e portanto são mutuamente ortogonais; isto é, Q_0 tem colunas ortogonais, algumas que podem ser iguais a 0. Seja Q e R matrizes obtidas pela eliminação das colunas zero de Q_0 e as linhas correspondentes de R_0, e dividindo cada coluna não nula de Q_0 pela norma-2 e multiplicando cada linha correspondente de R_0 pela mesma norma-2. Então (2.8) torna-se

$$A = QR \tag{2.9}$$

com R triangular superior e Q tendo colunas ortonormais. A forma (2.8) é chamada de *decomposição QR não normalizada* de A, enquanto (2.9) é a *decomposição QR normalizada*.

Teorema 16 *Seja A uma matriz $p \times q$ de posto k. Então:*

1. *A pode ser escrita pela decomposição QR não normalizada como $A = Q_0R_0$, onde:*

   - *Q_0 é $p \times q$ e tem colunas ortogonais (de que k são não nulos e $q - k$ são zeros) que geram o espaço coluna de A.*
   - *R_0 é $q \times q$, triangular superior unitária, e não singular.*

- *A norma-2 da i-ésima coluna de Q_0 é igual a distância da i-ésima coluna de A para o espaço gerador pelas primeiras $i-1$ colunas de A.*

2. *A pode ser escrita decomposição QR normalizada como $A = QR$, onde:*

- *Q é $p \times k$ e tem colunas ortonormais que geram o espaço coluna de A.*
- *R é $k \times q$, triangular superior, e tem posto k.*
- *Se $k = q$, então $|\langle R \rangle_{ii}|$ é igual a distância da i-ésima coluna de A para o espaço gerador pelas primeiras $i-1$ colunas de A.*

2.13 Autovalores e Autovetores

Para qualquer matriz quadrada A, existe pelo menos um número λ e um vetor não nulo associado u tal que:

$$Au = \lambda u, \tag{2.10}$$

ou equivalentemente,

$$(A - \lambda I)\, u = 0. \tag{2.11}$$

A equação em (2.11) indica que a subtração de λ de cada elemento diagonal de A produz uma matriz singular; a primeira relação mostra que a premultiplicação de u por A não altera a direção de u. O valor λ é chamado de *autovalor* de A, e o vetor correspondente u é chamado de *autovetor de A.*

Qualquer matriz $A_{n\times n}$ possui n autovalores $\{\lambda_1, ..., \lambda_n\}$ não necessariamente distintos, os quais são as n-ésimas raízes 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,

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

O produto dos autovalores é igual ao determinante de A:

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

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 de SAS^{-1}.

Se a matriz A é singular, existe um vetor não nulo x tal que $Ax = 0$, o que mostra que a 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 $\mathbb{R}^n$.

2.13.1 Convergência

Definição 18 *Seja* $||\cdot||$ *uma norma em* V*. Uma sequência de vetores* v_i *é dito ser convergente para o vetor* v_∞ *se, e somente se, a sequência de números reais* $||v_i - v_\infty||$ *converge para* 0*.*

2.14 Número de condição

O sistema linear

$$Ax = b, \tag{2.12}$$

tem uma única solução somente se A é quadrada e não singular. Antes de considerar como resolver (2.12), é interessante ver como a solução é afetada por pequenas perturbações no lado direito e nos elementos da matriz.

A solução exata de (2.12) é dado por

$$x = A^{-1}b.$$

Suponha que o lado direito de (2.12) é perturbado por $b + \delta b$, e que a solução exata do sistema perturbado é $x + \delta x$, isto é,

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

onde "δ" denota uma pequena mudança em um vetor ou matriz. Portanto,

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

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

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

Para medir δx, invocando a propriedade de vetor compatível e norma de matriz:

$$||x|| \ \leq ||A^{-1}|| \ ||\delta b||, \tag{2.13}$$

com possível igualdade para algum vetor δb. A perturbação na solução exata é ainda limitada por cima por $||A^{-1}||$ vezes a perturbação na lado direito.

Para determinar o *efeito relativo* desta mesma perturabação, observe que

$$||b|| \ \leq ||A|| \ ||x||. \tag{2.14}$$

Combinando as desigualdades (2.13) e (2.14) e reorganizando-as, segue

$$\frac{||\delta x||}{||x||} \leq ||A|| \ ||A^{-1}|| \ \frac{||\delta b||}{||b||} \ . \tag{2.15}$$

Se (2.12) é perturbada por uma matriz A, um processo similar produz $(A + \delta A)(x + \delta x) = b$. Esta equação pode ser reescrita como

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

de modo que

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

ou

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

Quando a mudança $||\delta A||$ é considerada relativa, isto torna

$$\frac{||\delta x||}{||x + \delta x||} \leq ||A|| \ ||A^{-1}|| \ ||\delta A||. \tag{2.16}$$

Em ambos (2.15) e (2.16), a mudança relativa na solução exata é limitado pelo fator $||A||\ ||A^{-1}||$ multiplicado pela perturbação relativa nos dados (matriz ou lado direito). O número $||A||\ ||A^{-1}||$ é definido como *número de condição de* A e é denotado por $cond(A)$.

Visto que $||I|| = 1$ para alguma norma subordinada, e $I = AA^{-1}$, segue que

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

de modo que

$$cond(a) \geq 1$$

para toda matriz.

O número de condição de uma matriz A indica o *efeito máximo* de perturbações em b ou A na solução exata de $Ax = b$. Se a matriz A é *mal condicionada,* $cond(A)$ é "grande". Se a matriz A é *bem condicionada,* $cond(A)$ é "pequeno".

3 Tópicos de Programação Não Linear e Otimização

3.1 Funções de Classe C^k

Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função contínua de valor real. f é dita ser da classe C^k em $\mathbb{R}^n$ se f tem derivadas contínuas de todas ordens até k .

3.2 Gradiente

Considere uma função $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, suave. Se a primeira derivada parcial de f com respeito às n variáveis existem e são contínuas, o vetor coluna dessas n derivadas parciais é chamado de gradiente de f em x, e representado por $\nabla f(x)$.

Definição 19 *Uma função contínua $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ é dita ser continuamente diferenciável se para todo $x \in \mathbb{R}^n$, $(\partial f/\partial xi)(x)$, $i = 1, ..., n$, existe e é contínua; o gradiente de f em x é definido como*

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, ..., \frac{\partial f(x)}{\partial x_n}\right]^T.$$

A função f é dita ser continuamente diferenciável em uma região aberta $D \subset \mathbb{R}^n$, denotando $f \in C^1(D)$, se é continuamente diferenciável para todo ponto em D.

3.3 Hessiana

Considere uma função contínua $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. Se a primeira e a segunda derivadas de f com respeito às n variáveis existem e são contínuas, a matriz contendo essas derivadas

parciais é chamada de *Hessiana* de f em x, $\nabla^2 f(x)$, assim,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & & \vdots \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

é chamada de hessiana da função f no ponto x. $H = \nabla^2 f(x)$.

3.4 Teorema de Taylor

O teorema de Taylor é de fundamental importância no que se segue porque ele mostra que se a função e suas derivadas são conhecidas em um ponto, então pode-se calcular aproximações para a função em todos os pontos pertencentes à vizinhança deste ponto.

Teorema 17 *(Teorema de Taylor) Se $f \in C^k$ então existe um escalar θ $(0 \leq \theta \leq 1)$, tal que*

$$f(x+h) = f(x) + hf\,'(x) + \tfrac{1}{2}h^2 f''(x) + ...$$

$$+ \tfrac{1}{(k-1)!}h^{k-1}f^{(k-1)}(x) + \tfrac{1}{k!}h^k f^{(k)}(x+\theta h),$$

onde $f^{(k)}(x)$ denota a k-ésima derivada de f calculado em x.

Para a prática computacional, é interessante somente os três primeiros termos da expansão:

$$F(x+hp) = F(x) + hg(x)^T p + \tfrac{1}{2}h^2 p^T G(x) p + O(h^3).$$

Note que o raio de mudança de F no ponto x ao longo da direção p é dada pela quantidade $g(x)^T p$, que é chamado de *derivada direcional* (ou a primeira derivada ao longo de p). Similarmente, o escalar $p^T G(x)$ pode ser interpretado como a segunda derivada de F ao longo de p, e é conhecido como a *curvatura de F ao longo de p*. Uma direção p tal que $p^T G(x) p > 0$ (< 0) é conhecido como *direção de curvatura positiva* (*curvatura negativa*).

3.5 Teorema Fundamental do Cálculo

Teorema 18 *Seja $f \in C^2$ e D um conjunto convexo aberto. Se $x, y \in D$, então*

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle dt.$$

Demonstração. Considere uma função unidimensional $\phi(t) = f(x+t(y-x))$. O teorema fundamental do Cáculo afirma que

$$\phi(1) - \phi(0) = \int_0^1 \phi'(t)dt.$$

Visto que $\phi(1) = f(y), \phi(0) = f(x)$ e

$$\phi'(t) = \langle \nabla f(x + t(y - x)), y - x \rangle,$$

tem-se que $\phi(1) - \phi(0) = \int_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle dt.$ ■

3.6 Derivadas Direcionais

Definição 20 *Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ ser continuamente diferenciável em um conjunto convexo $D \subset \mathbb{R}^n$. Então, para $x \in D$ e alguma perturbação não nula $p \in \mathbb{R}^n$, a derivada direcional de f em x na direção de p, definido por*

$$\frac{\partial f}{\partial p}(x) = \lim_{\varepsilon \longrightarrow 0} \frac{f(x + \varepsilon p) - f(x)}{\varepsilon}.$$

Lema 19 *Se $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ existe, então $\forall x \in \mathbb{R}^n$,*

$$\lim_{\varepsilon \longrightarrow 0} \frac{f(x + \varepsilon p) - f(x)}{\varepsilon} \nabla f(x)^T p,$$

$$f(x + p) = f(x) + \int_0^1 \nabla f(x + tp)^T p dt \equiv f(x) + \int_x^{x+p} \nabla f(z) dz. \tag{3.1}$$

Além disso existe $z \in (x, x + p)$ tal que

$$f(x + p) = f(x) + \nabla f(z)^T p. \tag{3.2}$$

Demonstração. Parametrizando f ao longo da linha formada pelo intervalo de $(x, x+p)$ tem-se a função g de uma variável definida por

$$g : \mathbb{R} \to \mathbb{R}, \quad g(t) = f(x + tp)$$

e recorrendo para o cálculo de uma variável. Definindo $x(t) = x + tp$. Então pela regra da cadeia, para $0 \leq \alpha \leq 1$,

$$\begin{aligned} \frac{dg}{dt}(\alpha) &= \sum_{i=1}^{n} \frac{\partial f(x(t))}{\partial x(t)_i}\left(x(\alpha)\right) \frac{dx(t)_i}{dt}(\alpha) \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x(\alpha)).p_i \\ &= \nabla f(x + \alpha p)^T p. \end{aligned} \tag{3.3}$$

Substituindo $\alpha = 0$ reduzindo (3.3) para

$$\frac{\partial f(x)}{\partial p} = \nabla f(x)^T p.$$

Pelo teorema fundamental do cálculo ou teorema de Newton,

$$g(1) = g(0) + \int_0^1 g\prime(t)dt$$

que, pela definição de g e (3.3), é equivalente

$$f(x + p) = f(x) + \int_0^1 f(x + tp)^T p dt$$

e prova (3.1). Finalmente, pelo teorema para funções de uma variável,

$$g(1) = g(0) + g\prime(\xi), \qquad \xi \in (0, 1)$$

que pela definição de g e (3.3), é equivalente para

$$f(x + p) = f(x) + \nabla f(x + \xi p)^T p, \qquad \xi \in (0, 1),$$

e prova (3.2). ■

3.7 Direção de Descida

Definição 21 *As direções $d \in \mathbb{R}^n$, tais que*

$$\nabla f(x)^T d < 0$$

são chamadas direções de descida a partir de x.

Lema 20 *Sejam $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $f \in C^1$ e $x \in \mathbb{R}^n$ tal que $\nabla f(x) \neq 0$, $d \in \mathbb{R}^n$, $\nabla f(x)^T d < 0$. Então, existe $\bar{\alpha} > 0$ tal que*

$$f(x + \alpha d) < f(x), \ \forall \alpha \in (0, \bar{\alpha}).$$

Demonstração. Considere $\phi : \mathbb{R} \to \mathbb{R}$,

$$\phi(\alpha) = f(x + \alpha d)$$

e

$$\phi(0) = f(x).$$

Pela regra da cadeia,

$$\phi'(\alpha) = \nabla f(x + \alpha d)^T d$$

$$\phi'(0) = f(x)^T d$$

e

$$\phi'(0) = \lim_{\alpha \to 0} \frac{\phi(0 + \alpha) - \phi(0)}{\alpha}.$$

Para $0 < \alpha < \bar{\alpha}$, com $\bar{\alpha}$ suficientemente pequeno, o sinal de $\phi'(0)$ e $\phi(\alpha) - \phi(0)$ deve ser o mesmo, ou seja, negativo. Como $\phi'(0) = f(x)^T d < 0$, logo

$$\phi(\alpha) - \phi(0) < 0,$$

$$f(x + \alpha d) - f(x) < 0$$

$$f(x + \alpha d) < f(x).$$

■

3.8 Condições de Otimalidade

Os problemas de otimização envolvem minimizar ou maximizar uma função objetivo, sujeita a um conjunto de restrições impostas nas variáveis. Em, essencialmente, todos os problemas de interesse as restrições serão expressas em termos de relações envolvendo as variáveis das funções contínuas. O problema geral a ser considerado é conhecido como *otimização linear restrita* e é expresso em termos matemáticos por:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{minimizar} & f(x) \\ \text{s.a.} & \begin{array}{l} c_i(x) = 0,\ i = 1, 2, ..., m'; \\ c_i(x) \geq 0,\ i = m' + 1, ..., m. \end{array} \end{array} \tag{3.4}$$

Um ponto $\widehat{x}$ é dito ser viável se satisfaz todas as restrições do problema (3.4). O conjunto de todos os pontos viáveis é denominado região viável. Um problema para o qual não existe pontos viáveis é chamado de problema inviável.

Para selecionar um método eficiente para resolver um problema particular da forma (3.4), é necessário analisar e classificar o problema. Antes de considerar um método para resolver um problema é necessário definir uma "solução para o (3.4). Para tal solução existir considere os seguintes teoremas e as definições de conjuntos e funções convexas.

3.8.1 Teorema de Weierstrass

Teorema 21 *Seja S um conjunto não vazio, compacto, e seja $f : S \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ contínua em S. Então, o problema*

$$\min_{x\in S} \quad f(x) \tag{3.5}$$

atinge seu mínimo, isto é, existe um minimizador para o problema (3.5).

Demonstração. Visto que f é contínua em S e S é fechado e limitado, f é limitado inferiormente em S. Conseqüentemente, visto que $S \neq \emptyset$, existe o menor limite inferior $\alpha \equiv \inf\{f(x)\} : x \in S\}$.

Agora, seja $0 < \varepsilon < 1$, e considere o conjunto $S_k = \{x \in S : \alpha + \varepsilon^k\}$ para todo $k = 1, 2, ...$ Pela definição de ínfimo, $S_k \neq \emptyset$ para todo k e então constrói-se uma seqüência de pontos $\{x_k\} \subseteq S$ selecionando um ponto $x_k \in S_k$ para todo $k = 1, 2, ...$ Como S é limitado, existe uma subseqüência convergente $\{x_k\}_K \longrightarrow \overline{x}$, indexado pelo

conjunto K. Como S é fechado, tem-se $\overline{x} \in S$; pele continuidade da função f,desde que $\alpha \leq f(x_k) \leq \alpha + \varepsilon^k$ para todo k,tem-se que

$$\alpha = \lim_{k \longrightarrow +\infty,\ k \in K} f(x_k) = f(\overline{x}).$$

Portanto, foi mostrado que existe uma solução $\overline{x} \in S$ tal que $f(\overline{x}) = \alpha = \inf\{f(x) : x \in S\}$, e então $\overline{x}$ é um minimizador. ■

3.8.2 Conjuntos Convexos

Definição 22 *Um conjunto S em $\mathbb{R}^n$ é convexo se para todo x e y em S o segmento de reta que une x e y também está em S.*

Definindo o segmento de linha $[x, y]$ que une x e y por:

$$[x, y] = \{x + \lambda(y - x) : 0 \leq \lambda \leq 1\}.$$

Note que este conjunto pode também ser descrito como segue:

$$[x, y] = \{\lambda y + (1 - \lambda)x : 0 \leq \lambda \leq 1\}.$$

Logo, um subconjunto S de $\mathbb{R}^n$ é convexo se, e somente se, para todo x e y em S e todo λ com $0 \leq \lambda \leq 1$, o vetor $\lambda x + (1 - \lambda)y$ está também em S.

Teorema 22 *Seja S um conjunto convexo em $\mathbb{R}^n$. Deixe $x^{(1)}, x^{(2)}, ..., x^{(k)}$ estar em S. Se $\lambda_1, \lambda_2, ..., \lambda_k$ são números não negativos cuja soma é 1, então a combinação convexa $\sum_{i=1}^{k} \lambda_i x^{(i)}$ está também em S.*

3.8.3 Funções Convexas

Um conjunto convexo se caracteriza por conter todos os segmentos cujos extremos são pontos do conjunto. Se x e y são pontos de $\mathbb{R}^n$, o segmento que os une está formado pelo z da forma $y + \lambda(x - y) \equiv \lambda x + (1 - \lambda)y$ com $\lambda \in [0, 1]$. Isto justifica a seguinte definição.

Definição 23 *Suponha que $f(x)$ é uma função real definida em um conjunto convexo $S \subset \mathbb{R}^n$. Então:*

- *a função $f(x)$ é convexa em S se*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y),$$

para todo $x, y \in S$ *e todo* λ *com* $0 \leq \lambda \leq 1$;

- *a função* $f(x)$ *é estritamente convexa em* S *se*

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y),$$

para todo $x, y \in S$ *com* $x \neq y$ *e todo* λ *com* $0 < \lambda < 1$.

Teorema 23 *Todo minimizador local de uma função convexa* $f(x)$ *definida em um subconjunto* S *de* $\mathbb{R}^n$ *é também um minimizador global. Todo minimizador local de uma função estritamente convexa* $f(x)$ *definida em um conjunto convexo* $S \subset \mathbb{R}^n$ *é o único minimizador global estrito de* $f(x)$ *em* S.

Demonstração. Suponha que x^* é um minimizador local para a função convexa $f(x)$ em S. Então existe um número positivo r tal que $f(x) \geq f(x^*)$ sempre que $x \in S$ e $\|x - x^*\| < r$.

Dado algum $y \in S$, quer provar-se que $f(y) \geq f(x^*)$. Para este fim, seleciona-se λ com $0 < \lambda < 1$ e tão pequeno que $x^* + \lambda(y - x^*) \in S$ e

$$\|x^* + \lambda(y - x^*) - x^*\| < r.$$

Por conseguinte,

$$f(x^*) \leq f(x^* + \lambda(y - x^*)) = f(\lambda y + [1-\lambda]x^*) \leq \lambda f(y) + [1-\lambda]f(x^*),$$

pois $f(x)$ é convexo em S. Nota-se que a última desigualdade é estrita se $y \neq x^*$ e $f(x)$ é estritamente convexa. Agora, subtrai-se $f(x^*)$ de ambos os lados da desigualdade anterior e divide-se o resultado por λ

$$\frac{f(x^*) - f(x*)}{\lambda} \leq \frac{\lambda f(y) + [1-\lambda]f(x^*) - f(x^*)}{\lambda},$$

para obter

$$0 \leq f(y) - f(x^*)$$

desigualdade estrita se $f(x)$ é estritamente convexa e $y \neq x^*$. Isto estabelece o resultado desejado. ■

Teorema 24 .

(a) Se $f_1(x), ..., f_k(x)$ são funções convexas no conjunto convexo S em $\mathbb{R}^n$, então

$$f(x) = f_1(x) + f_2(x) + \cdots + f_k(x)$$

é convexo. Além disso, se pelo menos um $f_i(x)$ é estritamente convexo em S, então a soma $f(x)$ é estritamente convexa.

(b) Se $f(x)$ é convexa (resp. estritamente convexa) em um conjunto S em $\mathbb{R}^n$ e se α é um número positivo, então $\alpha f(x)$ é convexo (resp. estritamente convexo) em S.

(c) Se $f(x)$ é uma função convexa (resp. estritamente convexa) definida em um conjunto S em $\mathbb{R}^n$, e se $g(y)$ é um função convexa aumentada (resp. estritamente aumentada) definida no range de $f(x)$ em $\mathbb{R}$, então a função composta $g(f(x))$ é convexa (resp. estritamente convexa) em S.

Demonstração.

(a) Para mostrar que qualquer soma finita de funções convexas em S é convexa em S, é suficiente mostrar que a soma $(f_1 + f_2)(x)$ de duas funções convexas $f_1(x)$ e $f_2(x)$ em S é de novo convexa em S. Se y, z pertence a S e $0 \leq \lambda \leq 1$, então

$$\begin{aligned}(f_1 + f_2)(\lambda y + [1-\lambda]z) &= f_1(\lambda y + [1-\lambda]z) + f_2(\lambda y + [1-\lambda]z) \\ &\leq \lambda f_1(y) + [1-\lambda]f_1(z) + \lambda f_2(y) + [1-\lambda]f_2(z) \\ &= \lambda(f_1 + f_2)(y) + [1-\lambda](f_1 + f_2)(z).\end{aligned}$$

Portanto, $(f_1 + f_2)(x)$ é convexo em S. Além disso, está claro deste cálculo que se $f_1(x)$ ou $f_2(x)$ é estritamente convexa, então $(f_1 + f_2)(x)$ é estritamente convexa porque a convexidade estrita de um função apresenta uma desigualdade estrita no lugar certo.

(b) Este resultado segue de um argumento similar usado em (a).

(c) Se y, z pertence a S e se $0 \leq \lambda \leq 1$, então

$$f(\lambda y + [1-\lambda]z) \leq \lambda f(y) + [1-\lambda]f(z)$$

visto que $f(x)$ é convexo em S. Conseqüentemente, visto que g é um aumento, a função convexa no range de $f(x)$, segue que

$$\begin{aligned} g(f(\lambda y + [1-\lambda]z)) &\leq g(\lambda f(y) + [1-\lambda]f(z)) \\ &\leq \lambda g(f(y)) + [1-\lambda]g(f(z)). \end{aligned}$$

Ainda, a função composta $g(f(x))$ é convexa em S. Se $f(x)$ é estritamente convexa e g é estritamente aumentada, a primeira desigualdade no cálculo anterior é estrito para $y \neq z$ e $0 < \lambda < 1$, assim $g(f(x))$ é estritamente convexa em S.

■

3.8.4 Caracterização de um ponto de mínimo

Um ponto $\bar{x} \in \mathbb{R}^n$ é solução de 3.4) se $\bar{x}$ é um ponto que pertença a região viável. Então, somente pontos viáveis podem ser ótimos.

Seja x^* um ponto viável para o problema (3.4), e defina $N(x^*, \delta)$ como sendo o conjunto dos pontos viáveis em uma vizinhança δ de x^*.

Definição 24 *O ponto x^* é um mínimo local se existe $\delta > 0$ tal que:*

1. *$f(x)$ é definido em $N(x^*, \delta)$; e*
2. *$f(x^*) < f(y)$, $\forall y \in N(x^*, \delta)$, $y \neq x^*$.*

3.8.5 Otimização com restrição linear

Em muitos problemas práticos, nem todos os valores possíveis das variáveis são aceitáveis, e são com freqüência necessárias ou desejáveis para as restrições impostas. Uma forma freqüente de restrição envolve especificamente que uma certa função linear de variáveis pode ser exatamente zero, não negativa, ou não positiva. A forma geral da função não linear é $l(x) = \alpha^T x - \beta$, para algum vetor linha α^T e escalar β. Por linearidade, o vetor coluna α é o gradiente (constante) de $l(x)$. Os tipos de restrições lineares são:

(i) $\alpha^T x - \beta = 0$ (restrição de igualdade);

(ii) $\alpha^T x - \beta \geq 0$ (restrição de desigualdade).

3.8.6 Condições de Otimalidade de uma Função com Restrições Lineares de Igualdade

Considerando as condições de otimalidade para um problema que contém somente *restrições de igualdade linear* (LEP), isto é,

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimizar}} \quad & F(x) \\ \text{sujeito a} \quad & Ax = b. \end{aligned}$$

A i-ésima linha da matriz $A_{m\times n}$ será denotada por a_i^T, e contém os coeficientes da i-ésima restrição linear:

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

Um ponto viável x^* é um mínimo local da LEP se e somente se $F(x^*) \leq F(x)$ para todo x na vizinhança de x^*. Não existe ponto viável se as restrições são inconsistentes. Assume-se que b situa-se no range de A.

Considere o passo entre dois pontos factíveis $\overline{x}$ e $\widehat{x}$; por linearidade $A(\overline{x}-\widehat{x}) = 0$, visto que $A\overline{x} = b$ e $A\widehat{x} = b$. Por razões similares o passo p de um ponto viável para outro ponto viável pode ser ortogonal as linhas de A, isto é, pode satisfazer

$$Ap = 0. \tag{3.6}$$

Algum vetor p tal que (3.6) mantém-se é chamado de direção viável com respeito as restrições de igualdade do LEP. Algum passo de um ponto viável ao longo da direção não viola as restrições, visto que $A(\widehat{x} + \alpha p) = A\widehat{x} = b$. Se um passo infinitesimal ao longo do vetor p provoca $Ap \neq 0$, causa um ponto perturbado e assim a infactibilidade.

Deixe as colunas da matriz Z formar uma base, então $AZ = 0$, e toda direção viável pode ser escrita como uma combinaçãqo linear das colunas de Z. Portanto, se p satisfaz (3.6), p pode ser escrito como Zp_z para algum vetor p_z.

Examinando a expansão da série de Taylor de F em relação a x^* ao longo de uma direção viável p ($p = Zp_z$):

$$F(x^* + \epsilon Zp_z) = F(x^*) + \epsilon p_z^T Z^T g(x^*) + \frac{1}{2}\epsilon^2 p_z^T Z^T G(x^* + \epsilon\theta p) Zp_z \tag{3.7}$$

onde θ satisfaz $0 \leq \theta \leq 1$, e ϵ (3.7) mostra que $p_z^T Z^T g(x^*)$ é negativo, então toda vizinhança de x^* conterá pontos viáveis com valor da função estritamente inferior. Ainda, a condição necessária para x^* seja um mínimo local da LEP é que $p_z^T Z^T g(x^*)$ pode sumir para todo p_z, o que implica que

$$Z^T g(x^*) = 0. \tag{3.8}$$

O vetor $Z^T g(x^*)$ é chamado de gradiente projetado de F em x^*.

O resultado (3.8) implica que $g(x^*)$ pode ser uma combinação linear das linhas de A, isto é,

$$g(x^*) = \sum_{i=1}^{m} \alpha_i \lambda_i^* = A^T \lambda^*,$$

para algum vetor λ^*, que é chamado vetor multiplicador de Lagrange. Os multiplicadores de Lagrange são únicos somente se as linhas de A são linearmente independentes.

Visto que $Z^T g(x^*) = 0$, a expansão da série de Taylor (3.7) torna-se

$$F(x^* + \epsilon Z p_z) = F(x^*) + \tfrac{1}{2}\epsilon^2 p_z^T Z^T G(x^* + \epsilon\theta p) Z p_z$$

indicando que se a matriz $Z^T G(x^*) Z$ é indefinida, toda vizinhança de x^* contém pontos viáveis com um valor estritamente inferior de F. Portanto a segunda condição necessária para que x^* seja ótimo para LEP é que a matriz $Z^T G(x^*) Z$, que chamada de matriz hessiana projetada pode ser semidefinida positiva.

Condições necessárias para um mínimo de LEP.

- $Ax^* = b$;
- $Z^T g(x^*) = 0$; ou, equivalentemente, $g(x^*) = A^T \lambda^*$; e
- $Z^T G(x^*) Z$ é semi-definida positiva.

Condições suficientes para um mínimo de LEP.

- $Ax^* = b$;
- $Z^T g(x^*) = 0$; ou, equivalentemente, $g(x^*) = A^T \lambda^*$; e
- $Z^T G(x^*) Z$ é definida positiva.

3.8.7 Convergência de Seqüências e Rapidez de Convergência

Definição 25 *Dado um método iterativo que produz uma seqüência de pontos* x_1, x_2, ... *a partir de um ponto inicial* x_0, *quer-se saber se o método converge para uma solução* x^*, *e caso afirmativo, com que rapidez. Sejam* $x^* \in \mathbb{R}^n$, $x_k \in \mathbb{R}^n$, $k = 1, 2, 3, ...$ *Então a seqüência* $\{x_k\}$ *converge para* x^* *se*

$$\lim_{k \longrightarrow +\infty} \|x_k - x^*\| = 0.$$

Se além disso, existir uma constante $c \in [0, 1)$ *e um inteiro* $\hat{k} \geq 0$ *tal que para todo* $k \geq \hat{k}$,

$$\|x_{k+1} - x^*\| \leq c\|x_k - x^*\|$$

então $\{x_k\}$ *é dita ser linearmente convergente. Se para alguma seqüência* $\{c_k\}$ *que converge para* 0,

$$\|x_{k+1} - x^*\| \leq c_k\|x_k - x^*\|,$$

então $\{x_k\}$ *converge superlinearmente para* x^*. *Se existirem constantes* $p > 1$, $c \geq 0$, *e* $\hat{k} \geq 0$ *tais que* $\{x_k\}$ *converge para* x^* *e para todo* $k \geq \hat{k}$,

$$\|x_{k+1} - x^*\| \leq c\|x_k - x^*\|^2,$$

então a convergência é dita ser quadrática.

Um método iterativo que convergirá para a resposta correta em uma certa taxa, contanto que seja inicializado perto o suficiente da resposta correta, é dito ser localmente convergente naquela taxa.

3.9 Funções Quadráticas em $\mathbb{R}^n$

Seja n um inteiro fixo e seja F uma função quadrática

$$F(x) = \frac{1}{2}x^T Ax - h^T x + c \tag{3.9}$$

em $\mathbb{R}^n$, onde A é uma matriz simétrica $n \times n$, h é um vetor de dimensão n, e c é um escalar.

O gradiente de F em x é o vetor

$$\nabla F(x) = Ax - h. \tag{3.10}$$

Um ponto crítico de F é um ponto x tal que $\nabla F(x) = 0$. Ainda x é um ponto crítico da função quadrática F se e somente se x é uma solução do sistema de equações lineares

$$Ax = h. \tag{3.11}$$

O sistema (3.11) pode, ou não pode, ter uma solução. No entanto, se A é não singular existe exatamente uma solução, isto é,

$$x_0 = A^{-1}h \tag{3.12}$$

e x_0 é o único ponto crítico de F. Se A é singular e x_0 é uma solução (3.11) então toda solução de x de (3.11) é expressa da forma $x = x_0 + z$, onde z é um vetor do espaço nulo de A, isto é, um vetor z tal que $Az = 0$. Em outras palavras se x_0 é um ponto crítico de F, então todo ponto crítico de F difere de x_0 por um vetor z pertencente ao espaço nulo de A.

Como F é uma função quadrática, tem-se a identidade

$$F(x + p) = F(x) + p^T(Ax - h) + \frac{1}{2}p^T Ap \tag{3.13}$$

em x e p. Esta identidade é obtida substituindo x por $x + p$ em (3.9). Observe que (3.13) expressa o *Teorema de Taylor* para o ponto x. Reescrevendo (3.13) tem-se,

$$F(x + p) = F(x) + p^T \nabla F(x) + \frac{1}{2}p^T \nabla^2 F(x)p. \tag{3.14}$$

A matriz $\nabla^2 F(x) = A$ é a *Hessiana* de F. Se x_0 é um ponto crítico de F, isto é, $\nabla F(x_0) = 0$, então substituindo x por x_0 e p por $x - x_0$ em (3.14), obtém-se a equação

$$F(x) = F(x_0) + \frac{1}{2}(x - x_0)^T A(x - x_0) \tag{3.15}$$

para F relativa a um ponto crítico de F. Geometricamente esta equação nos diz que, quando A é não singular, o ponto crítico $x_0 = A^{-1}h$ de F é o centro da superfície quadrática

$$F(x) = \gamma, \tag{3.16}$$

onde γ é uma constante.

Por (3.15) um ponto crítico x_0 de F é um ponto mínimo de F se e somente se A é semidefinida positiva, isto é, se e somente se a desigualdade $p^T Ap \geq 0$ mantém-se para todo vetor p. Se $p^T Ap > 0$ sempre $p \neq 0$, isto é, se A é definida positiva, então $x_0 = A^{-1}h$ é o único ponto mínimo de F. Diz-se que F é definida positiva se A é definida positiva.

As superfícies de nível (3.16) para uma função quadrática definida positiva F são elipsóides tendo $x_0 = A^{-1}h$ como o centro comum.

No estudo de um sistema linear $Ax = b$, define-se o *residual* do sistema de x como o vetor

$$r = h - Ax.$$

Quando A é simétrica, o vetor r é o gradiente negativo

$$r = -\nabla F(x) = h - Ax$$

da função quadrática

$$F(x) = \tfrac{1}{2}x^T Ax - b^T x + c.$$

Segue que o residual $r(x)$ é a *máxima direção de descida* de F para o ponto x.

A preocupação é não somente minimizar F globalmente, mas também minimizar F sujeita a um conjunto de restrições lineares

$$q_i^T x = \rho_i \quad (i = 1, ..., \ n-k, \ 1 \leq k < n), \tag{3.17}$$

onde $q_1, ..., q_{n-k}$ são vetores linearmente independentes. O conjunto de pontos satisfazendo (3.17) é denominado *Hiperplano*, denotado por H_k.

O teorema a seguir apresenta as condições de existência de pontos mínimos de uma função quadrática em hiperplanos.

Teorema 25 *Suponha que a hessiana A de uma função quadrática F é definida positiva. Um ponto x^* em um hiperplano H_k minimiza F em H_k se e somente se o gradiente $\nabla F(x^*)$ é ortogonal a H_k. Existe um único ponto mínimo x^* de F em H_k. Um hiperplano H_k cortado por um elipsóide E_{n-1} de dimensão $(n-1)$*

$$E_{n-1} : F(x) = \gamma$$

corta E_{n-1} em um elipsóide de dimensão $(k-1)$ cujo centro é o ponto mínimo x^* de F em H_k.

3.9.1 Propriedades Básicas de Funções Quadráticas

Teorema 26 *Os pontos mínimos de F sobre retas paralelas estão num hiperplano H_{n-1} que contém o ponto mínimo x_0 de F. O hiperplano H_{n-1} é definido pela equação*

$$p^T(Ax - b) = 0$$

onde p é um vetor direção para essas linhas paralelas. O vetor Ap é normal à H_{n-1}.

Teorema 27 *Dado um vetor não nulo p e seja x_2 e x_2^*, respectivamente, os pontos mínimos de F em duas linhas L e L^* cuja direção é p. O vetor $q = x_2^* - x_2$ é conjugado a p no sentido que a relação $p^T Aq = 0$ se mantém.*

Teorema 28 *O ponto mínimo x_2 de F na linha $x = x_1 + \alpha p$ é dado pela equação*

$$x_2 = x_1 + \alpha p$$

onde

$$\alpha = -\frac{p^T g_1}{p^T Ap}$$

e

$$g_1 = F'(x_1) = Ax_1 - b$$

Teorema 29 *Os ponto mínimos de F em hiperplanos paralelos de dimensão $n-1$ estão numa linha L conjugada a esses planos e passando pelo ponto mínimo x_0 de F. Em outras palavras, se q é um vetor não nulo dado, então para todo número real ρ o ponto mínimo x_1 de F no hiperplano*

$$H_{n-1} : q^T x = \rho$$

está na linha

$$L : x = x_0 + \alpha A^{-1} q$$

passando pelo ponto mínimo x_0 de F na direção $p = A - 1q$. O vetor p, ou equivalentemente a linha L, é conjugada a H_{n-1}.

Teorema 30 *Os pontos mínimos de F em hiperplanos paralelos H_k estão em um hiperplano conjugado a esses hiperplanos e passando pelo ponto mínimo x_0 de F. Em outras palavras, dado um conjunto de $n-k$ vetores linearmente independentes $q_1, ..., q_{n-k}$, então para todo conjunto de números reais $\rho_1, ..., \rho_{n-k}$ o ponto mínimo x_1 de F no hiperplano*

$$H_k : q_i^T x = i \qquad (i = 1, ..., n-k),$$

está no hiperplano

$$H_{n-k} : x = x_0 + \alpha_1 A^{-1} q_1 + \ldots + \alpha_{n-k} A^{-1} q_{n-k}$$

passando pelo ponto mínimo x_0 de F. Os vetores $p_1 = A^{-1}q_1, ..., p_{n-k} = A^{-1}q_{n-k}$ são conjugados a H_k de maneira que H_{n-k} é conjugado a H_k.

Corolário 31 *Se x_1 e x_1^* são respectivamente os pontos mínimos em hiperplanos paralelos H_k e H_k^*, então o vetor $p = x_1^* - x_1$ é conjugado a H_k e H_k^*.*

Teorema 32 *Dado um hiperplano H_k e um vetor w, o conjunto $H_k + w$ de pontos x^* expressível na forma $x^* = x + w$ com x em H_k é um hiperplano H_k^* paralelo a H_k. Reciprocamente, se H_k^* é um hiperplano paralelo a H_k e $w = x^* - x$ é um vetor unindo um ponto x em H_k a um ponto x^* em H_k^*, então $H_k^* = H_k + w$. De fato, existe um único vetor p conjugado a H_k de maneira que $H_k^* = H_k + p$. Se x_1 minimiza F em H_k, então $x_1^* = x_1 + p$ minimiza F em H_k^*.*

Teorema 33 *Dado um hiperplano H_k e um vetor $w \notin H_k$, o conjunto $H_k + \beta w$ de todos os pontos $x+\beta w$, determinado pelos pontos x em H_k e os números reais β, é um hiperplano de dimensão $k+1$ que gera H_k e $H_k^* = H_k + w$. Existe um único vetor p conjugado a H_k tal que $H_k^* = H_k + p$ e $H_k = H_k + \beta p$. O ponto mínimo x_2 de F em H_k está na linha $x = x_1 + \alpha p$, onde x_1 é o ponto mínimo de F em H_k.*

3.9.2 Minimização de uma Função Quadrática em Hiperplanos

Seja x_{k+1} o ponto mínimo de F num hiperplano H_k. Usar-se-á um conjunto de vetores não nulos $p_1, ..., p_k$ em H_k os quais são mutuamente conjugados no sentido que as relações

$$p_i^T A p_j = 0 \quad (i \neq j,\ i = 1, ..., k)$$

se mantém. Note que

$$p_i = p_i^T A p_i > 0, \quad (i = 1, ..., k)$$

porque $p_i \neq 0$ e A é definida positiva. Um conjunto de vetores mutuamente conjugados não nulos é um sistema conjugado (também denominados de $A-ortogonais$) e os vetores são linearmente independentes.

Teorema 34 *(direções conjugadas) Seja $p_0, p_1, ..., p_{n-1}$ um conjunto de vetores não nulos $A-ortogonais$. Para qualquer $x_0 \in \mathbb{R}^n$ a seqüência $\{x_k\}$ gerada de acordo com*

$$x_{k+1} = x_k + \alpha_k p_k \tag{3.18}$$

para $k \geq 0$, com

$$\alpha_k = -\frac{g_k^T p_k}{p_k^T A p_k} \tag{3.19}$$

e

$$g_k = A x_k - b \tag{3.20}$$

converge para a solução única x^*, de $Ax = b$ após n passos, isto é, $x_n = x^*$.

Demonstração. Já que os p_k são linearmente independentes, pode-se escrever

$$x^* - x_0 = \alpha_0 d_0 + \alpha_1 d_1 + ... + \alpha_{n-1} d_{n-1} \tag{3.21}$$

para algum conjunto de coeficientes α_k. Multiplicando por A e tomando o produto escalar com p_k encontra-se

$$\alpha_k = \frac{p_k^T A(x^* - x_0)}{p_k^T A p_k}. \tag{3.22}$$

Agora seguindo o processo iterativo de (3.18) de x_0 até x_k obtém-se

$$x_k - x_0 = \alpha_0 d_0 + \alpha_1 d_1 + ... + \alpha_{k-1} p_{k-1} \tag{3.23}$$

e portanto pela $A-ortogonalidade$ dos p_k segue que

$$p_k^T A(x_k - x_0) = 0. \tag{3.24}$$

Substituindo em (3.22) obtém-se

$$\alpha_k = \frac{p_k^T A(x^* - x_k)}{p_k^T A p_k} = -\frac{g_k^T p_k}{p_k A p_k} \tag{3.25}$$

que é idêntico a (3.19). ■

Teorema 35 *(subespaço expandido) Seja $p_0, p_1, p_2, ..., p_i$, uma seqüência de vetores $A-$ ortogonais não nulos em $\mathbb{R}^n$. Então para qualquer $x_0 \in \mathbb{R}^n$ a seqüência $\{x_k\}$ gerada de acordo com*

$$x_{k+1} = x_k + \alpha_k p_k \tag{3.26}$$

com

$$\alpha_k = -\frac{g_k^T p_k}{p_k^T A p_k} \tag{3.27}$$

possui a propriedade que x_k minimiza $F(x) = \frac{1}{2}x^T Ax - b^T x$ na linha $x = x_{k-1} + \alpha p_{k-1}$, $-\infty < \alpha < \infty$, bem como na variedade linear $x_0 + H_k$.

Demonstração. É necessário mostrar somente que x_k minimiza F na variedade linear $x_0 + H_k$, já que ela contém a linha $x = x_{k-1} + \alpha_k p_{k-1}$. Já que F é uma função estritamente convexa, a conclusão será mantida se puder ser demonstrado que g_k é ortogonal a H_k (isto é, o gradiente de F em x_k é ortogonal ao subespaço H_k).

Prova-se $g_k \perp H_k$ por indução. Já que H_0 é vazio essa hipótese é verdadeira para $k = 0$. Assumindo que é verdade para k, isto é, assumindo $g_k \perp H_k$, mostra-se que $g_{k+1} \perp H_{k+1}$. Tem-se

$$g_{k+1} = g_k + \alpha_k A p_k \tag{3.28}$$

e portanto

$$p_k^T g_{k+1} = p_k^T g_k + \alpha_k p_k^T A p_k = 0 \tag{3.29}$$

pela definição de α_k.Também para $i < k$

$$p_i^T g_{k+1} = p_i^T g_k + \alpha_k p_i^T A p_k \tag{3.30}$$

O primeiro termo do lado direito da última equação desaparece devido à hipótese da indução, enquanto o segundo termo desaparece pela $A-$*ortogonalidade* dos d_i. Portanto

$$g_{k+1} \perp H_{k+1}. \tag{3.31}$$

■

3.9.3 Método das Direções Conjugadas

O método das Direções Conjugadas é um método iterativo para resolver um sistema linear de equações

$$Ax = b, \tag{3.32}$$

onde A é uma matriz $n \times n$ simétrica e definida positiva. O problema (3.32) pode ser escrito equivalentemente como o seguinte problema de minimização:

$$\phi(x) = \frac{1}{2}x^T Ax - b^T x, \tag{3.33}$$

isto é, ambos (3.32) e (3.33) tem a mesma solução única. Essa equivalência permitirá interpretar o método do gradiente conjugado ou como um algoritmo que resolve sistema lineares ou como uma técnica para minimização de funções convexas quadráticas. Nota-se que o gradiente de ϕ é igual ao resíduo do sistema linear, isto é,

$$\nabla\phi(x) = Ax - b \stackrel{def}{=} r(x). \tag{3.34}$$

Uma das notáveis propriedades do método das Direções Conjugadas é sua capacidade de gerar, de um modo muito econômico, um conjunto de vetores com uma propriedade conhecida como *conjugacidade*. Um conjunto de vetores não-nulos $\{p_0, p_1, ..., p_k\}$ é dito ser conjugado com respeito a matriz definida positiva A se

$$p_j^T Ap_j = 0, \quad \text{para todo } i \neq 0. \tag{3.35}$$

A importância da conjugacidade surge do fato de que pode-se minimizar $\phi(\cdot)$ em n passos minimizando-a sucessivamente ao longo de direções individuais de um conjunto conjugado. Para verificar essa afirmação considera-se o seguinte método de direção conjugada. Dado um ponto inicial $x_0 \in \mathbb{R}^n$ e um conjunto de direções conjugadas $\{p_0, p_1, ..., p_{n-1}\}$, gera-se a seqüência $\{x_k\}$ fazendo-se

$$x_{k+1} = x_k + \alpha_k p_k, \tag{3.36}$$

onde α_k é um minimizador unidimensional da função quadrática $\phi(\cdot)$ ao longo de $x_k + \alpha_k p_k$, dado explicitamente por

$$\alpha_k = -\frac{r_k^T p_k}{p_k^T Ap_k}. \tag{3.37}$$

A seguir, tem-se o seguinte resultado.

Teorema 36 *Para qualquer $x_0 \in \mathbb{R}^n$ a seqüência $\{x_k\}$ gerada pelo algoritmo de direção conjugada (3.36), (3.37) converge para a solução x^* do sistema linear (3.32) em no máximo n passos.*

Demonstração. Visto que as direções $\{p_i\}$ são linearmente independentes, elas podem gerar todo o espaço $\mathbb{R}^n$. Portanto, pode-se escrever a diferença entre x_0 e a solução x^* como:

$$x^* - x_0 = \sigma_0 p_0 + \sigma_1 p_1 + \cdots + \sigma_{n-1} p_{n-1}, \tag{3.38}$$

para a mesma escolha dos escalares σ_k. Pre-multiplicando esta expressão por $p_k^T A$ e utilizando a propriedade de conjugacidade (3.35), obtém-se

$$\sigma_k = \frac{p_k^T A(x^* - x_0)}{p_k^T A p_k}. \tag{3.39}$$

Estabelecendo agora o resultado mostrado que estes coeficientes σ_k coincidem com o tamanho do passo α_k gerado pela fórmula (3.37).

Se x_k é gerado pelo algoritmo (3.36), (3.37), então tem-se

$$x_k = x_0 + \alpha_0 p_0 + \alpha_1 p_1 + \cdots + \alpha_{k-1} p_{k-1}. \tag{3.40}$$

Pré-multiplicando esta expressão por $p_k^T A$ e usando a propriedade de conjugacidade, tem-se que

$$p_k^T A(x_k - x_0) = 0, \tag{3.41}$$

e portanto

$$p_k^T A(x^* - x_0) = p_k^T A(x^* - x_k) = p_k^T (b - Ax_k) = -p_k^T r_k. \tag{3.42}$$

Comparando esta relação com (3.37) e (3.39), encontra-se que $\sigma_k = \alpha_k$, dando o resultado. ■

Teorema 37 *Seja $x_0 \in \mathbb{R}^n$ algum ponto inicial e suponha que a seqüência $\{x_k\}$ é gerada pelo algoritmo de direção conjugada (3.36), (3.37). Então*

$$r_k^T p_i = 0 \quad \textit{para } i = 0, ..., k-1 \tag{3.43}$$

e x_k é o mínimo de $\phi(x) = \frac{1}{2} x^T A x - b^T x$ sobre o conjunto

$$\{x | x = x_0 + span\{p_0, p_1, ..., p_{k-1}\}\}. \tag{3.44}$$

Demonstração. Mostrando que o ponto $\widetilde{x}$ minimiza ϕ sobre o conjunto (3.44) se e somente se $r(\widetilde{x})^T p_i = 0$, para cada $i = 0, 1, ..., k-1$. Seja definida $h(\sigma) = \phi(x_0 + \sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1})$, onde $\sigma = (\sigma_0, \sigma_1, \ldots, \sigma_{k-1})^T$. Visto que $h(\sigma)$ é uma quadrática estritamente convexa, tendo um único minimizador σ^* que satisfaz

$$\frac{\partial h(\sigma^*)}{\partial \sigma_i} = 0, \qquad i = 0, 1, ..., k-1. \tag{3.45}$$

Pela regra da cadeia, isto implica que

$$\nabla\phi(x_0 + \sigma_0^* p_0 + \cdots + \sigma_{k-1}^* p_{k-1})^T p_i = 0, \qquad i = 0, 1, ..., k-1. \tag{3.46}$$

Lembrando da definição (3.33), obtém-se o resultado desejado.

Usando agora a indução para mostrar que x_k satisfaz (3.43). Visto que α_k é sempre o minimizador unidimensional, tem-se imediatamente que $r_1^T p_0 = 0$. Fazendo a hipótese por indução, isto é, que $r_{k-1}^T p_i = 0$ para $i = 0, ..., k-2$. Segue

$$r_k = r_{k-1} + \alpha_{k-1} A p_{k-1}, \tag{3.47}$$

e tem-se

$$p_{k-1}^T r_k = p_{k-1}^T r_{k-1} + \alpha_{k-1} p_{k-1}^T A p_{k-1} = 0, \tag{3.48}$$

pela definição (3.37) de α_{k-1}. Enquanto isso, para os outros vetores p_i, $i = 0, 1, ..., k-2$, tem-se

$$p_i^T r_k = p_i^T r_{k-1} + \alpha_{k-1} p_i^T A p_{k-1} = 0 \tag{3.49}$$

pela hipótese de indução e a conjugacidade de p_i. Conclui-se que $r_k^T p_i = 0$, para $i = 0, 1, ..., k-1$, e a prova esta completa. ■

3.9.4 Propriedades básicas do método do Gradiente Conjugado

O método do Gradiente Conjugado é um método de direção conjugada com uma propriedade muito especial: na geração do conjunto de vetores conjugados, pode-se calcular um novo vetor p_k usando somente o vetor anterior p_{k-1}. Não se precisa conhecer todos os elementos anteriores $p_0, p_1, ..., p_{k-2}$ do conjunto conjugado; p_k é automaticamente conjugado destes vetores. Esta notável propriedade implica que o método requer pouco cálculo e armazenagem.

Agora, os detalhes do método do Gradiente Conjugado (CG). Toda direção p_k é escolhida como uma combinação linear da máxima direção de descida $-\nabla\phi(x_k)$ (que é o mesmo que o residual negativo $-r_k$) e direção anterior p_{k-1}. Escrevendo

$$p_k = -r_k + \beta_k p_{k-1} \tag{3.50}$$

onde o escalar β_k é determinado para exigir que p_{k-1} e p_k sejam conjugadas com respeito a A. Pré-multiplicando (3.50) por $p_{k-1}^T A$ e impondo a condição $p_{k-1}^T A p_k = 0$, tem-se que

$$\beta_k = \frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}. \tag{3.51}$$

Escolhe-se da primeira direção p_0 sendo a máxima direção de descida para o ponto inicial x_0.

Como no método de direção conjugada geral, o método de direções conjugadas representa minimizações sucessivas unidimensionais ao longo de cada direção. O algoritmo CG é expresso a seguir:

Algoritmo - CG

Dado x_0*;*

Faça $r_0 = Ax_0 - b$, $p_0 \longleftarrow -r_0$, $k \longleftarrow 0$*;*

Enquanto $r_k \neq 0$

$$\alpha_k \longleftarrow -\frac{r_k^T p_k}{p_k^T A p_k}; \tag{3.52}$$

$$x_{k+1} \longleftarrow x_k + \alpha_k p_k; \tag{3.53}$$

$$r_{k+1} \longleftarrow Ax_{k+1} - b; \tag{3.54}$$

$$\beta_{k+1} \longleftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k}; \tag{3.55}$$

$$p_{k+1} \longleftarrow -r_{k+1} + \beta_{k+1} p_k; \tag{3.56}$$

$$k \longleftarrow k + 1; \tag{3.57}$$

Fim (enquanto)

Teorema 38 *Suponha que a k-ésima iteração gerada pelo método do gradiente conjugado não seja a solução x^*. As quatro propriedades seguintes mantém-se:*

$$r_k^T r_i = 0, \quad i = 0, \ldots, k-1 \tag{3.58}$$

$$span\{r_0, r_1, \ldots, r_k\} = span\{r_0, Ar_0, \ldots, A^k r_0\}, \tag{3.59}$$

$$span\{p_0, p_1, \ldots, p_k\} = span\{r_0, Ar_0, \ldots, A^k r_0\}, \tag{3.60}$$

$$p_k^T A p_i = 0, \quad para \quad i = 0, 1, ..., k-1 \tag{3.61}$$

Portanto, a seqüência $\{x_k\}$ converge para x^* em no máximo em n passos.

Demonstração. A prova é por indução. A expressão (3.59) e (3.60) mantém-se trivialmente para $k = 0$ enquanto (3.61) mantém-se por construção para $k = 1$. Assumindo agora que estas três expressões são verdadeiras para algum k (a hipótese indução), elas continuam a manter-se para $k + 1$.

Para provar (3.59), mostra-se primeiro que o conjunto do lado esquerdo está contido no conjunto do lado direito. Por causa da hipótese de indução, tem-se de (3.59) e (3.60) que

$$r_k \in span\{r_0, Ar_0, \ldots, A^k r_0, \}, \quad p_k \in span\{r_0, Ar_0, \ldots, A^k r_0\}, \tag{3.62}$$

enquanto multiplicando a segunda dessas expressões por A, segue que

$$Ap_k \in span\{Ar_0, \ldots, A^{k+1} r_0\}. \tag{3.63}$$

Obtém-se que

$$r_{k+1} \in span\{r_0, Ar_0, \ldots, A^{k+1} r_0\}. \tag{3.64}$$

Combinando esta expressão com a hipótese de indução para (3.59), conclui-se que

$$span\{r_0, r_1, \ldots, r_k, r_{k+1}\} \in span\{r_0, Ar_0, \ldots, A^{k+1} r_0\}, \tag{3.65}$$

Para provar que a inclusão reversa mantém-se, usando a hipótese de indução em (3.60) para deduzir que

$$A^{k+1} r_0 = A(A^k r_0) \in span\{Ap_0, Ap_1, \ldots, Ap_k\}. \tag{3.66}$$

Visto que se tem $Ap_i = (r_{i+1} - r_i)/\alpha_i$ para $i = 0, 1, ..., k$, segue que

$$A^{k+1} r_0 \in span\{r_0, r_1, \ldots, r_{k+1}\}. \tag{3.67}$$

Combinando esta expressão com a hipótese de indução para (3.59), encontra-se que

$$span\{r_0, Ar_0, \ldots, A^{k+1} r_0\} \subset span\{r_0, r_1, \ldots, r_k, r_{k+1}\}.$$

Portanto, a relação (3.59) continua para manter-se quando k é substituído por $k+1$, como afirmado.

Mostrando agora que (3.60) continua a manter-se quando k é substituído por $k+1$ pelo seguinte argumento:

$$\begin{aligned} &span\{p_0, p_1, \ldots, p_k, p_{k+1}\} \\ &\qquad = span\{p_0, p_1, \ldots, p_k, p_{k+1}\} \quad \text{por} \quad (3.56) \\ &\qquad = span\{r_0, Ar_0, \ldots, A^k r_0, r_{k+1}\} \quad \text{por hipótese de indução de (3.60)} \\ &\qquad = span\{r_0, r_1, \ldots, r_k, r_{k+1}\} \quad \text{por (3.59)} \\ &\qquad = span\{r_0, Ar_0, \ldots, A^{k+1} r_0\} \text{ por (3.59) para } k+1. \end{aligned}$$

Provando agora a condição de conjugacidade (3.61) com k substituido por $k+1$. Multiplicando (3.56) por Ap_i, $i = 0, 1, \ldots, k$, obtém-se

$$p_{k+1}^T A p_i = -r_{k+1}^T A p_i + \beta_{k+1} p_k^T A p_i \tag{3.68}$$

Pela definição (3.55) de β_k, o lado direito de (3.68) desaparece quando $i = k$. Para $i \leq k-1$ precisa-se coletar um número de observações. Observe primeiro que a hipótese de indução para (3.61) implica que as direções $p_0, p_1, \ldots, pk$ são conjugadas, deduzir que

$$r_{k+1}^T p_i = 0, \quad \text{para } i = 0, 1, \ldots, k. \tag{3.69}$$

Segundo, repetidamente aplicando (3.60), encontra-se para $i = 0, 1, \ldots, k-1$, a seguinte inclusão mantém-se:

$$\begin{aligned} Ap_i &\in A\ span\{r_0, Ar_0, \ldots, A^i r_0\} = span\{Ar_0, A^2 r_0, \ldots, A^{i+1} r_0\} \\ &\subset span\{p_0, p_1, \ldots, p_{i+1}\}. \end{aligned} \tag{3.70}$$

Combinando (3.69) e (3.70), deduz-se que

$$r_{k+1}^T A p_i = 0, \quad \text{para} \quad i = 0, 1, \ldots, k+1, \tag{3.71}$$

e o primeiro termo do lado direito de (3.68) desaparece para $i = 0, 1, \ldots, k-1$. Por causa da hipótese de indução de (3.61), o segundo termo desaparece, e conclui-se que $p_{k+1}^T A p_i = 0$, $i = 0, 1, , k$. Portanto o argumento de indução mantém-se para (3.68) também.

Segue que o conjunto de direções geradas pelo método do gradiente conjugado é de fato um conjunto de direções conjugadas, assim o algoritmo termina em no máximo n iterações.

Finalmente, provando (3.58) por argumento não indutivo. Devido o conjunto direção ser conjugado, tem-se que $r_k^T p_i = 0$ para todo $i = 0, 1, \ldots, k-1$ e algum $k = 1, 2, \ldots, n-1$. Reorganizando (3.56), tem-se que

$$p_i = -r_i + \beta_i p_{i-1}, \tag{3.72}$$

e que $r_i \in span\{p_i, p_{i-1}\}$ para todo $i = 1, \ldots, k-1$. Conclui-se que $r_k^T r_i = 0$ para todo $i = 1, \ldots, k-1$, como afirmado. ■

A prova deste teorema depende da primeira direção p_0 que é a máxima direção de descida $-r_0$; de fato, o resultado não mantém-se para outra escolha de p_0. Assim, os gradientes r_k são mutuamente ortogonais.

3.9.5 Uma forma prática do método do Gradiente Conjugado

A seguir, a forma prática padrão do método do gradiente conjugado.

$$\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}. \tag{3.73}$$

Segundo, tem-se que $\alpha_k A p_k = r_{k+1} - r_k$, logo

$$\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}. \tag{3.74}$$

Algoritmo - CG Prático

Dado x_0*;*

Faça $r_0 = Ax_0 - b$, $p_0 \longleftarrow -r_0$, $k \longleftarrow 0$*;*

Enquanto $r_k \neq 0$

$\alpha_k \longleftarrow \frac{r_k^T r_k}{p_k^T A p_k}$;

$x_{k+1} \longleftarrow x_k + \alpha_k p_k$;

$r_{k+1} \longleftarrow r_k + \alpha_k A p_k$;

$\beta_{k+1} \longleftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$;

$p_{k+1} \longleftarrow -r_{k+1} + \beta_{k+1} p_k$;

$k \longleftarrow k + 1$;

Fim (enquanto)

O método CG é recomendado somente para problemas grandes; caso contrário, eliminação Gaussiana ou outro algoritmo de fatoração como decomposição de valor singular são preferidos, visto que eles são menos sensíveis aos erros de arredondamento na implementação computacional.

4
Minimização de uma Função Quadrática Sujeita a Restrições Lineares de Igualdade

Vamos considerar o problema

$$\begin{array}{cc} minimizar & f(x) = \frac{1}{2}x^T Gx + h^T x + c, \\ s.a. & Ax = b \end{array}$$

onde f é uma função quadrática, definida positiva, e A é uma matriz $m \times n$, com $n > m$. Usualmente o algoritmo de direções conjugadas para resolver esse problema gera uma seqüência de direções viáveis no núcleo de A, e então uma seqüência de pontos viáveis resolvendo a cada iteração o sistema reduzido, $Z^T \nabla^2 f(x) Z d_k = -Z^T \nabla f(x)$, onde Z é uma matriz cujas colunas formam uma base para o núcleo de A. Mais especificamente,

$$\begin{array}{cc} minimizar & f(x) = \frac{1}{2}x^T Gx + h^T x + c \\ s.a. & Ax = b \end{array}$$

é equivalente a,

$$\underset{d \in \mathbb{R}^{n-m}}{minimizar} \quad \nabla f(x)^T Z d + \frac{1}{2} d^T Z^T \nabla^2 f(x) Z d$$

que é equivalente a resolver o sistema linear

$$Z^T \nabla^2 f(x) Z d = -Z^T \nabla f(x). \tag{4.1}$$

A solução ótima do problema é então $x^* = x_0 + Zd$.

A abordagem do problema a ser feita aqui é diferente. Conquanto ainda utilize o algoritmo de direções conjugadas, o procedimento não implica em redução da dimensionalidade do problema, isto é, não se utiliza a hessiana reduzida ($Z^T\nabla^2 f(x)Z$), nem o gradiente reduzido ($Z^T\nabla f(x)$) para a montagem do sistema (4.1). Além disso, as direções conjugadas não são obtidas em $\mathbb{R}^{n-m}$, mas diretamente no espaço nulo de A, isto é, em um subespaço de $\mathbb{R}^n$. Nesse sentido o algoritmo resolve o problema na variedade $x_0 + B_i$, $i = 1, 2, ..., n-m$, onde B_i é o subespaço gerado pelas direções conjugadas obtidas pelo algoritmo.

4.1 O Modelo Quadrático

Um problema de programação quadrática (QP) surge, muitas vezes, como um subproblema em métodos de otimização restrita geral, razão pela qual deve ser sempre resolvido de maneira mais simples e barata possível.

Seja o problema de programação quadrática restrito dado por

$$\begin{array}{cc} minimizar & f(x) = \frac{1}{2}x^T Gx + h^T x + c, \\ s.a. & Ax = b \end{array} \tag{4.2}$$

onde G é uma matriz simétrica e definida positiva $n \times n$, h é um vetor de $\mathbb{R}^n$, A uma matriz $m \times n$ com $m < n$, de posto m, e b um vetor de $\mathbb{R}^m$. As condições de otimalidade de primeira ordem para o poblema (4.2) são,

$$\begin{aligned} \nabla_x l(x^*, \lambda^*) &= 0 \\ Ax^* &= b. \end{aligned}$$

Como $l(x, \lambda) = \frac{1}{2}x^T Gx - h^T x + \lambda^T(b - Ax)$, então as condições anteriores resultam no sistema,

$$\begin{aligned} Gx^* - h - A^T\lambda^* &= 0 \\ Ax^* - b &= 0 \end{aligned} \tag{4.3}$$

logo, as condições de otimalidade de primeira ordem são dadas pelo seguinte sistema linear,

$$\begin{pmatrix} G & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} h \\ b \end{pmatrix}. \tag{4.4}$$

A matriz do sistema (4.2) é chamada de Karush-Kuhn-Tucker (KKT) e o sistema de equações lineares é chamado de sistema KKT.

Sob certas condições sobre as matrizes G e A, as condições de otimimalidade de primeira ordem do sistema assume que existe uma única solução para o sistema (4.4), como pode-se observar no seguinte teorema,

Teorema 39 *Suponha que A seja uma matriz $m \times n$, com $m < n$, de posto completo m, e que G seja definida positiva no espaço nulo de A. Então, a matriz Karush-Kuhn-Tucker é não-singular e existe uma única solução (x^*, λ^*) que satisfaça o sistema (4.4).*

Demonstração. Sejam x e y dois vetores arbitrários tais que

$$\begin{pmatrix} G & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0. \tag{4.5}$$

Se $x = 0$, $A^T y = 0$, então $y = 0$ desde que A tenha posto completo. Se $x \neq 0$, $Ax = 0$, então, $x = Zw$ para qualquer vetor não nulo w, onde Z é uma matriz cujas colunas formam uma base para o espaço nulo de A, assim

$$0 = \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} G & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^T G x.$$

Segue que,

$$0 = x^T G x = w^T Z^T G Z w, \tag{4.6}$$

e tal w sendo zero, implicaria em, indo contra todas as hipóteses anteriores que $x = 0$. Assim, a equação (4.5) tem uma única solução se e somente se o vetor $[x\ y]^T = 0$. ■

Levando em consideração a não singularidade da matriz KKT, uma solução para o sistema (4.4) seria muito problemática do ponto de vista computacional visto que se tem de resolver um sistema quadrado aumentado de ordem m para $m + n$, sem saber com antecedência se as matrizes A e G satisfazem as condições do sistema acima. Além disso, do ponto de vista de otimização a matriz KKT é uma hessiana de uma função indefinida, a qual é bem sabido ter problemas na manipulação [19]. Agora será explorado como resolver o sistema (4.3) para x^* e λ^* ótimos.

4.1.1 Uma Abordagem do Espaço Nulo

O método do espaço nulo, descrito a seguir, não exige a não singularidade da matriz de G, mas somente que as condições do Teorema 39 se cumpram, isto é, A deve ter posto completo e G deve ser definida positiva no espaço nulo de A.

Seja Z uma matriz de dimensões $n \times (n-m)$, cujas colunas formam uma base para o espaço nulo de A, e seja R uma matriz de dimensão $n \times m$ cujas colunas formam uma base para o espaço imagem de A^T. Observe que como A é uma matriz de dimensão $m \times n$, com $m < n$, e posto completo m, então uma escolha trivial para R é fazer $R = A^T$. Como toda solução x^* para o problema (4.4) é um ponto do $\mathbb{R}^n$, então ele pode ser escrito de maneira única como segue,

$$x^* = Ru + Zv,$$

onde $u \in R^m$ e $v \in R^{n-m}$. Além disso, como a matriz A tem posto completo m e a matriz $[R|Z]$ tem posto completo n, então $A[\ R \mid Z\ ]$ tem posto completo m. Como $A[\ R \mid Z\ ] = [\ AR \mid 0\ ]$, então AR é uma matriz $m \times m$ e não singular. Agora, considere o sistema Karush-Kuhn-Tucker (4.3), o qual é dado pelas equações

$$\begin{array}{rcl} Gx^* - A^T\lambda^* & = & h \\ Ax^* & = & b. \end{array} \tag{4.7}$$

Como $x^* = Ru^* + Zv^*$, então temos que,

$$Ax^* = ARu^* + AZv^*,$$

donde se tem que,

$$ARu^* = b \quad \text{ou} \quad u^* = (AR)^{-1}\, b.$$

Multiplicando-se a primeira equação do sistema (4.7) por Z^T e usando o fato de que $x^* = Ru^* + Zv^*$, segue-se que,

$$Z^TGRu^* + Z^TGZv^* - Z^TA^T\lambda^* = Z^Th,$$

donde,

$$Z^TGRu^* + Z^TGZv^* = Z^Th,$$

e finalmente v^* é solução do sistema,

$$Z^T G Z v^* = Z^T h - Z^T G R u^*,$$

ou seja,

$$v^* = \left(Z^T G Z\right)^{-1} \left[Z^T h - Z^T G R u^*\right].$$

Com a obtenção de u^* e v^*, x^* fica determinado como sendo,

$$x^* = R\left(AR\right)^{-1} b + Z\left(Z^T G Z\right)^{-1} \left[Z^T h - Z^T G R u^*\right].$$

Para determinar o valor de λ^*, multiplica-se a primeira equação de (4.7) por R^T, obtendo-se,

$$R^T G x^* - R^T A^T \lambda^* = R^T h,$$

donde vem que,

$$\begin{aligned} \left(AR\right)^T \lambda^* &= R^T \left(G x^* - h\right) \\ &= R^T \nabla f(x^*), \end{aligned}$$

e finalmente,

$$\lambda^* = \left(AR\right)^{-T} R^T \nabla f(x^*).$$

Observe que a obtenção de λ^* não é cara do ponto de vista computacional tendo em vista que a matriz AR já foi anteriormente fatorada para a obtenção da solução x^*. Portanto, λ^* pode ser obtido ao custo da ordem $O(n^2)$ operações.

Esse método, claramente, é bastante atraente se a dimensão do núcleo de A for pequeno. Por outro lado, como temos que resolver um sistema cuja matriz é $Z^T G Z$, é conveniente que Z seja escolhida ortogonal, para não aumentar o número de condição do sistema. Se Z é ortogonal então o número de condição da matriz $(Z^T G Z)$ não é maior do que o número de condição da matriz G, isto é,

$$cond(Z^T G Z) \leq cond(G).cond(Z)^2,$$

donde então,

$$cond(Z^T G Z) \leq cond(G),$$

ou seja, a condição do sistema reduzido, $Z^T G Z$, é pelo menos tão boa quanto a condição do sistema de G. As equações anteriores podem ser bastante simplificadas se introduzirmos nessa abordagem alguns elementos de conjugacidade.

4.1.2 Abordagem de Conjugacidade

A idéia fundamental por trás dessa abordagem de conjugacidade é construir uma matriz M tal que:

$$\begin{array}{rrcl} (i) & M^T G M & = & I \\ (ii) & AM & = & [\ 0 \quad U\ ]. \end{array}$$

Suponha que G, obtida desse modo, seja definida positiva e a matriz M é uma matriz cujas colunas são vetores G-conjugados, que pode ser construída do seguinte modo. Suponha M decomposta em blocos, como sendo,

$$M = [\ M_1 \quad M_2\ ],$$

onde M_1 é uma matriz $n \times n-m$ e M_2 é uma matriz $m \times m$ tal que,

$$\begin{array}{ll} (i) & AM_1 = 0, \text{ isto é, as } m-n \text{ primeiras colunas de } M \text{ devem estar no espaço nulo de } A; \\ (ii) & AM_2 = U, \text{ onde } U \text{ é uma } m \times m \text{ matriz triangular superior.} \end{array} \tag{4.8}$$

Como A é uma matriz $m \times n$, $m < n$, com posto completo, isto é, $posto(A) = m$, fazemos a fatoração QR de A^T para obter,

$$A^T = QR,$$

onde Q é uma matriz ortogonal de dimensões $n \times n$ cujas primeiras m colunas de Q geram o espaço vetorial $\mathbb{R}(A^T)$, e as $n-m$ últimas colunas geram o núcleo de A, $\aleph(A)$. Então particionamos Q do seguinte modo,

$$Q = [\ Q_1 \quad Q_2\ ].$$

onde Q_1 é uma matriz $m \times m$, e Q_2 é uma matriz $m \times n-m$, e correspondentemente particionamos R como,

$$R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}.$$

onde R_1 é uma matriz $m \times m$ triangular superior, e 0 é a matriz nula de dimensão $n-m \times m$.

Definimos agora uma matriz de permutação P que permuta os blocos de Q do seguinte modo.

$$\overline{Q} = QP = [\ Q_2 \quad Q_1\ ],$$

onde, $Q_1 = [q_j]$, para $j = 1, 2, ..., m$ e $Q_2 = [q_j]$, para $j = m+1, ..., n$.

Agora seja X uma matriz definida como

$$\begin{aligned} X &= A^T(AA^T)^{-1}R^1 \\ &= (Q_1R_1)(R_1^TQ_1^TQ_1R_1)^{-1}R_1 \\ &= Q_1R_1R_1^{-1}R_1^{-T}R_1 \\ &= Q_1R_1^{-T}R_1. \end{aligned}$$

A matriz X tem as seguintes propriedades relevantes:

1. $AX = AA^T(AA^T)^{-1}R_1 = R_1$; isto é, AX é uma matriz triangular superior.
2. $Q_2^TX = Q_2^TQ_1(R_1^{-T}R_1) = 0$; isto é, as colunas de Q_2 são ortogonais as colunas de X.
3. $X^TQ_2 = (R_1^TR_1^{-1})Q_1^TQ_2 = 0$; isto é, as colunas de X são ortogonais as colunas de Q_2.

Defina agora a matriz $\widetilde{Q}$ como $\widetilde{Q} = [Q_2|X]$. $\widetilde{Q}$ é quadrada de posto completo n, portanto não singular. Logo, se G é uma matriz definida positiva de qualquer tamanho n, $\widetilde{Q}^TG\widetilde{Q}$, também é definida positiva, e podemos fazer sua fatoração de Cholesky

$$\widetilde{Q}^TG\widetilde{Q} = LL^T.$$

Seja $M = \widetilde{Q}L^{-T}$, então

$$\begin{aligned} M^TGM &= (\widetilde{Q}L^{-T})^TG(\widetilde{Q}L^{-T}) \\ &= L^{-1}\widetilde{Q}^TG\widetilde{Q}L^{-T} \\ &= L^{-1}(LL^T)L^{-T} \\ &= I. \end{aligned}$$

Além disso,

$$\begin{aligned} AM &= A\widetilde{Q}L^{-T} \\ &= A\begin{pmatrix} Q_2 & | & X \end{pmatrix} L^{-T} \\ &= \begin{pmatrix} 0 & | & AX \end{pmatrix} L^{-T} \\ &= \begin{pmatrix} 0 & | & R_1 \end{pmatrix} L^{-T} \\ &= \begin{pmatrix} 0 & | & U \end{pmatrix} \end{aligned} \tag{4.9}$$

Portanto, M como definida acima, satisfaz as condições (i) e (ii) da equação (4.8). Por construção M pode ser particionada como sendo,

$$M = \begin{bmatrix} Z & | & R \end{bmatrix},$$

onde Z é uma base para o espaço nulo de A e R é uma base para o espaço imagem de A^T. Tem-se então que,

$$\begin{aligned} I &= M^T G M \\ &= \begin{pmatrix} Z & | & R \end{pmatrix}^T G \begin{pmatrix} Z & | & R \end{pmatrix} \\ &= \begin{pmatrix} Z^T \\ R^T \end{pmatrix} G \begin{pmatrix} Z & | & R \end{pmatrix} \\ &= \begin{pmatrix} Z^T \\ R^T \end{pmatrix} \begin{pmatrix} GZ & | & GR \end{pmatrix} \\ &= \begin{pmatrix} Z^T G Z & Z^T G R \\ R^T G Z & R^T G R \end{pmatrix} \end{aligned}$$

donde então

$$\begin{aligned} Z^TGZ &= I \\ Z^TGR &= 0 \\ R^TGZ &= 0 \\ R^TGR &= I. \end{aligned}$$

Essas relações obtidas facilitam sobremaneira a solução dos sistemas lineares obtidas na seção 4.1.1 pela abordagem do espaço nulo. De fato, a equação $(AR)u^* = b$, pode ser simplificada do seguinte modo,

$$AM = A\left[\begin{array}{c|c} Z & R \end{array}\right] = \left[\begin{array}{c|c} AZ & AR \end{array}\right] = \left[\begin{array}{c|c} 0 & U \end{array}\right],$$

e portanto $(AR)u^* = b$ pode ser simplificado para $Uu^* = b$, que é um sistema triangular superior. A equação $(Z^TGZ)v^* = Z^Th - Z^TGRu^*$ fica reduzida a $v^* = Z^Th$, e finalmente, $(AR)^T\lambda^* = R^T\nabla f(x^*)$, fica reduzida a $U^T\lambda^* = R^T\nabla f(x^*)$.

4.2 Descrição do algoritmo

4.2.1 Inicialização

A inicialização do algoritmo depende da escolha de um ponto inicial viável x_0, isto é, $Ax_0 = b$, cuja escolha pode ser feita por qualquer método de solução de sistemas lineares. Aqui foi utilizado o seguinte esquema para encontrar um ponto inicial.

Seja A a matriz dada por (4.2) tal que por simplicidade, as m primeiras colunas de A sejam linearmente independentes, i.e., formem uma matriz básica B, e as $n - m$ colunas restante formem uma matriz não básica N. Portanto, a matriz A pode ser particionada da seguinte maneira:

$$A = \left(\begin{array}{cc} B & N \end{array}\right). \tag{4.10}$$

Segue que

$$Ax = b \Leftrightarrow (B \;\; N)\left(\begin{array}{c} x_B \\ x_N \end{array}\right) = b \tag{4.11}$$

o que implica

$$Bx_B + Nx_N = b \tag{4.12}$$

ou ainda,

$$Bx_B = b - Nx_N.$$

Fazendo $x_N = 0$, implica em resolver o sistema $Bx_B = b$, o qual tem solução x_{B_1}. O ponto inicial x_0 é dado por:

$$\begin{pmatrix} x_{B_1} \\ 0 \end{pmatrix}$$

De posse de x_0, calcula-se $g_0 = Gx_0 - h$.

A direção d_0 é uma direção inicial que pertence ao espaço nulo de A.

O comprimento de passo inicial é dado por: $\alpha_0 = -\frac{g_0^T d_0}{d_0^T G d_0}$.

4.2.2 Critério de Parada

Para o critério de parada, considere o seguinte problema

$$\begin{array}{cc} minimizar & \varphi(x) = \frac{1}{2}x^T Gx - h^T x + c \\ s.a. & Ax = b. \end{array}$$

Pela condição de otimalidade segue que,

$$\nabla\varphi(x^*) = A^T\lambda^* \Leftrightarrow \quad (\nabla\varphi(x^*) \in \Re(A^T)) \Leftrightarrow \quad \underset{z}{Minimizar} \quad \tfrac{1}{2}||A^T z - \nabla\varphi(x^*)||^2.$$

Seja o conjunto de direções conjugadas a $\aleph(A)$

$$B = \{d_1, d_2, ..., d_{n-m}\}, \quad [B] \equiv \aleph(A).$$

Seja $V_1 = x_0 + [d_1]$, $\nabla\varphi(x_1) \perp d_1$, mas $\nabla\varphi(x_1)$ não está necessariamente em $\Re(A^T)$. Se estiver, $x_1 = x^*$.

Observe que

$$V_1^\perp = [d_1]^\perp \supset \Re(A^T)$$

$$V_2 = x_0 + [\{d_1, d_2\}], \quad \nabla\varphi(x_2) \perp \{d_1, d_2\} \Rightarrow V_2^\perp \supset (A^T)$$

$$\vdots$$

$$V_{n-m}^\perp = [\{d_1, d_2, ..., d_{n-m}\}] \equiv \aleph(A) \Rightarrow V_{n-m}^\perp \equiv \aleph(A^T).$$

Solução do problema

$$\underset{z}{Minimizar} \quad \tfrac{1}{2}||A^T z - \nabla\varphi(x^*)||^2 \ .$$

Nós já temos $A^T = QR$

$$\begin{aligned} AA^T z &= A\nabla\varphi(x_k) \\ R^T Q^T Q R z &= A\nabla\varphi(x_k) \\ R^T R z &= A\nabla\varphi(x_k) \\ R^T Y &= A\nabla\varphi(x_k) \\ Rz &= Y \end{aligned}$$

Achamos z.

Então

$$\varphi(z) = \tfrac{1}{2}||A^T z - \nabla\varphi(x_k)||^2,$$

se $\varphi(z) = 0 \Rightarrow \nabla\varphi(x_k) \in \Re(A^T)$ e a condição de otimalidade se cumpre.

4.2.3 Iterações

Passo 1. Calcula-se o novo ponto: $x_{k+1} = x_k + \alpha_k d_k$;

Passo 2. Calcula-se o novo valor da função: f_{k+1};

Passo 3. Calcula-se o novo gradiente: g_{k+1};

Passo 4. Calcula-se o novo tamanho de passo: $\alpha_{k+1} = -\frac{g_{k+1}^T d_k}{d_k^T G d_k}$;

Passo 5. Calcula-se a nova direção $d_{k+1}^T G d_k = 0$;

4.2.4 Finalização

O algoritmo termina quando $\varphi(z) < tolerância$ ($tolerância = 10^{-8}$), isto é, quando o algoritmo verifica o critério de otimalidade, ou quando o número máximo de iterações for alcançado.

Se a condição $\varphi(z) < tolerância$ é violada, então $x_0 = x_{n-m}$ e retorna para as iterações.

4.2.5 Obtenção da G-conjugacidade das direções

Seja x_0 um ponto viável das restrições e $\{d_0, d_1, ..., d_{n-m}\}$ um conjunto de direções viáveis, linearmente independentes e considere que o algoritmo atualize a cada iteração

as direções viáveis. Se x_0 é um ponto viável de $Ax = b$, então todo algoritmo do tipo $x_{k+1} = x_k + \alpha_k d_k$ obriga que $d_k \in \aleph(A)$. As direções de busca, p_k, do algoritmo, são obtidas do seguinte modo:

$$p_o = d_o, \tag{4.13}$$

e

$$p_k = d_k - \sum_{i=1}^{k-1} \frac{\langle d_k, p_i \rangle_G}{\langle p_i, p_i \rangle_G} p_i, \quad \text{para} \quad k = 1, 2, ..., n - m + 1. \tag{4.14}$$

As direções p_k obtidas desse modo têm pelo menos duas propriedades desejáveis. A primeira propriedade é que $p_k \in \aleph(A)$.

De fato,

$$Ap_k = Ad_k - \sum_{i=1}^{k-1} \beta_i A p_i, \tag{4.15}$$

onde $\beta_i = \frac{\langle d_k, p_i \rangle_G}{\langle p_i, p_i \rangle_G}$ e $\langle x, y \rangle_G = x^T G y$.

Observe que para $k = 0$, $p_0 = d_0$ pertence ao $\aleph(A)$ para $k = 1$,

$$Ap_1 = Ad_1 - \beta_0 A p_0 = 0$$

Para $k = 2$

$$Ap_2 = Ad_2 - \beta_0 A p_0 - \beta_1 A p_1 = 0$$

E assim sucessivamente.

A segunda propriedade é que as direções p_k são G-conjugadas.

De fato, para todo i diferente de j, tem-se

$$\begin{aligned} p_1^T G p_2 &= p_1^T G d_2 - \frac{\langle d_2, p_1 \rangle_G}{\langle p_1, p_1 \rangle_G} p_1^T G p_1 \\ &= p_1^T G d_2 - \langle d_2, p_1 \rangle_G = 0 \end{aligned}$$

p_1 e p_2 são G-conjugados.

Logo,

$$\begin{aligned} p_1^T G p_3 &= p_1^T G d_3 - \frac{\langle d_3, p_1 \rangle_G}{\langle p_1, p_1 \rangle_G} p_1^T G p_1 - \alpha_2 d_1^T G d_2 \\ &= 0 \end{aligned}$$

$$\begin{aligned} p_2^T G p_3 &= p_2^T G d_3 - \alpha_1 d_2^T G d_1 - \frac{\langle d_3, p_2 \rangle_G}{\langle p_2, p_2 \rangle_G} p_2^T G p_2 \\ &= 0 \end{aligned}$$

p_1, p_2 e p_3 são G-conjugados.

E assim sucessivamente, pode-se ir gerando direções que são mutuamente G-conjugadas e que estão no espaço nulo de A.

O *novo algoritmo* a ser apresentado nesse trabalho resolve o problema (4.2), conquanto satisfaz as restrições impostas ao mesmo e avalia se a condição de otimalidade de primeira ordem se verifica.

4.3 Algoritmo I - Versão QR de A^T

Inicialização

Como descrito na página 64.

Escolha das Direções

Fazendo a decomposição QR de A^T, onde $A^T = QR = \begin{pmatrix} Q_1 & \vdots & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ \cdots \\ 0 \end{pmatrix}$,

- $[Q_1]$ gera $\Re(A^T)$;
- $[Q_2]$ gera $\aleph(A^T)$.

As colunas de Q_2 formam uma base para o núcleo de A,

$$Q_2 = \begin{pmatrix} q_1 & q_2 & \dots & q_{n-m} \end{pmatrix}.$$

Faça $Z = Q_2$. Logo, $z_1 = q_1, z_2 = q_2, \dots, z_{n-m} = q_{n-m}$ pertencem ao espaço nulo de A, pois $Az_1 = Az_2 = \cdots = Az_{n-m} = 0$.

Iterações

Como descrito na página 66.

Finalização

Como descrito na página 66.

Algoritmo 1

Passo 1. Encontre x_k *tal que* $Ax_k = b$*;*

Passo 2. Calcule a função para x_k*,* $f(x_k)$*;*

Passo 3. Calcule o gradiente $g(x_k)$*;*

Passo 4. Seja $d_1 = q_1$*;*

Passo 5. Calcule $\varphi(z)$*;*

Passo 6. Calcule o tamanho do passo α_k*;*

Enquanto $\varphi(z) > tolerância$*;*

Seja $x_{k+1} = x_k + \alpha_k d_k$*;*

Calcule a função para x_{k+1}*,* $f(x_{k+1})$*;*

Calcule o gradiente $g(x_{k+1})$*;*

Avalie o critério de parada $\varphi(z)$*;*

Calcule $d_{k+1} \in \aleph(A)$*;*

Calcule α_{k+1}*;*

Atualize o contador $i = i + 1$*;*

fim

4.4 Algoritmo II - Versão $(B \quad N)$

Inicialização

Como descrito na página 64.

Escolha das Direções

A matriz A pode ser particionada como visto em (4.10) e o sistema $Ax = b$ fica substituido por (4.12). Logo, se tem de resolver

$$Bx_B = b - Nx_N. \tag{4.16}$$

Fazendo $x_N = 0$ em (4.16), implica em resolver o sistema $Bx_B = b$, o qual tem solução x_{B_1}. O vetor x_0 é dado por:

$$\begin{pmatrix} x_{B_1} \\ 0 \end{pmatrix}$$

Fazendo $x_N = e_1$, onde e_1 é o primeiro vetor canônico, implica em resolver o sistema $Bx_B = b - Ne_1$, o qual tem solução x_{B_2}. O vetor x_1 é dado por:

$$\begin{pmatrix} x_{B_2} \\ e_1 \end{pmatrix}$$

Seja $z_1 = x_1 - x_0$. Para $j = 2, 3, \ldots, n - m$

Construa o conjunto de vetores Z tal que as colunas sejam dadas por:

$$z_j = x_j - x_{j-1},$$

Os vetores v_j, $j = 1, 2, 3, .., n - m$ obtidos pertencem ao espaço nulo de A pois, se $Ax_0 = b$ e $Ax_1 = b$, fazendo $z_1 = x_1 - x_0$, então $Az_1 = A(x_1 - x_0) = b - b = 0$. Logo, para todo $j = 1, 2, 3, .., n - m$, $z_j = xj - xj - 1$ tem-se $Az_j = 0$ que é equivalente a dizer que $zj \in \aleph A$, $\forall j$.

Iterações

Como descrito na página 66.

Finalização

Como descrito na página 66.

Algoritmo 2

Passo 1. Faça $x_N = 0$ para $Bx_B + Nx_N = b$ e encontre x_0:

Passo 2. Faça $x_B = e_1$ para $Bx_B + Nx_N = b$ e encontre x_1;

Passo 3. Seja $d_0 = x_1 - x_0$;

Passo 4. Calcule a função para x_k, $f(x_k)$;

Passo 5. Calcule o gradiente $g(x_k)$;

Passo 6. Calcule $\varphi(z)$;

Passo 7. Calcule o tamanho do passo α_k;

Enquanto $\varphi(z) > tolerância$;

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

Calcule a função para x_{k+1}, $f(x_{k+1})$;

Calcule o gradiente $g(x_{k+1})$;

Avalie o critério de parada $\varphi(z)$;

Calcule $d_{k+1} \in \aleph(A)$, tal que $d_{k+1}^T G d_k = 0$;

Calcule α_{k+1};

Atualize o contador $i = i + 1$;

fim

4.5 Algoritmo III - Versão B^{-1}

Inicialização

Como descrito na página 64.

Escolha das Direções

A matriz A e o vetor x podem ser particionados como feito em (4.11). Dessa forma de particionar a matriz resulta o seguinte sistema :

$$Bx_B + Nx_N = b. \tag{4.17}$$

A matriz B é uma matriz não-singular, a qual possui inversa. Logo, pré-multiplicando B^{-1} em (4.17), obtém-se

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

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

$$x = \begin{pmatrix} B^{-1}b - B^{-1}Nx_N \\ x_N \end{pmatrix},$$

Logo, pode-se escrever x como

$$x = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix} b + \begin{pmatrix} -B^{-1}N \\ I_{n-m} \end{pmatrix} x_N. \tag{4.18}$$

Desse modo, uma base Z para o espaço nulo é dada pela matriz $n \times n - m$ de (4.18),

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

Pode-se escolher vetores z_i, $i = 1, 2, 3, \ldots, n - m$ pertencentes ao espaço nulo de forma que z_i seja igual a i-ésima coluna de Z. Desse modo, para qualquer vetor z_i, $i = 1, 2, 3, \ldots, n - m$ tem-se $Az_i = 0$.

Iterações

Como descrito na página 66.

Finalização

Como descrito na página 66.

Algoritmo 3

Passo 1. Encontre x_k tal que $Ax_k = b$;

Passo 2. Calcule a função para x_k, $f(x_k)$;

Passo 3. Calcule o gradiente $g(x_k)$;

Passo 4. Seja $d_1 = z_1$;

Passo 5. Calcule $\varphi(z)$;

Passo 6. Calcule o tamanho do passo α_k;

Enquanto $\varphi(z) >$ tolerância;

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

Calcule a função para x_{k+1}, $f(x_{k+1})$;

Calcule o gradiente $g(x_{k+1})$;

Avalie o critério de parada $\varphi(z)$;

Calcule $d_{k+1} \in \aleph(A)$, *tal que* $d_{k+1}^T G d_k = 0$;

Calcule α_{k+1};

Atualize o contador $i = i + 1$;

fim

4.6 Algoritmo IV - Gradiente Conjugado Reduzido

Inicialização

A inicialização do algoritmo depende da escolha de um ponto inicial viável x_0, isto é, $Ax_0 = b$.

De posse de x_0, calcula-se $g_0 = Gx_0 - h$.

Escolha das Direções

Encontre todas as direções que são mutuamente G-conjugadas e que estão no núcleo de A. De posse de todas direções, monta-se a matriz Z que é uma base para o núcleo de A.

Faça $\overline{H} = Z^T G Z$ e $\overline{g} = Z^T \nabla f(x_0)$.

Faça $r_0 = \overline{H} x_0 + \overline{g}$.

Faça $d_0 = -r_0$.

Iterações

A cada iteração

1. Calcula-se o novo tamanho de passo: $\alpha_{k+1} = \frac{r_k^T r_k}{d_k^T \overline{H} d_k}$;
2. Calcula-se um novo ponto: $x_{k+1} = x_k + \alpha_k d_k$;
3. Calcula-se o gradiente: g_{k+1}.
4. Calcula-se $r_{k+1} = r_k + \alpha_k \overline{H} d_k$

5. Calcula-se $\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

6. Calcula-se $d_{k+1} = -r_{k+1} + \beta_{k+1} d_k$

Finalização

Se a condição $||r_0|| \leq$ *tolerância* é violada, $x_0 = x_{n-m}$ e retorna para as iterações.

Algoritmo

Algoritmo 4

Passo 1. Encontre x_k *tal que* $Ax_k = b$*;*

Passo 2. Calcule a função para x_k*,* $f(x_k)$*;*

Passo 3. Calcule o gradiente $g(x_k)$*;*

Passo 4. Encontre uma base G-conjugada para $\aleph(A)$*;*

Faça $\overline{H} = Z^T G Z$*;*

Faça $\overline{g} = Z^T \nabla f(x_0)$*;*

Faça $r_0 = \overline{H} x_0 + \overline{g}$*;*

Faça $d_0 = -r_0$*;*

Enquanto $||r_0|| >$ *tolerância;*

Calcule o tamanho do passo $\alpha_{k+1} = \frac{r_k^T r_k}{d_k^T \overline{H} d_k}$

Seja $x_{k+1} = x_k + \alpha_k d_k$*;*

Calcule $r_{k+1} = r_k + \alpha_k \overline{H} d_k$

Calcule $\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

Calcule $d_{k+1} = -r_{k+1} + \beta_{k+1} d_k$

Atualize o contador $i = i + 1$*;*

fim

5 Experimentos Números

Para analisar o comportamento do algoritmo foram efetuados vários experimentos numéricos. Na implementação do algoritmo, os problemas de minimização quadrática com restrições lineares são definidos como:

$$\begin{array}{ll} minimizar & f(x) = \frac{1}{2}x^T Gx + h^T x + c, \\ s.a. & Ax = b \end{array}$$

Na apresentação das tabelas contendo os resultados obtidos se utiliza a seguinte simbologia:

1. *Iterações*: número de iterações feito pelo algoritmo;
2. *Tempo(s)*: tempo de CPU em segundos;
3. $||Z^T \nabla f(x)||$: no caso dos algoritmos de direções conjugadas $norma - 2$ do produto interno entre a base Z formada iterativamente e o vetor gradiente, e o caso dos algoritmos de gradiente conjugado reduzido, Z é uma base para o espaço nulo de A;
4. $f(x^*)$: valor da função objetivo na solução obtida;

6

Conclusões

A abordagem mais promissora encontrada na literatura para resolver o problemas de programação quadrática com restrições lineares de igualdade usa como idéia básica a conjugacidade para reduzir a dimensão do espaço no qual o problema está imerso. Essa abordagem, contudo, não está isenta de de problemas, principalmente nos casos mais relevantes de sistemas de grande porte, esparsos. Como já foi observado, se G, a hessiana da função a ser minimizada, é esparsa, não existe um procedimento geral para escolha de de uma base Z do espaço nulo das restrições de modo que Z^TGZ seja também esparsa [3]. Portanto essa abordagem conduz geralmente a um sistema de grande porte, denso, o que é uma situações desfavorável. A idéia que apresentamos aqui para resolver o mesmo problema tem pelo menos duas vantagens sobre a abordagem descrita acima. Primeiro, não modifica a esparsidade da matriz G, evitando introduzir mal condicionamento no problema, e segundo, não modifica a estrutura dos autovalores da hessiana G quando essa estrutura é favorável. Propositalmente, na implementação do novo algoritmo, a obtenção dos vetores conjugados não fez uso de nenhuma técnica muito apurada, dando oportunidade a que o algoritmo se comparasse favoravelmente com os algoritmos existentes, baseada apenas nos aspectos conceituais. Uma questão em aberto é se as performances do algoritmo melhora com escolhas mais sofisticadas na sua confecção, e nesse caso, o quanto melhora. Outra questão em aberto é o quanto e de que maneira a escolha da base para o espaço nulo das restrições impacta na convergência do algoritmo. De todo modo os resultados numéricos oriundos dos testes são bastante promissores quanto às potencialidades da abordagem apresentada.

Referências Bibliográficas

[1] BORGES, C.L.T., FALCÃO, D.M. e COUTINHO, A.L.G.A.: *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.

[2] COLLEMAN, T.F.: *Linearly constrained optimization and projected preconditioned conjugate gradients,*in Proceedings of the Fifth SIAM Conference on Applied Linear Algebra, SIAM: Philadelphia, pp.118 - 122, 1994.

[3] COLLEMAN, T.F. and VERMA, A.: *A preconditioned conjugate gradient approach to linear equality constrained minimization,* Computer Science Department and Cornell Theory Center, Cornell University, Ithaca, New York, USA, pp. 61 - 72, 2001.

[4] COLLEMAN, T.F., LIU, Jianguo and YUAN, Wei: *A New Trust-Region Algorithm for Equality Constrained Optimization,* Computational Optimization and Applications, pp. 177 - 199, 2002.

[5] CUNHA, Cristina: *Métodos numéricos para engenharias e ciências aplicadas.* Editora da Unicamp, São Paulo - São Paulo, 2 edição, 2000.

[6] FLETCHER, Roger: *Practical methods of optimization,* A Wiley Interscience Publication, Chichester, N.Y.: Wiley, 2nd ed, 1987.

[7] GILBERT, J.R. and HEATH, M.: *Computing a sparse basis for the nullspace,* SIAM J. Alg. & Disc. Meth., vol. 8, pp.446 - 459, 1987.

[8] GILL, Philip E.; MURRAY, Walter and WRIGHT, Margareth H.: *Practical Optimization*, Academic Press, 1981.

[9] GILL, P., MURRAY, D. Ponceleon e SAUNDERS, M.: *Preconditiones for indefinite systems arising in optimization.* SIAMJ. Matrix Anal. Appl., pp. 292-311, 1992.

[10] GOLDBARG, Marco César e GOLDBARG, Elizabeth F.G.: *Transgenética Computacional: Uma Aplicação ao Problema Quadrático de Alocação.* Pesquiso Operacional, v.22, n.3, pp. 359-386, 2002

[11] GOLUB, Gene H. and VAN LOAN, Charles F.: *Matrix computation,* The Johns Hopkins University Press Ltd., London, 3rd ed.,1996.

[12] HESTENES, Magnus: *Applications of Mathematics: Conjugate Direction Methods in Optimization,* Springer-Verlag, New York, 1980.

[13] LUENBERGER, David G.: *Introduction to linear and nonlinear programming.* Addison-Wesley Publishing Company, Massachusetts, 1973.

[14] MARTINEZ, J.M. and SANTOS, S.A. *Métodos Computacionais de Otimização,* XX Colóquio Brasileiro de Matemática, IMPA, 1995.

[15] MORABITO, R. and FARAGO, R.: *A tight Lagrangean relaxation bound for the manufacturer's pallet loading problem,* Studia Informatica Universalis, pp. 57 - 76, 2002.

[16] NASH, S.G. e SOFER, A.: *Linear and Nonlinear Programming.* McGrawHill, Singapure, 1996.

[17] NASH, S.G. e SOFER, A.: *Preconditioning Reduced Matrices.* SIAMJ. Matrix Anal. Appl., pp. 47-68, 1996.

[18] NOCEDAL, J. e WRIGHT, Stephen J.: *Numerical Optimization.* Springer, New York, 1999.

[19] NOCEDAL, J. e WRIGHT, Stephen J. *On the solution of equality constrained quadratic programming problems arising in optimization,* SIAM, volume 23, Issue 4, pp. 1376 - 1395, 2001.

[20] PENNY, J.E.T. e LINDFIELD, G.R.: *Numerical methods usind matlab.* Prendice Hall, Upper Saddle River, 2nd ed., 1999.

[21] TRAFALIS, Theodore B. and COUELLAN, Nicolas P.: *Neural network training via an affine scaling quadratic optimization algorithm,* Volume 9 , Issue 3, pp. 475 - 481, 1996.

[22] SHARMA, D.K., PEER, S.K. and SHARMA, H.P.: *Quadratic Assignment Formulation for the Multifactor Facilities Layout Problem,* From Journal International of Modelling and Simulation, 2005.

[23] STRANG, Gilbert: *Linear Algebra and its applications,* Hartcour Brace, San Diego, 3rd ed., 1988.

[24] WRIGHT, Margaret H., MURRAY, Walter e WALTER, Philip E.: *Numerical linear algebra and optimization.* Addison-Wesley publishing Company, 1991.

A

Tabelas de Dados

Este anexo contém alguns resultados obtidos, utilizando o software MatLab. Utilizou-se para a matriz hessiana a matriz de Hilbert, a matriz Wathen (matriz do MatLab) e a matriz Minij (matriz do MatLab).

	Hessiana G	Matriz A	Vetor h	Vetor b
Tabela 1 - Hilbert	10×10	5×10	10×1	5×1
Tabela 2 - Hilbert	10×10	2×10	10×1	2×1
Tabela 3 - Hilbert	17×17	7×17	17×1	7×1
Tabela 4 - Hilbert	17×17	6×17	17×1	6×1
Tabela 5 - Hilbert	36×36	10×36	36×1	10×1
Tabela 6 - Hilbert	50×50	20×50	50×1	20×1
Tabela 7 - Hilbert	75×75	35×75	75×1	35×1
Tabela 8 - Hilbert	90×90	10×90	90×1	10×1
Tabela 9 - Wathen	8×8	2×8	8×1	2×1
Tabela 10 - Wathen	21×21	5×21	21×1	5×1
Tabela 11 - Wathen	40×40	15×40	40×1	15×1
Tabela 12 - Wathen	65×65	30×65	65×1	30×1
Tabela 13 - Wathen	96×96	20×96	96×1	20×1

Tabela A.1: Dados utilizados nos experimentos

	Hessiana G	Matriz A	Vetor h	Vetor b
Tabela 14 - Minij	5×5	2×5	5×1	2×1
Tabela 15 - Minij	15×15	5×15	15×1	5×1
Tabela 16 - Minij	30×30	7×30	30×1	7×1
Tabela 17 - Minij	50×50	10×50	50×1	10×1
Tabela 18 - Minij	75×75	20×75	75×1	20×1
Tabela 19 - Minij	100×100	25×100	100×1	25×1

Tabela A.2: Dados utilizados nos experimentos

B

Tabelas de Resultados

	Iterações	Tempo	$\|\|Z^T \nabla f(x)\|\|$	$f(x^*)$
Algorithm I	5	0.0300	7.7415e-09	-3.6163
Algorithm II	5	0.0300	9.7275e-09	-3.6163
Algorithm III	5	0.0300	2.1881e-09	-3.6163
Algorithm IV	7	0.0500	2.0807e-09	-3.6163

Tabela B.1: Experimento Numérico 1

	Iterações	Tempo	$\|\|Z^T \nabla f(x)\|\|$	$f(x^*)$
Algorithm I	8	0.0200	8.0787e-09	-9.7915e+06
Algorithm II	8	0.0300	3.5253e-09	-9.7915e+06
Algorithm III	8	0.0310	3.4131e-09	-9.7915e+06
Algorithm IV	28	0.0600	3.5640e-09	-9.7915e+06

Tabela B.2: Experimento Numérico 2

	Iterações	Tempo	$\|\|Z^T\nabla f(x)\|\|$	$f(x^*)$
Algorithm I	10	0.0700	3.5425e-09	-8.1531e+09
Algorithm II	10	0.1100	4.7191e-09	-8.1531e+09
Algorithm III	10	0.0700	9.4223e-09	-8.1531e+09
Algorithm IV	84	0.0700	2.2892e-09	-8.1531e+09

Tabela B.3: Experimento Numérico 3

	Iterações	Tempo	$\|\|Z^T\nabla f(x)\|\|$	$f(x^*)$
Algorithm I	11	0.0900	6.7843e-09	-1.9536
Algorithm II	11	0.1200	1.9314e-09	-1.9536
Algorithm III	11	0.0900	1.7800e-09	-1.9536
Algorithm IV	58	0.1410	9.2509e-09	-1.9536

Tabela B.4: Experimento Numérico 4

	Iterações	Tempo	$\|\|Z^T\nabla f(x)\|\|$	$f(x^*)$
Algorithm I	26	0.5110	2.3975e-09	-12.6047
Algorithm II	26	0.4809	8.7969e-09	-12.6047
Algorithm III	26	0.5110	4.7608e-09	-12.6047
Algorithm IV	31	0.5500	3.0564e-09	-12.6047

Tabela B.5: Experimento Numérico 5

	Iterações	Tempo	$\|\|Z^T\nabla f(x)\|\|$	$f(x^*)$
Algorithm I	30	1.0020	6.8791e-09	1.7781
Algorithm II	30	0.9620	1.9014e-09	1.7781
Algorithm III	30	1.1520	3.7524e-09	1.7781
Algorithm IV	35	1.4600	6.2778e-09	1.7781

Tabela B.6: Experimento Numérico 6

	Iterações	Tempo	$\|\|Z^T \nabla f(x)\|\|$	$f(x^*)$
Algorithm I	40	2.7740	6.4903e-09	2.9193
Algorithm II	40	2.5730	1.7593e-09	2.9193
Algorithm III	40	3.2650	6.4941e-09	2.9193
Algorithm IV	63	4.5100	6.7954e-09	2.9193

Tabela B.7: Experimento Numérico 7

	Iterações	Tempo	$\|\|Z^T \nabla f(x)\|\|$	$f(x^*)$
Algorithm I	80	11.1070	1.9056e-09	-0.5780
Algorithm II	80	30.2950	7.8087e-09	-0.5780
Algorithm III	80	11.0160	1.8506e-09	-0.5780
Algorithm IV	100	15.1110	1.2235e-09	-0.5780

Tabela B.8: Experimento Numérico 8

	Iterações	Tempo	$\|\|Z^T \nabla f(x)\|\|$	$f(x^*)$
Algorithm I	6	0.0160	4.5461e-09	1.1270
Algorithm II	6	0.0160	3.6515e-09	1.1270
Algorithm III	6	1.6630	1.3422e-09	1.1270
Algorithm IV	6	0.0320	1.6746e-09	1.1270

Tabela B.9: Experimento Numérico 9

	Iterações	Tempo	$\|\|Z^T \nabla f(x)\|\|$	$f(x^*)$
Algorithm I	16	0.0470	1.0650e-09	0.5505
Algorithm II	16	0.0620	7.5269e-09	0.5505
Algorithm III	16	0.0620	2.8211e-09	0.5505
Algorithm IV	19	0.0470	7.1355e-09	0.5505

Tabela B.10: Experimento Numérico 10

	Iterações	Tempo	$\|\|Z^T \nabla f(x)\|\|$	$f(x^*)$
Algorithm I	25	0.2660	7.6475e-09	1.3322
Algorithm II	25	0.2820	2.2826e-09	1.3322
Algorithm III	25	0.2970	7.1521e-09	1.3322
Algorithm IV	28	0.3070	3.9585e-09	1.3322

Tabela B.11: Experimento Numérico 11

	Iterações	Tempo	$\|\|Z^T \nabla f(x)\|\|$	$f(x^*)$
Algorithm I	35	0.9060	8.0157e-09	1.7917
Algorithm II	35	0.9370	5.4822e-09	1.7917
Algorithm III	35	0.9220	1.0244e-09	1.7917
Algorithm IV	39	0.9950	6.6381e-09	1.7917

Tabela B.12: Experimento Numérico 12

	Iterações	Tempo	$\|\|Z^T \nabla f(x)\|\|$	$f(x^*)$
Algorithm I	76	7.2350	1.6274e-09	0.2841
Algorithm II	76	6.8750	3.3310e-09	0.2841
Algorithm III	76	6.8130	2.4117e-09	0.2841
Algorithm IV	83	7.7630	3.6830e-09	0.2841

Tabela B.13: Experimento Numérico 13

	Iterações	Tempo	$\|\|Z^T\nabla f(x)\|\|$	$f(x^*)$
Algorithm I	3	0.0150	4.3566e-09	-0.7750
Algorithm II	3	0.0100	2.7371e-09	-0.7750
Algorithm III	3	0.0160	1.9665e-09	-0.7750
Algorithm IV	3	0.0310	3.1830e-09	-0.7750

Tabela B.14: Experimento Numérico 14

	Iterações	Tempo	$\|\|Z^T\nabla f(x)\|\|$	$f(x^*)$
Algorithm I	10	0.0310	7.0857e-09	-1.0428
Algorithm II	10	0.0310	1.1311e-09	-1.0428
Algorithm III	10	0.0310	1.3609e-09	-1.0428
Algorithm IV	11	0.0320	1.8257e-09	-1.0428

Tabela B.15: Experimento Numérico 15

	Iterações	Tempo	$\|\|Z^T\nabla f(x)\|\|$	$f(x^*)$
Algorithm I	23	0.1250	3.5183e-09	-2.3556
Algorithm II	23	0.1410	8.7309e-09	-2.3556
Algorithm III	23	0.1125	1.6371e-09	-2.3556
Algorithm IV	24	0.1570	3.2317e-09	-2.3556

Tabela B.16: Experimento Numérico 16

	Iterações	Tempo	$\|\|Z^T\nabla f(x)\|\|$	$f(x^*)$
Algorithm I	40	0.6560	1.7272e-09	-3.5049
Algorithm II	40	0.6250	6.4955e-09	-3.5049
Algorithm III	40	0.5940	9.2087e-09	-3.5049
Algorithm IV	47	0.7260	7.1166e-09	-3.5049

Tabela B.17: Experimento Numérico 17

	Iterações	Tempo	$\|\|Z^T\nabla f(x)\|\|$	$f(x^*)$
Algorithm I	55	2.2500	1.3735e-09	-3.5625
Algorithm II	55	2.3590	2.5106e-09	-3.5625
Algorithm III	55	2.1250	2.1092e-09	-3.5625
Algorithm IV	67	2.9470	8.1996e-09	-3.5625

Tabela B.18: Experimento Numérico 18

	Iterações	Tempo	$\|\|Z^T\nabla f(x)\|\|$	$f(x^*)$
Algorithm I	75	7.9060	0.5224e-09	-6.1546
Algorithm II	75	7.5930	0.1086e-09	-6.1546
Algorithm III	75	6.7820	0.1367e-09	-6.1546
Algorithm IV	85	8.3620	0.5672e-09	-6.1546

Tabela B.19: Experimento Numérico 19