

DÉBORA CINTIA MARCILIO

**LAGRANGEANO AUMENTADO APLICADO NA RESOLUÇÃO DE
SUBPROBLEMAS GERADOS PELO MÉTODO DE
PROGRAMAÇÃO QUADRÁTICA SEQÜENCIAL**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre, pelo Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Setores de Ciências Exatas e Tecnologia, Universidade Federal do Paraná.

Orientador:

Prof. Dr. Luiz Carlos Matioli.

Curitiba

2006

TERMO DE APROVAÇÃO

*Eu quase de nada sei,
mas desconfio de muita coisa.*

João Guimarães Rosa

Aos meus filhos.

Agradecimentos

A Deus.

Sem o apoio de algumas pessoas seria impossível a conclusão desse trabalho, a essas pessoas meus sinceros agradecimentos:

Ao professor Matioli, pela orientação durante o meu aprendizado. A nossa caminhada como orientador e orientando começou em 2002, já realizamos alguns trabalhos juntos e com certeza a dedicação e a paciência do professor Matioli foram essenciais para a conclusão de todos.

Aos professores Yuan Jin Yun e Ricardo Biloti, pela leitura cuidadosa da dissertação, participação na banca examinadora e sugestões as quais enriqueceram nossa proposta.

Aos meus lindos, Gabriel e Daniel, pelo incentivo e estímulo para concluir este trabalho e por me fazer a mulher mais amada do mundo.

Aos meus pais e aos meus irmãos, por estarem sempre presentes na minha vida me apoiando em todos os momentos.

Ana Paula obrigada pela amizade que você me devota, pela presença em todos os momentos, por ficar triste quando estou triste e por rir comigo quando estou feliz, por guardar os meus segredos e principalmente obrigado por me achar merecedora de guardar os teus segredos.

À todos do CESEC, os amigos das intermináveis horas de estudos e os amigos da cozinha também e principalmente obrigado Maris.

Resumo

Estamos propondo a aplicação do algoritmo de Lagrangeano Aumentado com penalidade quadrática, em problemas de programação quadrática convexos. Os problemas de programação quadrática são compostos de função objetivo quadrática e restrições lineares (no nosso caso essas restrições serão de desigualdade). Essa importante classe de problemas, será gerada através do algoritmo de programação quadrática seqüencial, onde a cada iteração o problema quadrático é formado fazendo-se, no ponto atual, uma aproximação quadrática da função Lagrangeana associada ao problema original e uma aproximação linear das restrições. Em seguida utilizamos penalização para resolver esses subproblemas usando o algoritmo de Lagrangeano Aumentado com penalidade quadrática, o que resulta em uma seqüência de problemas quadráticos irrestritos, que comparado com o problema original pode ser considerado de mais fácil solução. Através dessa nova metodologia mostramos que, asseguradas algumas hipóteses, se a função Lagrangeano associada ao problema original é estritamente convexa (convexa), então a matriz hessiana da função Lagrangeano Aumentado é definida positiva (semidefinida positiva), logo satisfaz a condição de otimalidade suficiente de segunda ordem.

Abstract

We are considering the application of the Augmented Lagrangian algorithms with quadratic penalty, to convex problems of quadratic programming. The problems of quadratic programming are composites of quadratic objective function and linear constraints (in our case, only inequality constraints will be taken into account inequality). This important class of problems will be generated through the algorithm of sequential quadratic programming, where at each iteration the quadratic problem is formed by, in the current point, a quadratic approach of the Lagrangian function associated with the original problem and the linearization of approach of the constraints. After that we use penalization to solve these subproblems using the Augmented Lagrangian algorithms with quadratic penalty, what results in a sequence of unconstrained quadratic problems, that compared with the original problem can be considered of easier solution. Through this new methodology we show that, under some hypotheses, if the Lagrangian function associated with the original problem is strict convex (convex), then the hessian matrix of Augmented Lagrangian function is defined positive (semidefined positive), then satisfies the sufficient condition optimality of second order.

Sumário

Introdução	1
1 Conceitos Fundamentais	4
1.1 Matriz e Vetor	4
1.2 Funções	7
1.3 Convexidade	8
1.4 Condições de Otimalidade para Problemas de Minimização Restrita e Irrestrita	9
1.4.1 Minimização Irrestrita	9
1.4.2 Minimização Restrita	13
2 Penalidade e Lagrangeano Aumentado	15
2.1 Funções de Penalidades Coercivas Pela Direita	16
2.2 Construindo Penalidades da Família \mathcal{P}	16
2.3 Métodos de Lagrangeano Aumentado	18
3 Programação Quadrática	21
3.1 Programação Quadrática com Restrições de Igualdade	22
3.1.1 Resolvendo o Sistema KKT	25
3.2 Programação Quadrática com Restrições de Igualdade e Desigualdade	27
3.3 Métodos para Resolver o Problema (PQ)	29
3.4 Lagrangeano Aumentado com Penalidade Quadrática	31
4 Programação Quadrática Seqüencial	37
4.1 Programação Quadrática Seqüencial Pura	37
4.2 Estrutura de Programação Quadrática Seqüencial	39
4.2.1 Região de Confiança	42
4.3 Programação Quadrática Seqüencial com Região de Confiança	46

5	Implementação e Testes Numéricos	50
5.1	Implementação	50
5.2	Resultados Numéricos	53
	Conclusão	59
	Referências Bibliográficas	60

Lista de Figuras

3.1	G indefinida	28
3.2	G definida negativa	29
3.3	Função Quadrática Clássica	31
3.4	Derivada	31
3.5	Função Quadrática	32
4.1	Um passo de Região de Confiança	45
4.2	Passo do método Região de Confiança	49

Lista de Tabelas

5.1	Problemas da coleção CUTE	53
5.2	Comparação entre as penalidades	55
5.3	Resumo de Iterações	57
5.4	Aumento do Parâmetro de Penalidade	57
5.5	Resumo de Iterações em Problemas Não Convexos	58

Introdução

Segundo Martínez [11], otimização é um problema matemático com muitas aplicações “no mundo real”. Consiste em encontrar os mínimos e os máximos de uma função de várias variáveis, definida numa determinada região do espaço multidimensional. A linguagem utilizada pela otimização para expressar os problemas é conhecida como programação matemática, que trata do estudo de problemas de otimização e o desenvolvimento de métodos para resolvê-los.

Cada vez mais o processo de otimização tem sido utilizado em problemas de diversas áreas de engenharia, economia e indústrias, onde os responsáveis pela tomada de decisões buscam na otimização ferramentas para maximizar lucros e minimizar custos.

O problema geral de otimização (sem restrição no número de variáveis) é expresso em programação matemática como:

$$(PR) \quad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & h_i(x) = 0 \quad i = 1, 2, \dots, p \\ & g_i(x) \leq 0 \quad i = 1, 2, \dots, m \\ & x \in \mathbb{R}^n \end{array}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Caso o problema seja de *maximizar* f tem-se que *maximizar* $f(x)$ é equivalente a *minimizar* $-f(x)$.

A função f é chamada de função objetivo e as funções h e g são as restrições do problema, o conjunto formado pelas restrições é chamado de conjunto viável. Dependendo das características da função objetivo, das restrições e do conjunto viável, têm-se os diferentes problemas de otimização.

Discutiremos nesse trabalho o caso onde as funções são, normalmente, não-lineares e continuamente diferenciáveis, ou seja, estudaremos o problema geral de programação não-linear diferenciável. Mais especificamente, estudaremos o caso onde a função objetivo é quadrática e as restrições são lineares, ou seja, problemas de Programação Quadrática.

Problemas de otimização com as características dos problemas de Programação Quadrática são muito importantes, pois eles surgem na formulação matemática de várias situações práticas ou ainda como subproblemas gerados por métodos de programação não-linear como é o caso de Programação Quadrática Sequencial.

Os métodos baseados em Programação Quadrática Sequencial são bastante utilizados para resolver problemas de programação não-linear. São métodos que consistem no tratamento de problemas como (PR), descrito acima, através de uma seqüência de problemas mais simples da seguinte forma: gera-se um processo iterativo que em cada iteração, o problema (PR) é transformado em um problema quadrático. O problema quadrático é formado fazendo-se, no ponto atual, uma aproximação quadrática da função Lagrangeana associada ao problema (PR) e uma aproximação linear das restrições.

Outra metodologia bastante utilizada para resolver problemas de programação não-linear são os métodos de Lagrangeano Aumentado que são baseados nos métodos de penalização, é uma classe de métodos robustos e aplicados a problemas de pequeno e grande porte. Originalmente, foram aplicados em problemas com restrições de igualdade e posteriormente generalizados para problemas com restrições de desigualdade, neste trabalho, será aplicado em problemas com restrições de desigualdade.

A metodologia dos métodos de Lagrangeano Aumentado ou, mais geral, dos métodos de penalização é bastante simples. Consiste em transformar o problema restrito em uma seqüência de subproblemas irrestritos, formados pela função objetivo do problema original acrescida a um múltiplo das restrições. São iterativos e a cada iteração resolvem o subproblema irrestrito gerado e atualizam os multiplicadores e o parâmetro de penalização.

O foco central de nosso trabalho é a aplicação do método Lagrangeano Aumentado usando a função penalidade $p \in \mathcal{P}$ **tipo 1**

$$(y, \mu) \in \mathbb{R} \times \mathbb{R}_{++} \mapsto p(y, \mu) = \theta(\mu y)$$

e a função θ quadrática

$$y \in \mathbb{R} \mapsto \theta(y) = \frac{1}{2}y^2 + y,$$

introduzida em [12] e [10], em problemas de Programação Quadrática.

Para gerar os subproblemas quadráticos, onde será aplicado o algoritmo de Lagrangeano Aumentado com penalidade quadrática, usamos o algoritmo de programação quadrática sequencial.

Logo estamos propondo um algoritmo iterativo, onde a cada iteração primeiramente transformamos um problema de programação não-linear com restrições de

desigualdade em um problema de programação quadrática com restrições lineares de desigualdade, que em seguida será penalizado através do método de Lagrangeano Aumentado com penalidade quadrática tornando-se um problema quadrático irrestrito. Que comparado com o problema original pode ser considerado de mais fácil solução.

Descrição por Capítulos

O trabalho está dividido em cinco capítulos, a seguir descreveremos o que será tratado em cada capítulo da dissertação.

No capítulo 1, fazemos uma revisão de alguns conceitos, que mais frequentemente aparecem ao longo do trabalho. Ainda neste primeiro capítulo apresentamos a minimização restrita e irrestrita, onde fazemos uma revisão sobre as condições de otimalidade referentes a esses problemas. No capítulo 2, descrevemos o método Lagrangeano Aumentado, definindo funções penalidades e indicando uma metodologia para a construção dessas.

No capítulo 3, apresentamos o resultado mais relevante deste trabalho. Primeiramente, apresentamos o problema quadrático e discutimos metodologias para a resolução do mesmo. Dentre essas metodologias propomos a utilização do método Lagrangeano Aumentado com penalidade quadrática para a resolução de problemas quadráticos, através dessa nova metodologia mostramos que, assegurada algumas hipóteses, se a função Lagrangeano associado ao problema original é estritamente convexa (convexa), então a matriz hessiana da função Lagrangeano Aumentado é definida positiva (semidefinida positiva), logo satisfaz a condição de otimalidade suficiente de segunda ordem.

No capítulo 4 discutimos a metodologia da Programação Quadrática Sequencial, a qual é usada nesse trabalho como ferramenta na geração de problemas quadráticos, nos quais serão aplicados o método que estamos propondo. No capítulo 5 apresentamos questões referentes à implementação dos algoritmos de Lagrangeano Aumentado e Programação Quadrática Sequencial. Neste capítulo apresentamos também, os testes numéricos realizados com esses algoritmos.

Capítulo 1

Conceitos Fundamentais

Neste capítulo, serão revistos alguns conceitos básicos e definições que serão necessários para o desenvolvimento dos demais capítulos dessa dissertação.

Começamos introduzindo sistemas de equações lineares, matrizes, vetores, gradientes e matrizes hessianas. Revisamos algumas propriedades de matrizes, bem como funções convexas, sendo estas muito importantes em otimização.

Este capítulo contém os problemas de minimização restrita e irrestrita, onde será feita uma revisão sobre as condições de otimalidade referentes a esses problemas. Veremos brevemente alguns métodos para resolver problemas de minimização irrestrita e por fim, a condição de qualificação (LICQ) e a condição de otimalidade Karush-Kuhn-Tucker (KKT).

1.1 Matriz e Vetor

Iniciaremos definindo uma **Equação Linear** de n variáveis x_1, x_2, \dots, x_n sendo uma expressão que pode ser escrita da seguinte forma:

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = b,$$

onde a_1, a_2, \dots, a_n e b são constantes reais. As variáveis x_j , $j = 1, 2, \dots, n$ na equação linear são chamadas de incógnitas.

Um conjunto finito de equações lineares nessas variáveis é chamado de **Sistema Linear**. E, dado um sistema linear arbitrário com m equações e n incógnitas podemos escrevê-lo da seguinte maneira:

$$\begin{aligned}
 a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
 a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
 \vdots & \\
 a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
 \end{aligned}
 \tag{1.1}$$

sendo $x_j, j = 1, 2, \dots, n$ as incógnitas e a_{ij} e $b_j, i = 1, 2, \dots, m, j = 1, 2, \dots, n$ constantes reais.

Matriz é um conjunto de números reais ou complexos dispostos em linhas e colunas. O tamanho de uma matriz pode ser descrito pelo número de linhas e colunas que ela contém. Geralmente as matrizes são denotadas por letras maiúsculas. Uma matriz em $\mathbb{R}^{m \times n}$ será indicada por

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

onde m é o número de linhas e n o número de colunas da mesma. Neste trabalho será utilizada somente matriz real.

$A^t = (a_{ji})$ é a matriz Transposta de A , e a matriz Inversa de A , neste caso $m=n$, é denotada por A^{-1} .

Uma matriz que contém somente uma coluna é chamada de vetor coluna, assim como a matriz que contém somente uma linha é chamada de vetor linha. Nessa dissertação trabalharemos somente com vetores colunas, denotados por letras minúsculas.

Um vetor $x \in \mathbb{R}^n$ será indicado por

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

onde $x_j \in \mathbb{R}, j = 1, 2, \dots, n$ é a j -ésima componente do vetor x .

O vetor $x^t = [x_1, x_2, \dots, x_n]$ é o vetor transposto de x dado na relação acima.

Considere o sistema de m equações lineares e n incógnitas dado na relação (1.1), pode-se reescrever as m equações desse sistema por uma única equação matricial:

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

As constantes reais a_{ij} , $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$ compõem a matriz A e b_j , $j = 1, 2, \dots, n$ o vetor b , assim pode-se escrever essas equações lineares matricialmente, ou seja, $Ax = b$.

A seguir serão apresentadas algumas notações e definições relacionadas a matrizes e vetores que serão utilizadas ao longo do texto.

O produto escalar entre dois vetores x e y em \mathbb{R}^n é definido por

$$\langle x, y \rangle = \sum_{i=1}^n x_i \cdot y_i = x^t \cdot y \quad (1.2)$$

A norma Euclidiana de um vetor $x \in \mathbb{R}^n$ é definida por

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^t \cdot x} \quad (1.3)$$

Uma matriz quadrada é dita Simétrica se $A = A^t$, ou seja, $a_{ij} = a_{ji}$ para todo i e j .

Definição 1.1 *Seja A uma matriz real simétrica. Dizemos que A semidefinida positiva se*

$$x^t Ax \geq 0, \quad \forall x \in \mathbb{R}^n. \quad (1.4)$$

A é dita definida positiva se

$$x^t Ax > 0, \quad \forall x \in \mathbb{R}^n \text{ tal que } x \neq 0. \quad (1.5)$$

A é dita semidefinida negativa se

$$x^t Ax \leq 0, \quad \forall x \in \mathbb{R}^n. \quad (1.6)$$

A é dita definida negativa se

$$x^t Ax < 0, \quad \forall x \in \mathbb{R}^n \text{ tal que } x \neq 0. \quad (1.7)$$

Finalmente, A é dita indefinida se

$$x^t Ax \text{ pode assumir valores negativos e positivos.} \quad (1.8)$$

1.2 Funções

As funções com as quais trabalharemos serão definidas no \mathbb{R}^n com valores em \mathbb{R} , isto é, $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Definição 1.2 Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $x \in \text{dom}f$. A derivada direcional de f em x na direção $d \in \mathbb{R}^n$ é definida por

$$f'(x, d) = \lim_{t \rightarrow 0} \frac{f(x + td) - f(x)}{t}. \quad (1.9)$$

O vetor das derivadas parciais de f em um ponto x , é denominado de vetor **Gradiente**, e indicado por

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_i} \right] = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}. \quad (1.10)$$

A matriz das derivadas parciais de segunda ordem de f em um ponto x , é denominada de **matriz Hessiana**, e indicada por

$$\nabla^2 f(x) = \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right] = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{bmatrix}. \quad (1.11)$$

O conjunto de várias funções a valores reais g_1, g_2, \dots, g_m em \mathbb{R}^n , pode ser visto como uma função vetorial $g(\cdot)$. Essa função determina o vetor g definido por $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, ou seja,

$$g(x)^t = \left[g_1(x), g_2(x), \dots, g_m(x) \right]. \quad (1.12)$$

Supondo que g seja contínua com derivadas parciais contínuas. A matriz J das derivadas parciais de g denomina-se **matriz Jacobiana** e será indicada por

$$\nabla g(x) = \left[\frac{\partial g_i(x)}{\partial x_j} \right] = \begin{bmatrix} \frac{\partial g_1(x)}{\partial x_1} & \frac{\partial g_1(x)}{\partial x_2} & \dots & \frac{\partial g_1(x)}{\partial x_n} \\ \frac{\partial g_2(x)}{\partial x_1} & \frac{\partial g_2(x)}{\partial x_2} & \dots & \frac{\partial g_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m(x)}{\partial x_1} & \frac{\partial g_m(x)}{\partial x_2} & \dots & \frac{\partial g_m(x)}{\partial x_n} \end{bmatrix}. \quad (1.13)$$

1.3 Convexidade

A convexidade de um problema é muito importante em otimização, pois está relacionada com o conceito de mínimo global. O termo convexo pode ser usado tanto para conjunto quanto para funções. Estamos nos baseando nos livros [8] e [5].

Definição 1.3 *O conjunto $S \subset \mathbb{R}^n$ é chamado convexo se para todo $x_1, x_2 \in S$ e todo $\alpha \in [0, 1]$, tem-se $\alpha x_1 + (1 - \alpha)x_2 \in S$.*

Geometricamente, está definição quer dizer que se escolhermos dois pontos quaisquer em S e os unirmos por um segmento de reta, todos os pontos desse segmento devem estar em S , para que S seja convexo. Caso contrário, S será não convexo.

Definição 1.4 *Considere S um subconjunto não vazio e convexo do \mathbb{R}^n . A função $f : S \rightarrow \mathbb{R}$ é dita convexa sobre S , se para todo $x_1, x_2 \in S$ e todo $\alpha \in [0, 1]$ se verifica*

$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2). \quad (1.14)$$

Se a desigualdade em (1.14) é estrita a função é chamada de **estritamente convexa**.

Teorema 1.5 *Seja $S_j, j = 1, \dots, n$, conjuntos convexos. Então o conjunto*

$$S = \bigcap_{j=1}^n S_j \quad (1.15)$$

é convexo.

Prova. Ver [6] ■

Teorema 1.6 *Se $f \in C^1$ então (i) e (ii) abaixo são equivalentes, e se $f \in C^2$ então (i), (ii) e (iii) abaixo são equivalentes.*

(i) f é convexa;

(ii) $\forall x, y \in \mathbb{R}^n : f(y) \geq f(x) + \nabla f^t(x) \cdot (y - x)$;

(iii) $\nabla^2 f(x) \geq 0, \forall x \in \mathbb{R}^n$.

1.4 Condições de Otimalidade para Problemas de Minimização Restrita e Irrestrita

Nesta seção apresentamos os problemas de minimização restrita e irrestrita e as condições de otimalidade referentes a esses problemas. As condições de otimalidade são relações entre as derivadas da função objetivo e as derivadas das funções que definem as restrições. As condições necessárias devem ser obrigatoriamente satisfeitas por minimizadores, enquanto as condições suficientes, quando satisfeitas, asseguram que o ponto em questão é um minimizador local ou global dependendo de o problema ser ou não convexo.

Começamos com a minimização irrestrita, que é um caso particular do problema geral de programação não linear, veremos os tipos de minimizadores e as condições de otimalidade para esse tipo de problema. A seguir, apresentamos a minimização restrita, veremos que se o problema satisfaz uma condição de qualificação, há uma extensão ao caso irrestrito.

1.4.1 Minimização Irrestrita

A formulação matemática de um problema de minimização irrestrita é a seguinte:

$$(PI) \quad \begin{array}{ll} \text{minimizar} & f(x) \\ & x \in \mathbb{R}^n \end{array}$$

sendo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continuamente diferenciável. E f é chamada de função objetivo.

No processo de minimização irrestrita, pode-se obter como solução vários tipos de minimizadores, a seguir veremos como podemos reconhecê-los. Lembrando que todas as definições a seguir são análogas para o processo de maximização, ou seja, maximizar f é equivalente a minimizar $-f$.

Definição 1.7 *Um ponto $x^* \in \mathbb{R}^n$ é um minimizador local de f se, e somente se, existe $\varepsilon > 0$ tal que $f(x) \geq f(x^*)$ para todo $x \in \mathbb{R}^n$ tal que $\|x - x^*\| < \varepsilon$.*

Definição 1.8 Um ponto $x^* \in \mathbb{R}^n$ é um minimizador local estrito de f se, e somente se, existe $\varepsilon > 0$ tal que $f(x) > f(x^*)$ para todo $x \in \mathbb{R}^n$ tal que $x \neq x^*$ e $\|x - x^*\| < \varepsilon$.

Definição 1.9 Um ponto $x^* \in \mathbb{R}^n$ é um minimizador global de f se, e somente se, $f(x) \geq f(x^*)$ para todo $x \in \mathbb{R}^n$.

Definição 1.10 Um ponto $x^* \in \mathbb{R}^n$ é um minimizador global estrito de f se, e somente se, $f(x) > f(x^*)$ para todo $x \in \mathbb{R}^n$ tal que $x \neq x^*$.

A partir das definições, dadas acima, pode parecer que a única forma de achar um minimizador é analisando todos os pontos de sua vizinhança, certificando-se que nenhum deles tem um valor de função objetivo menor. Entretanto, quando a função f é suave, existe maneiras mais práticas e eficientes de identificar minimizadores. Em particular, se f é continuamente diferenciável até segunda ordem, pode-se dizer se um ponto x^* é um minimizador local, examinando o vetor gradiente $\nabla f(x^*)$ e a matriz hessiana $\nabla^2 f(x^*)$ no ponto.

A seguir será apresentado um resultado baseado na série de Taylor, a principal ferramenta utilizada para o estudo de minimização de funções.

Teorema 1.11 Suponha que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tem derivadas contínuas pelo menos até segunda ordem, e que $p \in \mathbb{R}^n$. Então tem-se que

$$f(x + p) = f(x) + \nabla f(x + tp)^t p + \frac{1}{2} p^t \nabla^2 f(x + tp) p \quad (1.16)$$

para algum $t \in (0, 1)$.

Prova. Ver [15] ■

Teorema 1.12 (Condição Necessária de Primeira Ordem) Se x^* é um minimizador local e f é continuamente diferenciável numa vizinhança de x^* , então $\nabla f(x^*) = 0$.

Prova. Ver [15] ■

Se $\nabla f(x^*) = 0$, x^* é chamado de ponto estacionário da função f , podendo ser um ponto de máximo, mínimo ou de sela.

Para os teoremas seguintes serão usadas as definições de matriz semidefinida positiva (1.4) e definida positiva (1.5).

Teorema 1.13 (Condição Necessária de Segunda Ordem) Se x^* é um minimizador local de f e $\nabla^2 f$ é contínua numa vizinhança de x^* , então $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ é semi-definida positiva.

Prova. Ver [15] ■

Teorema 1.14 (*Condição Suficiente de Segunda Ordem*) *Suponha que $\nabla^2 f$ é contínua numa vizinhança de x^* e que $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ é positiva definida. Então x^* é um minimizador local estrito de f .*

Prova. Ver [15] ■

Teorema 1.15 *Quando f é convexa, um minimizador local x^* é um minimizador global de f . Se f é diferencialmente contínua e convexa, então um ponto estacionário x^* é um minimizador global de f .*

Prova. Ver [15] ■

As condições necessárias de otimalidade devem ser obrigatoriamente satisfeitas por minimizadores, enquanto as condições suficientes, quando satisfeitas, asseguram que o ponto em consideração é um minimizador local. A garantia de que o ponto é um minimizador global, em geral, é muito mais complicada.

Os algoritmos utilizados para encontrar a solução do problema (PI), quase sempre iterativos, iniciam com um ponto dado x^0 e determinam uma direção d^k . Escolhe-se direções ao longo das quais é possível fazer decrescer o valor da função objetivo, portanto espera-se que o ponto encontrado seja um minimizador local. Assim as direções $d \in \mathbb{R}^n$ tais que $\nabla f(x)^t d < 0$, são chamadas de **direções de descida** a partir do ponto x , ou seja, as direções que formam um ângulo maior que 90 graus com o vetor gradiente da função objetivo.

Ao longo dessa direção, calcula-se um **comprimento do passo** λ^k que permita uma redução razoável no valor da função objetivo. Logo, resolve-se um subproblema para encontrar o comprimento do passo:

Dada uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$, um ponto x e uma direção de descida $d \in \mathbb{R}^n$, encontrar $\lambda \geq 0$ (se existir) tal que

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

Geralmente o critério de parada desses algoritmos é a condição de otimalidade de primeira ordem, isto é, o algoritmo para quando um ponto estacionário da função objetivo é encontrado.

Os algoritmos geram, portanto, uma seqüência (x^k) em \mathbb{R}^n com uma seqüência associada $(f(x^k))$ de valores reais monótona decrescente [11]. O processo a seguir é um modelo geral de algoritmo para resolver o problema (PI):

Algoritmo 1.1 (*Minimização Irrestrita*)

Dado $x^0 \in \mathbb{R}^n$

Para $k=0$

Enquanto $\nabla f(x^k) \neq 0$

Escolha uma direção $d^k \in \mathbb{R}^n$ tal que $\nabla f(x^k)^t d^k < 0$;

Calcule o comprimento do passo $\lambda^k > 0$ tal que

$$f(x^k + \lambda^k d^k) < f(x^k);$$

$$x^{k+1} = x^k + \lambda^k d^k$$

$$k = k + 1$$

Existem na literatura vários métodos para a escolha da direção d^k a ser tomada, assim como também existem várias metodologias para a determinação do comprimento do passo λ^k .

Como, por exemplo, para a escolha da direção tem-se o método de Cauchy que usa o oposto do gradiente da função objetivo como direção:

$$d^k = -\nabla f(x^k). \tag{1.17}$$

Esse é um dos métodos mais conhecidos e também um dos mais antigos, que resolve para o problema irrestrito.

Outra escolha usual da direção d^k é a direção de Newton:

$$d^k = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k), \tag{1.18}$$

como veremos, mais adiante, esta iteração está bem definida se a matriz hessiana da função f for não singular.

Para a determinação do comprimento do passo λ^k , existem vários métodos, denominados métodos de busca unidirecional, os quais garantem que o decréscimo no valor da função seja proporcional ao comprimento do passo na direção d^k . Para exemplificar podemos citar dois métodos de busca unidirecional: busca exata e a busca de Armijo.

Uma discussão mais detalhada, sobre esse assunto, pode ser encontrada em [11] e [7].

1.4.2 Minimização Restrita

O problema geral de programação não linear diferenciável com restrições de igualdade e desigualdade, é:

$$(PR) \quad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & h_i(x) = 0 \quad i \in E \\ & g_i(x) \leq 0 \quad i \in I \\ & x \in \mathbb{R}^n \end{array}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, e $E = 1, 2, \dots, p$ e $I = 1, 2, \dots, m$ são dois conjuntos finitos de índices, sendo que o conjunto E corresponde as restrições de igualdade e o conjunto I corresponde as restrições de desigualdade.

Se o problema satisfaz uma ‘‘Condição de Qualificação’’ há uma extensão ao caso geral da condição necessária de otimalidade no caso irrestrito visto anteriormente. Esta extensão consiste na condição de otimalidade Karush-Kuhn-Tucker. Se a condição de qualificação não for satisfeita não podemos garantir que as condições de otimalidade são realmente necessárias [7].

Definição 1.16 *Conjunto viável Ω é o conjunto de pontos x que satisfazem as restrições de igualdade e desigualdade, ou seja,*

$$\Omega = \{x \mid h_i(x) = 0, i \in E; g_i(x) \leq 0, i \in I\}, \quad (1.19)$$

Define-se, a função Lagrangeano para o problema (PR) como:

$$l(x, \lambda, \mu) = f(x) + \lambda^t h(x) + \mu^t g(x)$$

onde, λ e μ são os multiplicadores de Lagrange associados às restrições de igualdade e desigualdades respectivamente.

Definição 1.17 *Conjunto ativo $\mathbf{A}(x)$ para algum ponto viável x é a união dos índices referentes as restrições de igualdade e os índices referentes as restrições de desigualdade que estão satisfeitas na igualdade no ponto x , isto é,*

$$\mathbf{A}(x) = E \cup \{i \in I \mid g_i(x) = 0\}$$

Definição 1.18 (LICQ) - *Dado o ponto x^* e o conjunto ativo $\mathbf{A}(x^*)$ definido em (1.17), dizemos que a condição de qualificação de independência linear é satisfeita se os gradientes das restrições, pertencentes ao conjunto ativo, são linearmente independentes.*

Teorema 1.19 *KKT* - Suponha que x^* é solução local de (PR) e que satisfaz a condição de qualificação (LICQ). Então há vetores de multiplicadores de Lagrange λ^* e μ^* , com componentes λ_i^* , $i \in E$ e μ_i^* , $i \in I$, tal que as seguintes condições são satisfeitas

$$\begin{aligned}\nabla_x l(x^*, \lambda^*) &= 0, \\ h_i(x^*) &= 0 \quad \forall i \in E, \\ g_i(x^*) &\leq 0 \quad \forall i \in I, \\ \mu_i^* &\geq 0 \quad \forall i \in I, \\ \mu_i^* g_i(x^*) &= 0 \quad \forall i \in I.\end{aligned}\tag{1.20}$$

Prova. Ver [1]

■

Capítulo 2

Penalidade e Lagrangeano Aumentado

Neste capítulo será abordada uma importante classe de métodos para programação não-linear que são os métodos baseados em penalidades o princípio básico desses métodos é converter um problema relativamente difícil em um problema de mais fácil solução, utilizando procedimentos que resolvem problemas restritos através de uma seqüência de problemas irrestritos.

Os algoritmos determinados por esses métodos são iterativos e cada iteração consiste em resolver subproblemas formados pela função objetivo do problema original acrescida a um múltiplo das restrições, e atualização do parâmetro de penalidade.

Iniciamos apresentando funções de penalidade, definindo-as, e indicando uma metodologia para a construção dessas. As funções de penalidades serão utilizadas no método de Lagrangeano Aumentado.

O método de Lagrangeano Aumentado na sua forma original foi introduzido para resolver problemas com restrições de igualdade e posteriormente generalizados para problemas com restrições de desigualdade. Neste trabalho, iremos aplicá-los a problemas com restrições de desigualdade.

Lagrangeano Aumentado tem como objetivo reverter o mal condicionamento proveniente da atualização do parâmetro de penalidade, nos métodos baseados em penalidades clássicas, e a perda de estrutura de minimização quando as restrições são acrescidas na função objetivo.

2.1 Funções de Penalidades Coercivas Pela Direita

Definimos uma família de funções de penalidades que são coercivas à direita e posteriormente serão necessárias para definir o algoritmo de Lagrangeano Aumentado. As penalidades desta família, pelas características da própria definição, são mais indicadas para problemas com restrição de desigualdade.

Usamos $p'(y, \mu)$ para denotar a derivada da função p em relação á primeira variável.

Definição 2.1 *Definimos a família \mathcal{P} de funções de penalidades*

$$y \in \mathbb{R}, \mu \in \mathbb{R}_{++} \mapsto p(y, \mu) \in \mathbb{R}$$

com $\text{dom } p(\cdot, \mu) = (-\infty, b)$, $b > 0$, possivelmente $+\infty$ e p satisfazendo as seguintes propriedades: Para todo $\mu \in \mathbb{R}_{++}$,

$$(P1) \quad p(0, \mu) = 0;$$

$$(P2) \quad p(\cdot, \mu) \text{ é estritamente convexa e diferenciável em } (-\infty, b);$$

$$(P3) \quad p'(0, \mu) = \mu;$$

$$(P4) \quad \lim_{y \rightarrow b} p'(y, \mu) = +\infty \text{ (coercividade pela direita);}$$

$$(P5) \quad \lim_{y \rightarrow -\infty} p'(y, \mu) = 0.$$

Observe que para $\mu > 0$ as funções de penalidades $p(\cdot, \mu)$ da família \mathcal{P} , como definimos, passam pela origem, têm derivada μ na origem, são coercivas pela direita e são estritamente convexas no interior de seu domínio [12].

2.2 Construindo Penalidades da Família \mathcal{P}

Vimos, na última seção, como são as funções de penalidades que vamos utilizar na definição do algoritmo de Lagrangeano Aumentado. Em seguida veremos metodologias de construir penalidades da família \mathcal{P} . Basicamente, todas usam uma função auxiliar de uma variável real.

Definição 2.2 *Considere $\theta : \mathbb{R} \rightarrow \mathbb{R}$ uma função satisfazendo as seguintes propriedades*

$$(a) \quad \text{intdom } \theta = (-\infty, b), \quad b > 0, \text{ possivelmente } b = +\infty;$$

(b) $\theta(\cdot)$ é estritamente convexa e diferenciável em $(-\infty, b)$;

(c) $\lim_{y \rightarrow b} \theta'(y) = +\infty$ (coercividade pela direita);

(d) $\lim_{y \rightarrow -\infty} \theta'(y) = 0$;

(e) $\theta(0) = 0$ e $\theta'(0) = 1$.

Exemplo 2.3 A função exponencial, $y \in \mathbb{R} \mapsto \theta(y) = e^y - 1$, satisfaz as propriedades (a) – (e) acima e este é um caso em que $b = +\infty$, isto é, $\text{intdom } \theta = (-\infty, +\infty)$.

A seguir serão apresentadas duas metodologias de penalização, as quais serão denominadas **Tipo 1** e **Tipo 2**

Tipo 1: Considere $\theta : \mathbb{R} \rightarrow \mathbb{R}$ satisfazendo as propriedades (a)-(e) dadas acima. Então, p dada por

$$(y, \mu) \in \mathbb{R} \times \mathbb{R}_{++} \mapsto p(y, \mu) = \theta(\mu y) \in \mathbb{R} \quad (2.1)$$

é uma penalidade da família \mathcal{P} .

Exemplo 2.4 Considere $\theta : \mathbb{R} \rightarrow \mathbb{R}$ definida por

$$\theta(y) = e^y - 1$$

exibida anteriormente no exemplo (2.3). Assim p obtida a partir de θ fica

$$y \in \mathbb{R}, \mu \in \mathbb{R}_{++} \mapsto p(y, \mu) = e^{\mu y} - 1.$$

A penalidade tipo 1 dada na relação (2.1) pode ser encontrada em [12].

Tipo 2: Considere $\theta : \mathbb{R} \rightarrow \mathbb{R}$ satisfazendo as propriedades (a)-(e) dadas no início da seção. Então, p dada por

$$(y, \mu) \in \mathbb{R} \times \mathbb{R}_{++} \mapsto p(y, \mu) = \mu\theta(y) \in \mathbb{R} \quad (2.2)$$

é uma penalidade da família \mathcal{P} .

Exemplo 2.5 Considere $\theta : \mathbb{R} \rightarrow \mathbb{R}$ definida por

$$\theta(y) = e^y - 1.$$

Assim p dada em (2.2) obtida a partir de θ fica

$$y \in \mathbb{R} \quad \mu \in \mathbb{R}_{++} \mapsto p(y, \mu) = \mu(e^y - 1).$$

A penalidade tipo 2 da relação (2.2) é a mais utilizada na literatura e pode ser encontrada em [2]. Note que a diferença entre as penalidades tipo 1 e 2 está na posição em que o parâmetro $\mu \in \mathbb{R}_{++}$ ocupa em cada uma destas. Nas penalidades tipo 1, μ multiplica o argumento da função θ , já nas do tipo 2, μ multiplica a função θ [12].

Apresentaremos em seguida os métodos de Lagrangeano Aumentado.

2.3 Métodos de Lagrangeano Aumentado

Nesta seção trataremos do método Lagrangeano Aumentado, sua formulação e propriedades. Lembramos que na sua forma original foram introduzidos para resolver problemas com restrições de igualdade e posteriormente generalizados para problemas com restrições de desigualdade. Neste trabalho, iremos aplicá-los a problemas com restrições de desigualdade, ou seja

$$(P) \quad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & g_i(x) \leq 0, \quad i = 1, 2, \dots, m \\ & x \in \mathbb{R}^n \end{array}$$

onde as funções f e g_i para $i = 1, 2, \dots, m$ são contínuas e possuem derivadas, pelo menos até a primeira ordem.

Em relação ao problema (P), definimos a função Lagrangeano Aumentado como

$$(x, \mu, r) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_{++} \mapsto L(x, \mu, r) \in \mathbb{R}$$

onde

$$L(x, \mu, r) = f(x) + r \sum_{i=1}^m p\left(\frac{g_i(x)}{r}, \mu_i\right) \quad (2.3)$$

e p é uma função de penalidade da família \mathcal{P} .

Note que a função Lagrangeano Aumentado é formada pela função objetivo f somada a uma penalização das restrições.

O algoritmo de Lagrangeano Aumentado aplicado ao problema (P) com penalidade pertencente a família \mathcal{P} , gera seqüências (x^k) e (μ^k) , a partir de $\mu^0 > 0$ fixo e (r^k) uma seqüência de termos positivos e limitada, da seguinte forma:

Algoritmo 2.1

Dados $\mu^0 > 0$, $\bar{r} > 0$ e (r^k) uma seqüência em $(0, \bar{r})$.

$k = 0$

Repita

Encontre

$$x^{k+1} \in \operatorname{argmin}\{L(x, \mu^k, r^k) : x \in \mathbb{R}^n\}$$

Atualize

$$\mu_i^{k+1} = p'\left(\frac{g_i(x^{k+1})}{r^k}, \mu_i^k\right), \quad i = 1, 2, \dots, m.$$

$$k = k + 1$$

onde $L(x, \mu^k, r^k)$ é a função Lagrangeano Aumentado definida na relação (2.3), r^k é o parâmetro de penalidade e μ^k é o vetor dos multiplicadores de Lagrange.

A cada iteração o algoritmo do método Lagrangeano Aumentado resolve um subproblemas irrestritos, que geralmente chamamos de subproblemas gerados pelo algoritmo, e em seguida atualiza o parâmetro de penalidade.

Note que a parte mais pesada do algoritmo (2.1) é a minimização da função Lagrangeano Aumentado. O subproblema irrestrito formado a cada iteração pode ser resolvido por qualquer método de minimização irrestrita.

A seqüência (x^k) gerada pelo algoritmo está bem definida (ver afirmação em [12]).

A seqüência (x^k, μ^k) é chamada de primal - dual. Em geral (μ^k) é sempre viável (dual) para $k > 0$ e espera-se que (x^k) vá se tornando viável à medida que k cresce, como será visto a seguir,

Uma das vantagens de utilizar o método de Lagrangeano Aumentado é que a convergência do método não exige nenhuma restrição quanto ao parâmetro de penalidade, podendo diminuir, aumentar ou manter constante, o que não ocorre nos métodos de penalidades clássicos. Os métodos de penalidades clássicos exigem certas restrições sobre o parâmetro de penalidade para obter a convergência. Os métodos de Lagrangeano Aumentado têm por objetivo conciliar os aspectos de mal-condicionamento e evitar perda de estrutura de minimização proveniente do método de penalização [11].

No algoritmo, a cada iteração o ponto x^{k+1} é determinado por

$$x^{k+1} \in \operatorname{argmin}\left\{f(x) + r^k \sum_{i=1}^m p\left(\frac{g_i(x)}{r^k}, \mu_i^k\right)\right\} \quad (2.4)$$

Inicialmente fornecemos um valor arbitrário para o vetor μ (por exemplo, vetor nulo ou vetor de uns), e a cada iteração esses valores são atualizados forçando a

satisfazer a primeira e a última relação de KKT dadas em (1.19), como mostraremos a seguir.

Derivando, em relação a x , a função Lagrangeano Aumentado dada na relação (2.3) tem-se

$$(x, \mu, r) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_{++} \mapsto \nabla_x L(x, \mu, r) = \nabla f(x) + r^k \sum_{i=1}^m p' \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \nabla g_i(x). \quad (2.5)$$

Se escolhermos

$$\mu^{k+1} = p' \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \quad (2.6)$$

e substituirmos em (2.5) tem-se

$$\nabla f(x) + r^k \sum_{i=1}^m \underbrace{p' \left(\frac{g_i(x)}{r^k}, \mu_i^k \right)}_{\mu^{k+1}} \nabla g_i(x). \quad (2.7)$$

Observe que (2.7) é a primeira relação de KKT em (1.19). Além disso, sendo $p \in \mathcal{P}$ uma penalidade da família \mathcal{P} , é conhecido que $p'(\cdot, \mu^k) \geq 0$, ou seja, $\mu^{k+1} \geq 0$. Assim, os algoritmos de Lagrangeano Aumentado são construídos forçando-se a satisfazer algumas das condições de otimalidade.

Capítulo 3

Programação Quadrática

Em otimização, os problemas restritos que são compostos por função objetivo quadrática e restrições lineares, são chamados de problemas de Programação Quadrática. Esse tipo de problema é muito importante, pois eles surgem na formulação matemática de várias situações práticas ou ainda como subproblemas gerados por métodos de programação não linear como é o caso de Programação Quadrática Seqüencial, que veremos com mais detalhes adiante.

O problema quadrático geral com restrições de igualdade e desigualdade é definido da seguinte forma:

$$\begin{array}{ll} \text{minimizar} & q(x) = \frac{1}{2}x^t Gx + x^t d \\ (PQ) \quad \text{sujeito a} & a_i^t x = b_i \quad i \in E = 1, 2, \dots, p \\ & a_i^t x \leq b_i \quad i \in I = 1, 2, \dots, m \end{array}$$

sendo $[a_i]$ $i \in E \cup I$, x e $d \in \mathbb{R}^n$ e $G \in \mathbb{R}^{n \times n}$ simétrica.

Pode-se resolver um problema quadrático em um número finito de iterações mas o quão trabalhoso será para encontrar a solução para esse problema depende das características da função objetivo e do número de restrições. Uma das dificuldades na resolução de um problema quadrático é que esse problema pode não ter solução. Isto acontece em duas situações:

- Quando a região viável é vazia, ou seja, quando não tem intersecções entre as restrições;
- Quando a função objetivo não é limitada inferiormente na região viável.

Se a matriz hessiana G é semidefinida positiva, o problema é convexo, e nesse caso resolver esse tipo de problema, às vezes não é mais difícil do que resolver um problema linear. No caso da matriz hessiana G ser indefinida, o problema quadrático é não convexo, podendo ser mais desafiador resolver esse problema.

Primeiramente serão tratados os problemas de programação quadrática com restrições de igualdade e alguns métodos para minimizar esse tipo de problema. Na seção referente a programação quadrática com restrições de igualdade e desigualdade será visto que os métodos existentes na literatura recaem no caso onde o problema é composto somente por restrições de igualdade.

E por fim vamos propor a aplicação do método Lagrangeano Aumentado com penalidade quadrática para problemas quadráticos convexos com restrições de desigualdade.

3.1 Programação Quadrática com Restrições de Igualdade

Iniciamos a nossa discussão pelo algoritmo para programação quadrática, considerando o caso em que somente restrições de igualdade estão presentes.

Considere o problema que desejamos resolver

$$(PQE) \quad \begin{array}{ll} \text{minimizar} & q(x) = \frac{1}{2}x^t Gx + x^t d \\ \text{sujeito a} & Ax = b \end{array}$$

onde $A \in \mathbb{R}^{m \times n}$ é a matriz jacobiana das restrições definida por

$$A = [a_i]_{i \in E}^t.$$

Sendo o número de restrições igual a m , e assumindo que $m \leq n$, supondo que A tem posto linha completo, ou seja, $\text{posto}(A) = m$ e as restrições do problema (PQE) são consistentes, então as condições de Karush-Kuhn-Tucker para o problema quadrático com restrições de igualdade é da seguinte forma:

Teorema 3.1 *Suponha que x^* é solução local de (PQE) e que satisfaz a condição de qualificação (LICQ), dada na definição (1.18). Então há um vetor de multiplicadores de Lagrange λ^* , com componentes λ_i^* , $i \in E$, tal que as seguintes condições são satisfeitas em (x^*, λ^*)*

$$\begin{aligned} Gx^* + d + A^t \lambda^* &= 0, \\ Ax^* - b &= 0. \end{aligned} \tag{3.1}$$

A condição de otimalidade de primeira ordem para x^* ser a solução do (PQE) é especificar qual é o vetor λ^* que satisfaz o seguinte sistema de equações:

$$\begin{bmatrix} G & A^t \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix}. \tag{3.2}$$

Esse sistema de equações são as condições de primeira ordem (3.1), reescritas matricialmente. O vetor λ é o vetor de multiplicadores de Lagrange.

O sistema (3.2) pode ser escrito de uma forma mais usual para implementação computacional, fazendo

$$x^* = x + p$$

onde x é uma estimativa da solução e p é o passo desejado, substituindo em (3.2) tem-se:

$$\begin{bmatrix} G & A^t \\ A & 0 \end{bmatrix} \begin{bmatrix} x + p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix} \implies \begin{cases} Gx + Gp + A^t\lambda^* = -d \\ Ax + Ap = b \end{cases} \quad (3.3)$$

Fazendo

$$c = Ax - b, \quad g = d + Gx \quad e \quad p = x^* - x,$$

e substituindo em (3.3) tem-se:

$$\begin{bmatrix} G & A^t \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} g \\ c \end{bmatrix} \quad (3.4)$$

A matriz, no sistema (3.4), é chamada de matriz KKT, e o resultado seguinte prova a não-singularidade dessa matriz, para isso usa-se Z para denotar a matriz $n \times (n - m)$ cuja as colunas formam uma base para o espaço nulo de A . Ou seja, Z tem posto completo e $AZ = 0$.

Lema 3.2 *Considere a matriz A com posto completo e Z^tGZ positiva definida, então a matriz KKT*

$$KKT = \begin{bmatrix} G & A^t \\ A & 0 \end{bmatrix}$$

é não singular e em particular o par de vetores (x^, λ^*) satisfaz (3.2).*

Prova. Considere vetores p e v tais que satisfaçam o sistema homogêneo a seguir:

$$\begin{bmatrix} G & A^t \\ A & 0 \end{bmatrix} \begin{bmatrix} p \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \begin{cases} Gp + A^tv = 0 \\ Ap = 0 \end{cases} \quad (3.5)$$

Multiplicando a primeira equação do sistema (3.5) por p^t , tem-se

$$p^tGp + (Ap)^tv = 0.$$

Pela equação (3.5) $Ap = 0$, então $p^t G p = 0$. Se Z é base para o espaço nulo de A então pode-se escrever p como uma combinação linear de Z , $p = Z\mu$ para algum vetor $\mu \in \mathbb{R}^{n-m}$ assim:

$$0 = p^t G p = (Z\mu)^t G (Z\mu) = \mu^t Z^t G Z \mu$$

como por hipótese $(Z^t G Z)$ é definida positiva isso implica que $\mu = 0$, dessa forma $p = Z\mu = 0$, logo $p = 0$. Substituindo $p = 0$ na primeira equação do sistema (3.5), obtém-se $A^t v = 0$, logo $AA^t v = 0$. Como $\text{posto}(AA^t) = \text{posto}(A) = m$, a matriz (AA^t) é não singular e conseqüentemente $v = 0$. Assim o sistema homogêneo (3.5) só tem a solução trivial, portanto a matriz KKT é não-singular. ■

Vimos que se as condições (3.2) são satisfeitas, então há um único par de vetor (x^*, λ^*) que satisfaz a condição de otimalidade de primeira ordem para (PQE). A condição suficiente de otimalidade de segunda ordem é satisfeita em (x^*, λ^*) quando x^* é um minimizador local estrito para (PQE).

Teorema 3.3 *Considere o par de vetores (x^*, λ^*) do lema (3.2). Então o vetor x^* é a solução global única do problema (PQE).*

Prova. Seja x um ponto viável, isto é, $Ax = b$ e tome $p = x^* - x$, desde que $Ax^* = Ax = b$, tem-se que $Ap = 0$. Substituindo na função objetivo do problema (PQE), tem-se:

$$\begin{aligned} q(x) &= \frac{1}{2}x^t G x + x^t d \\ q(x^* - p) &= \frac{1}{2}(x^* - p)^t G (x^* - p) + d^t (x^* - p) \\ &= \frac{1}{2} [x^{*t} G x^* - x^{*t} G p - p^t G x^* + p^t G p] + x^{*t} d - p^t d \\ &= \frac{1}{2} p^t G p - p^t G x^* - p^t d + q(x^*). \end{aligned}$$

Da condição de otimalidade de primeira ordem para (PQE) tem-se $Gx^* = -d + A^t \lambda^*$ e desde que $Ap = 0$, substituindo em $q(x^* - p)$ tem-se:

$$\begin{aligned} q(x^* - p) &= \frac{1}{2} p^t G p - p^t (-d + A^t \lambda^*) - p^t d + q(x^*) \\ &= \frac{1}{2} p^t G p + p^t d - (Ap)^t \lambda^* - p^t d + q(x^*) \\ &= \frac{1}{2} p^t G p + q(x^*) \end{aligned}$$

desde que, p exista dentro do espaço nulo de A , pode-se escrever $p = Z\mu$, para algum vetor $\mu \in \mathbb{R}^{n-m}$, assim

$$\begin{aligned} q(x^* - p) &= \frac{1}{2} (Z\mu)^t G (Z\mu) + q(x^*) \\ &= \frac{1}{2} \mu^t Z^t G Z \mu + q(x^*) \end{aligned}$$

Por hipótese, $Z^t G Z$ é definida positiva, conclui-se que $q(x) > q(x^*)$ exceto quando $\mu = 0$, ou seja, quando $x = x^*$. Conseqüentemente, x^* é solução global única de (PQE). ■

Na próxima seção serão discutidos alguns métodos para resolver o sistema KKT ou alternativas para isso.

3.1.1 Resolvendo o Sistema KKT

Fatoração Simétrica Indefinida

Para o caso, onde o problema (PQE), é não convexo, ou seja, a matriz G é indefinida pode-se achar a solução fazendo a fatorização triangular total da matriz KKT e depois fazer as substituições, mas não pode-se usar o algoritmo de fatorização Cholesky, pois a matriz KKT é indefinida. Outra alternativa é usar a eliminação Gaussiana com pivoteamento parcial e obter as matrizes L e U , mas esse método tem a desvantagem de ignorar a simetria do sistema. A estratégia mais eficiente para esse caso é usar a Fatoração Simétrica Indefinida.

$$P^t \mathbf{K} P = L B L^t \quad (3.6)$$

onde P é a matriz permutação, L é a matriz triangular inferior e B é a matriz com blocos na diagonal, com blocos de (1×1) ou (2×2) e \mathbf{K} a matriz a ser fatorada.

As permutações simétricas definida pela matriz P servem para a estabilidade numérica dos cálculos e para o caso em que a matriz KKT é esparsa, essas permutações mantêm a esparsidade da mesma.

Para resolver o sistema KKT (3.2), primeiramente calcula-se a fatorização (3.6), utilizando a matriz KKT no lugar da matriz \mathbf{K} e depois realiza a seguinte seqüência de operações para se chegar à solução:

$$\text{resolva } Ly = P^t \begin{bmatrix} g \\ c \end{bmatrix} \text{ para obter } y;$$

$$\text{resolva } B\hat{y} = y \text{ para obter } \hat{y};$$

$$\text{resolva } L^t \bar{y} = \hat{y} \text{ para obter } \bar{y};$$

$$\text{e determine } \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = P\bar{y}.$$

Método Espaço Linha

O método Espaço Linha é um caso especial do método visto anteriormente - o método Fatoração Simétrica Indefinida, mas nesse método o problema precisa ser convexo, pois usa-se a matriz G para realizar eliminação em blocos no sistema (3.2), assumindo que $G > 0$ pode-se multiplicar a primeira equação do sistema (3.2) por AG^{-1} e o resultado subtraído da segunda equação do mesmo sistema para obter o seguinte sistema linear em λ^* :

$$(AG^{-1}A^t)\lambda^* = AG^{-1}g - c.$$

Resolvendo-se esse sistema para λ^* recupera-se p a partir da primeira equação de (3.2):

$$Gp = A^t\lambda^* - g.$$

Esses cálculos requerem operações com a matriz G^{-1} além da fatorização da matriz $AG^{-1}A^t$, e isso só é viável quando [15]:

- A matriz G é bem condicionada e fácil de inverter;
- A matriz G^{-1} é conhecida explicitamente;
- O número de restrições é pequeno, assim o número de pré-cálculos necessários para formar a matriz $(AG^{-1}A^t)$ não será grande.

Método Espaço Nulo

O método do espaço nulo, não requer a não singularidade da matriz G e conseqüentemente tem maior aplicação do que o método espaço linha, visto anteriormente. Ele assume somente as condições do lema (3.2), isto é, $\text{posto}(A) = m$ e $Z^tGZ > 0$. Isso requer, de qualquer modo, conhecer Z , matriz $n \times (n - m)$ cuja as colunas formam uma base para o espaço nulo de A , ou seja, $AZ = 0$.

Suponha o particionamento do vetor p em duas componentes:

$$p = Yp_y + Zp_z,$$

onde, $Y \in \mathbb{R}^{n \times m}$ tal que $[Y \mid Z]$ é não singular, $p_y \in \mathbb{R}^m$ e $p_z \in \mathbb{R}^{n-m}$.

Substituindo p na segunda equação do sistema (3.2), e lembrando que $AZ = 0$, tem-se:

$$(AY)p_y = -c. \tag{3.7}$$

Desde que A tenha posto completo e $[Y \mid Z]$ é não singular, o produto

$A[Y \mid Z] = [AY \mid 0]$ também tem posto completo . Portanto AY é não singular e

$$p_y = -(AY)^{-1}c$$

Substituindo $p = Yp_y + Zp_z$ na primeira equação do sistema (3.4) e multiplicando por Z^t tem-se:

$$(Z^tGZ)p_z = -(Z^tGYp_y + Z^tGZ), \quad (3.8)$$

esse sistema pode ser resolvido por meio da fatorização de Cholesky na matriz (Z^tGZ) , determinar p_z e conseqüentemente p . Para obter os multiplicadores de Lagrange, basta multiplicar a primeira equação do sistema (3.2) por Y^t , e tem-se:

$$(AY)^t\lambda^* = Y^t(g + Gp).$$

O sistema (3.8) também pode ser resolvido por meio do método Gradiente Conjugado. Uma das vantagens de usar esse método é que na ausência de uma representação explícita de Z pode-se calcular vetores arbitrários ao longo do processo.

Outros métodos, para resolver o sistema KKT, e uma maior discussão sobre esse assunto pode ser encontrado em [15] e [1].

3.2 Programação Quadrática com Restrições de Igualdade e Desigualdade

Nesta seção serão discutidas algumas classes de algoritmos para resolver problemas de programação quadrática com restrições de igualdade e desigualdade.

Primeiramente, vamos rever as condições de otimalidade para problemas de programação quadrática com restrições de desigualdade e igualdade. Considere o problema (PQ) definido anteriormente, o qual reescrevemos novamente.

$$(PQ) \quad \begin{array}{ll} \text{minimizar} & q(x) = \frac{1}{2}x^tGx + x^td \\ \text{sujeito a} & a_i^tx = b_i, \quad i \in E = 1, 2, \dots, p, \\ & a_i^tx \leq b_i, \quad i \in I = 1, 2, \dots, m, \end{array}$$

sendo x e $d \in \mathbb{R}^n$ e $G \in \mathbb{R}^{n \times n}$ simétrica.

A função Lagrangeano para esse problema pode ser definida como:

$$l(x, \lambda) = \frac{1}{2}x^tGx + x^td + \sum_{i \in I \cup E} \lambda_i(a_i^tx - b_i).$$

O conjunto ativo para o ponto ótimo (os índices das restrições de igualdade e os índices das restrições de desigualdade que estão satisfeitas na igualdade), é

$$\mathbf{A}(x^*) = \{i \in I \cup E \mid a_i^t x^* = b_i\}.$$

Para simplificar, as condições gerais (1.20), assumindo que a condição do teorema (1.19) são satisfeitas, conclui-se que a solução x^* de (PQ) satisfaz a condições de primeira ordem:

$$\begin{aligned} Gx^* + d + \sum_{i \in \mathbf{A}(x^*)} \lambda_i^* a_i &= 0, \\ a_i^t x^* &= b_i, \quad \forall i \in \mathbf{A}(x^*), \\ a_i^t x^* &\leq b_i, \quad \forall i \in I \cap \mathbf{A}(x^*), \\ \lambda_i^* &\geq 0, \quad \forall i \in I \cap \mathbf{A}(x^*). \end{aligned} \tag{3.9}$$

Omitindo detalhes da condição de otimalidade de segunda ordem, tem-se que para x^* ser um minimizador local, a condição de otimalidade de segunda ordem é satisfeita se $Z^t G Z \geq 0$, sendo Z base para o espaço nulo da matriz Jacobiana das restrições ativas:

$$[a_i]_{i \in \mathbf{A}(x^*)}^t.$$

Desde que se as condições do teorema (3.3) são satisfeitas então x^* é o minimizador global para o problema (PQ). Quando o problema é dito indefinido ou não convexo, ocorrem algumas complicações para os algoritmos. Nas figuras a seguir têm-se alguns exemplos de problemas de programação quadrática não convexos.

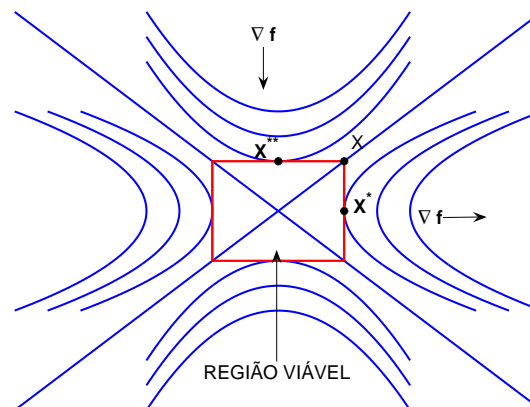


Figura 3.1: G indefinida

Na primeira figura (3.1), a matriz G tem um autovalor positivo e um autovalor negativo, note que x^{**} é um maximizador local, x^* é um minimizador local e x é um ponto estacionário.

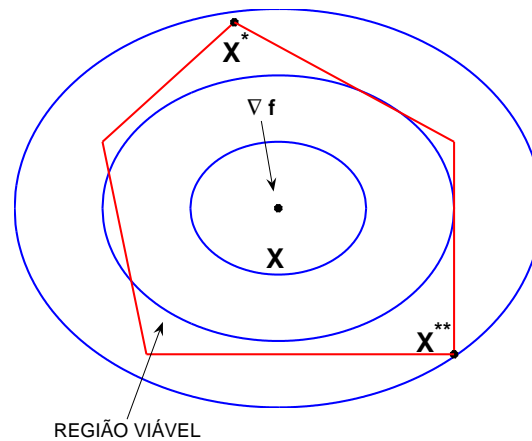


Figura 3.2: G definida negativa

Na segunda figura (3.2), a matriz G tem dois autovalores negativos, note que x é um maximizador global e os pontos x^* e x^{**} são minimizadores locais.

Uma segunda característica que causa dificuldades para os algoritmos é a degeneração. A degeneração aparece essencialmente em situações como estas:

- Os gradientes das restrições ativas $a_i, i \in \mathbf{A}(x^*)$, são linearmente dependentes.
- Existe multiplicador de Lagrange ótimo, $\lambda_i^* = 0$, para algum índice ativo $i \in \mathbf{A}(x^*)$.

Se $\mathbf{A}(x^*)$ fosse conhecido, poderíamos achar a solução aplicando uma das técnicas para resolver problemas de programação quadrática com restrições de igualdade (PQE), vistas na seção anterior. Mas como veremos, a determinação desse conjunto é o principal desafio dos algoritmos para resolver (PQ).

3.3 Métodos para Resolver o Problema (PQ)

Essa seção será iniciada pelo método das **Restrições Ativas** para problemas convexos. O método das restrições ativas começa com uma suposição de um conjunto ativo viável, mas essa suposição pode estar incorreta, nas iterações o método usa informações como o gradiente e os multiplicadores de Lagrange para corrigir a estimativa e assim adicionar restrições que estão ativas ou remover as restrições que não estão ativas.

Esse método tem algumas variações como primal, dual e primal-dual. Nós vamos restringir nossas discussões para o primal, que nas iterações em geral se mantém viável ao problema (PQ), enquanto decresce a função objetivo.

Dado um ponto x^k viável tem-se o conjunto das restrições ativa:

$$\mathbf{A}^k = \mathbf{A}^k(x^k) = \{i \in I \cup E \mid a_i^t x^k = b\}$$

A cada iteração tenta-se localizar a solução para o problema (PQ), onde somente as restrições ativas aparecem. Para isso, as vezes é mais adequado substituir o x^k original por uma correção $\delta^k = x - x^k$, ficando com o problema de programação quadrática com restrições de igualdade:

$$\begin{aligned} \text{minimizar} \quad & \frac{1}{2} \delta^t G \delta + \delta^t g^k \\ \text{sujeito a} \quad & a_i^t \delta = 0 \quad i \in A^k \end{aligned} \tag{3.10}$$

sendo, $g^k = d + Gx^k$, ou seja, $g^k = \nabla q(x^k)$. Assim, pode-se resolver (3.10) por qualquer método visto na seção anterior.

Se δ^k é viável, isto é, viável somente para as restrições que estão ativas em x^k , a próxima iteração é

$$x^{k+1} = x^k + \delta^k,$$

se não, fazemos uma busca linear na direção de δ^k para achar o melhor ponto viável e a próxima iteração é

$$x^{k+1} = x^k + \lambda^k \delta^k.$$

Pode-se calcular o passo λ^k usando uma busca unidirecional:

$$\lambda^k = \min \left(1, \min_{\substack{i \notin A^k \\ a_i^t p^k < 0}} \frac{b_i - a_i^t x^k}{a_i^t p^k} \right) \tag{3.11}$$

Se $\lambda^k < 1$ então uma nova restrição torna-se ativa. O método do conjunto ativo, usualmente acrescenta uma restrição ativa a cada iteração. Com isso, esse método, pode necessitar de várias iterações para chegar em um resultado.

O método do **Gradiente Projetado** é construído para fazer escolhas rápidas no conjunto ativo. Esse método acelera o processo para a solução por permitir mudanças rápidas dentro do conjunto ativo, o gradiente projetado é mais eficiente quando existem poucas restrições, em particular quando as restrições são de caixa, ou seja, quando existe um limite inferior e superior para as variáveis. Não é necessário assumir $G > 0$, pois esse método pode ser aplicado tanto para problemas convexos quanto para os não convexos.

O método do gradiente projetado usa a direção do método de Cauchy, isto é, $d = -\nabla f(x)$ combinado com a projeção ortogonal do gradiente, na variedade linear das restrições ativas.

Uma discussão mais detalhada sobre os métodos do Conjunto Ativo e Gradiente Projetado pode ser encontrada em [4], [1].

3.4 Lagrangeano Aumentado com Penalidade Quadrática

Penalidade Quadrática Clássica

A função de penalidade proposta por Rockafellar [16] com parâmetro de penalidade $r > 0$, tem a seguinte forma:

$$y \in \mathbb{R}, \mu \in \mathbb{R}_+ \mapsto p(y, \mu) = \frac{1}{2r}(\max\{0, ry + \mu\}^2 - \mu^2), \quad (3.12)$$

ou equivalentemente

$$p(y, \mu) = \begin{cases} \frac{r}{2}y^2 + \mu y, & \text{se } y \geq -\frac{\mu}{r} \\ -\frac{\mu^2}{2r}, & \text{se } y < -\frac{\mu}{r} \end{cases} \quad (3.13)$$

Para $\mu \in \mathbb{R}_+$ fixo a derivada em (3.13) tem a forma

$$p'(y, \mu) = \begin{cases} ry + \mu, & \text{se } y \geq -\frac{\mu}{r} \\ 0, & \text{se } y < -\frac{\mu}{r} \end{cases} \quad (3.14)$$

As figuras a seguir mostram o gráfico da função p (3.13) e da função p' (3.14) para o caso em que $r = \mu = 1$.

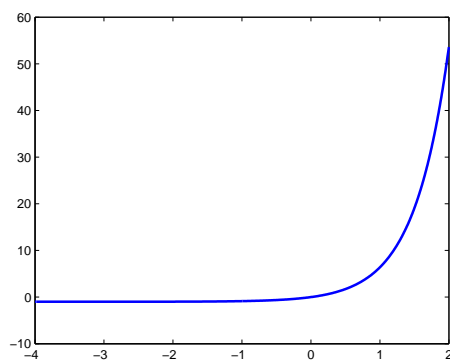


Figura 3.3: Função Quadrática Clássica

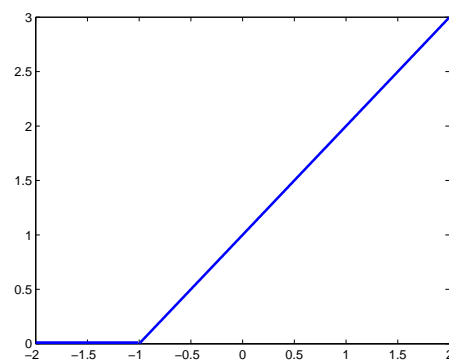


Figura 3.4: Derivada

Como a função de Rockafellar não possui segunda derivada no ponto $y = -\frac{\mu}{r}$, pode-se ter dificuldades ao usar métodos que utilizem derivada de ordem 2, como é o caso, por exemplo, do método de Newton e do método de Região de Confiança.

Função Quadrática

A função quadrática θ a ser utilizada no algoritmo de Lagrangeano Aumentado, foi introduzida em [10], é definida como

$$y \in \mathbb{R} \mapsto \theta(y) = \frac{1}{2} y^2 + y, \quad (3.15)$$

A figura a seguir mostra o gráfico da função θ dada na relação acima.

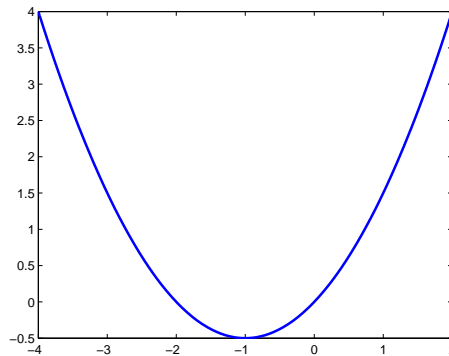


Figura 3.5: Função Quadrática

A função θ definida acima, não é crescente em todo seu domínio, isto é, no intervalo de $(-\infty, -1)$ é decrescente e no intervalo de $(-1, +\infty)$ é crescente. Por não ser crescente em todo o seu domínio gera derivadas direcionais negativas. Com a função θ dada na relação (3.15) as funções de penalidades construídas a partir dessa, não serão da família \mathcal{P} (Capítulo 2). Observa-se que a propriedade de limitação por baixo da derivada, dada no item (P5) da definição (2.1), não é satisfeita, ver [10].

Outro fato a destacar é com relação ao algoritmo de Lagrangeano Aumentado (2.1) com as penalidades tipo 1 e 2 e a função θ dada na relação (3.15), esse algoritmo poderá gerar multiplicadores negativos, pois θ é decrescente no intervalo $(-\infty, -1)$. Este fato é um problema, pois foge à filosofia da classe de métodos de Lagrangeano Aumentado. No entanto, em [12] foi mostrado que é possível contornar este problema através do aumento do parâmetro de penalidade. Além disso, foi mostrado em [12] que o algoritmo de Lagrangeano Aumentado com penalidade tipo 1 e θ dada na relação (3.15) possui boas propriedades de convergências.

Neste trabalho estamos propondo a utilização da função quadrática (3.15) juntamente com a penalidade p tipo 1 dada na relação (2.1), em problemas de programação quadrática.

Considere o problema de programação quadrática com restrições lineares de

desigualdade

$$(PQI) \quad \begin{array}{ll} \text{minimizar} & q(x) = \frac{1}{2}x^t Gx + x^t d \\ \text{sujeito a} & Ax \leq b \end{array}$$

onde $G \in \mathbb{R}^{n \times n}$ é simétrica e definida positiva, $A \in \mathbb{R}^{m \times n}$ é a matriz Jacobiana das restrições definida por

$$A = [a_i]^t \quad i = 1, \dots, m.$$

Em que o número de restrições é igual a m , com $m \leq n$, A tem posto linha completo, ou seja, $\text{posto}(A) = m$, e as restrições do problema (PQI) são consistentes.

A função Lagrangeano Aumentado associada ao problema quadrático (PQI) é da forma

$$(x, \mu, r) \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_{++} \mapsto L(x, \mu, r) = q(x) + r \sum_{i=1}^m p\left(\frac{a_i^t x - b_i}{r}, \mu_i\right) \quad (3.16)$$

Usando a função penalidade $p \in \mathcal{P}$ tipo 1 (2.1)

$$(y, \mu) \in \mathbb{R} \times \mathbb{R}_{++} \mapsto p(y, \mu) = \theta(\mu y)$$

e θ dada na relação (3.15)

$$y \in \mathbb{R} \mapsto \theta(y) = \frac{1}{2}y^2 + y$$

tem-se

$$\begin{aligned} p(y, \mu) = \theta(\mu, y) &= \frac{1}{2}(\mu y)^2 + (\mu y) \\ &= \frac{1}{2}(\mu^2 y^2) + \mu y \end{aligned} \quad (3.17)$$

Substituindo (3.17) em (3.16) obtém-se

$$\begin{aligned}
 L(x, \mu, r) &= q(x) + r \sum_{i=1}^m \theta \left(\mu_i \frac{(a_i^t x - b_i)}{r} \right) \\
 &= q(x) + r \sum_{i=1}^m \frac{1}{2} \left(\frac{\mu_i}{r} (a_i^t x - b_i) \right)^2 + \frac{\mu_i}{r} (a_i^t x - b_i) \\
 &= q(x) + \sum_{i=1}^m \frac{\mu_i^2}{2r} (a_i^t x - b_i)^2 + \mu_i (a_i^t x - b_i) \\
 &= \frac{1}{2} x^t G x + x^t d + \sum_{i=1}^m \frac{\mu_i^2}{2r} (a_i^t x - b_i)^2 + \mu_i (a_i^t x - b_i) \\
 &= \frac{1}{2} x^t G x + \sum_{i=1}^m \frac{\mu_i^2}{2r} (a_i^t x - b_i)^2 + x^t d + \sum_{i=1}^m \mu_i (a_i^t x - b_i) \quad (3.18)
 \end{aligned}$$

uma função quadrática.

Derivando, em relação a x , a função Lagrangeano Aumentado dada na relação (3.18) tem-se

$$\begin{aligned}
 \nabla_x L(x, \mu, r) &= \nabla q(x) + \sum_{i=1}^m \frac{\mu_i^2}{r} (a_i^t x - b_i) a_i + \mu_i a_i \\
 &= \nabla q(x) + \sum_{i=1}^m \left(\frac{\mu_i^2}{r} (a_i^t x - b_i) + \mu_i \right) a_i \quad (3.19)
 \end{aligned}$$

escolhendo-se

$$\bar{\mu}_i = \left(\frac{\mu_i^2}{r} (a_i^t x - b_i) + \mu_i \right)$$

e substituímos em (3.19) tem-se

$$\nabla q(x) + \underbrace{\sum_{i=1}^m \left(\frac{\mu_i^2}{r} (a_i^t x - b_i) + \mu_i \right) a_i}_{\bar{\mu}_i} \quad (3.20)$$

ou seja,

$$\nabla_x L(x, \mu, r) = Gx + d + \sum_{i=1}^m \bar{\mu}_i a_i, \quad (3.21)$$

a qual é a primeira condição de KKT dada na relação (1.19).

A matriz hessiana da função Lagrangeano Aumentado com penalidade qua-

drática para problemas de programação quadrática é da seguinte forma:

$$\nabla_{xx}^2 L(x, \mu, r) = G + \frac{1}{r} \sum_{i=1}^m \mu_i^2 a_i a_i^t \quad (3.22)$$

Com as relações obtidas acima, é possível mostrar o teorema a seguir que é a principal contribuição dessa dissertação.

Teorema 3.4 *Considere o método de Lagrangeano Aumentado com a penalidade tipo 1 dada em (2.1) e a função θ dada na relação (3.15). Assim:*

Se a matriz do problema quadrático (PQI) é semidefinida positiva, então, a matriz hessiana dada na relação (3.22), é semidefinida positiva.

Se a matriz do problema quadrático (PQI) é definida positiva, então, a matriz hessiana dada na relação (3.22), é definida positiva.

Prova.

Considere a matriz hessiana dada na relação (3.22), reescrita a seguir

$$\nabla_{xx}^2 L(x, \mu, r) = G + \frac{1}{r} \sum_{i=1}^m \mu_i^2 a_i a_i^t$$

e suponha um vetor $x \in \mathbb{R}^n$ arbitrário.

Multiplicando ambos os lados da equação acima pelo vetor x , obtém-se:

$$\begin{aligned} x^t \nabla_{xx}^2 L(x, \mu, r) x &= x^t G x + \frac{1}{r} \sum_{i=1}^m \mu_i^2 x^t a_i a_i^t x \\ x^t \nabla_{xx}^2 L(x, \mu, r) x &= x^t G x + \frac{1}{r} \sum_{i=1}^m \mu_i^2 \|a_i^t x\|^2 \end{aligned} \quad (3.23)$$

No segundo termo, o parâmetro de penalidade r é positivo, os multiplicadores de Lagrange são não negativos e $\|a_i x\|^2 > 0$. Desta forma:

$$\frac{1}{r} \sum_{i=1}^m \mu_i^2 \|a_i^t x\|^2 \geq 0$$

Agora, se G é definida positiva, ou seja $x^t G x > 0$ para $x \neq 0$, segue que

$$\nabla_{xx}^2 L(x, \mu, r) > 0.$$

Para o caso em que G é semidefinida positiva, ou seja $x^t G x \geq 0$ para todo $x \in \mathbb{R}^n$, segue que

$$\nabla_{xx}^2 L(x, \mu, r) \geq 0.$$

■

Pela condição suficiente de otimalidade de segunda ordem (1.14), se $\nabla_{xx}^2 L(x, \mu, r)$ é continua numa vizinhança de x^* , $\nabla f(x^*) = 0$ e $\nabla_{xx}^2 L(x^*, \mu, r) > 0$, então x^* é um minimizador local estrito de f .

Logo se o problema (PQI) é estritamente convexo, caso em que a matriz hessiana, G , da função objetivo é definida positiva, a função Lagrangeano Aumentado associada ao problema (PQI) possui hessiana definida positiva o que nos motivou a propor o método de Lagrangeano Aumentado com a função penalidade tipo 1 dada na relação (2.1) e θ quadrática dada na relação (3.15) para resolver os subproblemas quadráticos gerados pelo método de Programação Quadrática Seqüencial que é o tópico do próximo capítulo.

Capítulo 4

Programação Quadrática Seqüencial

Segundo os autores [15] e [11] a Programação Quadrática Seqüencial é um dos métodos mais eficientes para problemas de programação não linear.

A principal idéia dessa classe de métodos é a geração de subproblemas quadráticos e a cada passo resolvê-los. Os métodos de Programação Quadrática Seqüencial são as generalizações do método de Newton para o problema geral de otimização, onde geralmente tem-se um problema com função objetivo e restrições não lineares. O que a Programação Quadrática Seqüencial faz é substituir, a cada iteração, a função objetivo por uma aproximação quadrática da função lagrangeano do problema original num ponto x^k e as restrições por aproximações lineares também no ponto x^k . Dessa maneira o subproblema a ser resolvido a cada iteração k é um subproblema quadrático com restrições lineares que comparado com o problema original pode ser considerado de mais fácil resolução.

Esses métodos podem ser considerados métodos primais-duais, no sentido que eles trabalham simultaneamente no espaço das variáveis primais e no espaço dos multiplicadores de KKT (variáveis duais). A Programação Quadrática Pura basicamente consiste em resolver as equações de KKT pelo método de Newton, então começaremos descrevendo esse método. Para um melhor aprofundamento do métodos de programação quadrática seqüencial ver [15], [11], [13], [3] e [14].

4.1 Programação Quadrática Seqüencial Pura

Problemas que contém unicamente restrições de igualdade não são muito comuns na prática, mas inicialmente vamos restringir nossos estudos a esse caso. Considere o problema que desejamos resolver:

$$(PE) \quad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & h_i(x) = 0 \quad i = 1, \dots, m \\ & x \in \mathbb{R}^n \end{array}$$

sendo, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ funções continuamente diferenciáveis e h um vetor de m funções h_i .

A função Lagrangeano para esse problema é

$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i h_i(x)$$

A idéia principal da Programação Quadrática Seqüencial para (PE) é, a partir de x^k , fazer uma aproximação que gera um subproblema quadrático e, após resolver esse subproblema, definir o novo ponto x^{k+1} . Uma das maneira de achar a solução ótima desse subproblema é encontrar o ponto KKT. Essa busca é feita através da resolução do sistema com $n + m$ variáveis (x, λ) e $n + m$ equações

$$F(x, \lambda) = \begin{bmatrix} \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla h_i(x) \\ h_i(x) \end{bmatrix} = 0, \quad i = 1, \dots, m. \quad (4.1)$$

Será usado A^k para denotar a matriz jacobiana das restrições h no ponto x^k , isto é,

$$A^{k^t} = [\nabla h_1(x^k), \dots, \nabla h_m(x^k)],$$

e a matriz hessiana em x da função Lagrangeano associada ao problema (PE) no ponto (x^k, λ^k) será denotada por

$$W^k = \nabla_{xx}^2 L(x^k, \lambda^k) = \nabla^2 f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k).$$

Se A^* tem posto completo, a solução ótima (x^*, λ^*) do problema (PE) satisfaz (4.1).

O passo de Newton da iteração k é dado por

$$\begin{bmatrix} x^{k+1} \\ \lambda^{k+1} \end{bmatrix} = \begin{bmatrix} p^k \\ v^k \end{bmatrix} + \begin{bmatrix} x^k \\ \lambda^k \end{bmatrix}$$

onde, (p^k, v^k) é a solução do seguinte sistema KKT:

$$\begin{bmatrix} W^k & A^{k^t} \\ A^k & 0 \end{bmatrix} \begin{bmatrix} p^k \\ v^k \end{bmatrix} = \begin{bmatrix} -\nabla_x L(x^k, \lambda^k) \\ -h(x^k) \end{bmatrix} \quad (4.2)$$

Note que a matriz jacobiana de (4.1) no ponto (x^k, λ^k) é

$$J^k = J(x^k, \lambda^k) = \begin{bmatrix} W^k & A^{k^t} \\ A^k & 0 \end{bmatrix} \quad (4.3)$$

O passo de Newton está bem definido quando a matriz J^k é não-singular. Para mostrar que a matriz jacobiana de (4.1) é não-singular usa-se as seguintes hipóteses:

- (H1) A jacobiana das restrições tem posto completo, isto é, $\text{posto}(A^k) = m$;
- (H2) A matriz W^k é definida positiva no núcleo da jacobiana das restrições, isto é, $d^t W^k d > 0$ para todo $d \neq 0$ tal que $A^k d = 0$.

Lema 4.1 *Sob as condições acima a matriz dada por (4.3) é não singular.*

Prova. A prova é idêntica a do lema (3.2). ■

Com as condições estabelecidas acima, o algoritmo de Newton, para sistemas não lineares, converge quadraticamente para a solução. Entretanto, o método de Newton tem alguns inconvenientes:

- o sistema (4.2) poderá determinar não somente os possíveis minimizadores locais, mas também os maximizadores e os pontos de sela.
- a seqüência (x^k, λ^k) poderá não convergir se a escolha do ponto inicial não for suficientemente próxima da solução ótima (x^*, λ^*) .

A escolha de um ponto inicial próximo da solução ótima do problema é o principal inconveniente para a construção do algoritmo geral e real baseado no método de Newton. Desejando remediar esses inconvenientes e ainda fazer uso da convergência quadrática do passo de Newton, quando o ponto inicial está próximo da solução ótima, usa-se o método de Newton associado a outros métodos.

4.2 Estrutura de Programação Quadrática Seqüencial

Existe uma outra maneira de olharmos os sistemas (4.2). Suponha que na iteração k o subproblema quadrático definido seja

$$(SQ) \quad \begin{array}{ll} \text{minimizar} & \frac{1}{2} p^t W^k p + \nabla^t f^k p \\ \text{sujeito a} & A^k p + h(x^k) = 0 \end{array}$$

lembrando que $W^k = \nabla_{xx}^2 L(x^k, \lambda^k)$.

Ressalta-se que a função objetivo do problema (SQ) difere apenas por uma constante da aproximação quadrática da função Lagrangeano associada ao problema (SQ), definida da seguinte forma:

$$L(p, \lambda) = \frac{1}{2} p^t W^k p + \nabla^t f^k p + \lambda^t (A^k p + h(x^k)).$$

De fato, dado (x^k, λ^k) o modelo quadrático para a função Lagrangeano é

$$M_L(p) = L(x^k, \lambda^k) + \nabla^t L(x^k, \lambda^k) p + \frac{1}{2} p^t W^k p \quad (4.4)$$

mas

$$L(x^k, \lambda^k) = f(x^k) + \lambda^{k^t} h(x^k)$$

e

$$\nabla^t L(x^k, \lambda^k) p = \nabla^t f(x^k) p + \lambda^{k^t} A^k p,$$

então (4.4) pode ser reescrito como

$$M_L(p) = \frac{1}{2} p^t W^k p + \nabla^t f(x^k) p + \nu,$$

com a constante $\nu = L(x^k, \lambda^k) + \lambda^{k^t} A^k p = f(x^k) + \lambda^{k^t} (A^k p + h(x^k))$.

Considerando a restrição do subproblema (SQ), tem-se que $\nu = f(x^k)$. Assim, cada iteração de Programação Quadrática Seqüencial pura, consiste em minimizar o modelo quadrático da função Lagrangeano, sujeito a linearização das restrições.

Se as condições usadas para mostrar a não-singularidade de J^k se verificam, o subproblema (SQ) tem solução única (p^k, μ^k) que satisfaz

$$\begin{bmatrix} W^k p + \nabla f^k + A^{k^t} \mu^k \\ A^k p + h^k \end{bmatrix} = 0. \quad (4.5)$$

Nota-se que p^k e v^k podem ser identificados como solução do sistema de Newton (4.2). Subtraindo-se $A^{k^t} \lambda^k$ em ambos os lados da primeira equação em (4.2), obtêm-se

$$\begin{bmatrix} W^k & A^{k^t} \\ A^k & 0 \end{bmatrix} \begin{bmatrix} p \\ \lambda^{k+1} \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) \\ -h(x^k) \end{bmatrix}, \quad (4.6)$$

pela não singularidade dos coeficientes da matriz tem-se que $p = p^k$ e $\lambda^{k+1} = v^k$ [15].

Conclusão, o sistema KKT (4.2), para o problema (PE) é equivalente às condições de otimalidade para o problema (SQ).

A interpretação em termos do método de Newton facilita a análise de convergência, enquanto que a estrutura da Programação Quadrática Seqüencial, permite desenvolver algoritmos práticos para resolver problemas como (SQ).

A seguir o algoritmo que descreve o processo, [15].

Algoritmo 4.1 (*Programação Quadrática Seqüencial*)

Escolha o par inicial (x_0, λ_0) ;

PARA $k = 0, 1, 2, \dots$

Avaliar $f(x^k)$, $\nabla f(x^k)$, W^k , $h(x^k)$ e A^k ;

Resolver (SQ) para obter p^k e μ^k ;

$x^{k+1} \leftarrow x^k + p^k$;

$\lambda^{k+1} \leftarrow \mu^k$;

SE teste de convergência foi satisfeito

PARE com a solução aproximada (x^{k+1}, λ^{k+1}) ;

FIM

FIM

A estrutura da Programação Quadrática Seqüencial Local para problemas não-lineares com restrições de igualdade é facilmente estendida para os problemas com restrições de desigualdade do tipo

$$A^k p + g(x) \leq 0.$$

E a cada passo deve-se resolver o problema quadrático

$$(SQI) \quad \begin{array}{ll} \text{minimizar} & \frac{1}{2} p^t W^k p + \nabla^t f^k p \\ \text{sujeito a} & A^k p + g(x) \leq 0 \end{array}$$

Pode-se usar o algoritmo, do método de Programação Quadrática Seqüencial, descrito acima para resolver (SQI), mas com algumas modificações: O passo p^k e a nova estimativa do multiplicador λ^{k+1} são definidos através da solução e dos multiplicadores de Lagrange correspondente ao problema (SQI).

Como vimos no capítulo sobre Programação Quadrática (capítulo 3), a região viável de um problema quadrático com restrições de igualdade como o (SQ), reescrito a seguir

$$(SQ) \quad \begin{array}{ll} \text{minimizar} & \frac{1}{2}p^t W^k p + \nabla^t f^k p \\ \text{sujeito a} & A^k p + h(x^k) = 0 \end{array}$$

pode ser vazia, ou seja, quando não existe intersecções entre as restrições. Existem várias maneiras de contornar esta dificuldade. Em todas elas, o problema deve ser modificado de maneira tal que, por um lado, o novo subproblema tenha solução e, por outro lado, que a nova solução coincida com a solução do subproblema (SQ) nos casos em que aquela exista.

Persiste, porém, a possibilidade de que a função objetivo de (SQ) seja ilimitada inferiormente no seu conjunto de viabilidade. A modificação, nesse caso, vem do nosso objetivo final, que é resolver o problema de programação não linear com restrição de igualdade, e para isso nos baseamos em que perto de x^k , o subproblema (SQ) é parecido com o problema original. Para expressar essa necessidade acrescenta-se em (SQ) uma restrição adicional, essa restrição é chamada de **região de confiança**[11]. Com essa modificação tem-se o método de programação quadrática Seqüencial associado ao método de região de confiança.

A seguir será discutido o método Região de Confiança, e na próxima seção será apresentado o método de Programação Quadrática Seqüencial com Região de Confiança.

4.2.1 Região de Confiança

Região de Confiança é um método para resolver problemas de otimização irrestrita, a cada iteração esse método encontra simultaneamente a direção e o comprimento do passo a ser dado, diferentemente dos métodos vistos na seção referente a minimização irrestrita (1.4.1), onde primeiro era escolhido uma direção e depois calculado o comprimento do passo.

O método define uma região em torno do ponto x^k corrente, na qual é definido um modelo quadrático da função objetivo. Então se calcula o minimizador aproximado do modelo dentro dessa região. Assim, a direção e o comprimento são calculados simultaneamente dentro de uma região de confiança. Se o ponto não é aceito, reduz-se o tamanho da região e encontra-se um novo minimizador.

Considere o problema irrestrito

$$(PE) \quad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & x \in \mathbb{R}^n \end{array}$$

sendo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é de classe C^2 .

Um modelo quadrático de f em torno do ponto corrente x^k é:

$$m_k(p) = f_k + \nabla^t f_k p + \frac{1}{2} p^t B_k p, \quad (4.7)$$

onde $p = x - x^k$ e $B_k \in \mathbb{R}^{n \times n}$ é uma matriz simétrica, podendo ser uma aproximação da hessiana $\nabla^2 f(x^k)$ ou qualquer outra matriz que satisfaça

$$\|B_k\| \leq \beta,$$

para alguma constante $\beta > 0$, independentemente do índice k .

Como o modelo quadrático (4.7) deixa de ser confiável à medida que x se afasta de x^k , pode-se confiar em aproximar a função objetivo apenas numa vizinhança de x^k , ou seja, na região:

$$\{p \in \mathbb{R}^n \mid \|p\| \leq \Delta_k\}, \quad (4.8)$$

onde $\Delta_k > 0$ é chamado de raio da região de confiança e $\|\cdot\|$ é uma norma qualquer em \mathbb{R}^n .

Assim o modelo de região de confiança próximo ao ponto x^k , para resolver o problema (PE) é da forma:

$$\begin{aligned} &\text{minimizar} && m_k(p) \\ &\text{sujeito a} && \|p\| \leq \Delta_k \end{aligned} \quad (4.9)$$

sendo $m_k(\cdot)$ dada na relação (4.7)

Dessa forma, a solução $p^k \in \mathbb{R}^n$ para o problema (4.9) seria uma boa aproximação para o minimizador de f na região (4.8). O passo de região de confiança p^k depende do raio Δ_k da região e pode mudar de direção quando o raio é alterado.

Dado um passo p^k , definimos a redução **predita** produzida pelo passo p^k , como:

$$pred = m_k(0) - m_k(p^k)$$

e a redução **verdadeira** como:

$$ared = f(x^k) - f(x^k + p^k)$$

Para avaliar se o modelo está bom ou ruim, se é necessário alterar o raio da região de confiança ou não, analisa-se a taxa de variação que é a razão entre a

variação na função objetivo e a sua aproximação quadrática:

$$\rho_k = \frac{f(x^k) - f(x^k + p^k)}{m_k(0) - m_k(p^k)}. \quad (4.10)$$

Dessa forma, podem-se tirar algumas conclusões:

- Se ρ_k é negativo, o novo valor $f(x^k + p^k)$ é maior que o anterior $f(x^k)$ e o passo p^k deve ser rejeitado, pois como o passo p^k é obtido minimizando-se o modelo m_k sobre uma região que inclui o ponto x^k , a redução predita será sempre não negativa.
- Se ρ_k está próximo de 1, tem-se que a redução verdadeira está próxima da redução do modelo, assim pode-se aumentar a região de confiança na próxima iteração.
- Se ρ_k é positivo, mas não próximo de 1, mantém-se o raio da região, mas se ρ_k está próximo de zero ou é negativo diminui-se a região.

O algoritmo a seguir descreve o processo do método Região de Confiança, que é baseado em [15].

Algoritmo 4.2 (*Região de Confiança*)

Dado $\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta})$ e $\eta \in [0, \frac{1}{4})$

PARA $k = 0, 1, 2, \dots$

Obter p^k aproximadamente resolvendo (4.9)

Calcular ρ_k por (4.10)

Se $\rho_k < \frac{1}{4}$

$$\Delta_{k+1} = \frac{1}{4}\Delta_k$$

Senão

Se $\rho_k > \frac{3}{4}$ e $\|p^k\| = \Delta_k$

$$\Delta_{k+1} = \min\{2\Delta_k, \bar{\Delta}\}$$

Senão

$$\Delta_{k+1} = \Delta_k$$

Se $\rho_k > \eta$

$$x^{k+1} = x^k + p^k$$

Senão

$$x^{k+1} = x^k$$

$$k = k + 1$$

FIM

No algoritmo acima, $\bar{\Delta}$ é um limitante superior para o raio da região de confiança, assim, para todo k tem-se $\Delta_k \leq \bar{\Delta}$.

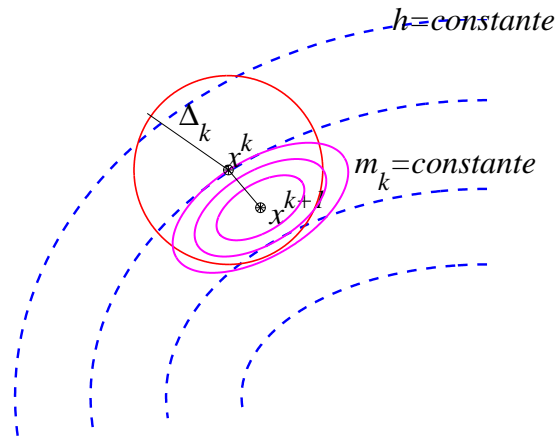


Figura 4.1: Um passo de Região de Confiança

A figura (4.1) mostra uma iteração do algoritmo de região de confiança, as curvas de nível da função objetivo do problema original estão representadas por linhas tracejadas. As linhas cheias representam as curvas de nível do modelo em x^k . O ponto x^{k+1} é a minimização do modelo na região de confiança centrada em x^k .

4.3 Programação Quadrática Seqüencial com Região de Confiança

Acrescentando a restrição de região de confiança ao problema quadrático (SQ), obtém-se o seguinte problema

$$\begin{aligned} \text{minimizar} \quad & \frac{1}{2}p^t W^k p + \nabla^t f^k p \\ \text{sujeito a} \quad & A^k p + h(x^k) = 0 \\ & \|p\| \leq \Delta^k \end{aligned} \tag{4.11}$$

Com a inclusão dessa restrição aumenta-se a dificuldade para resolver o subproblema quadrático (SQ), além disso, o subproblema (4.11), pode nem sempre ter solução, porque as restrições podem não ter nenhum ponto em comum. Para satisfazer as restrições lineares e a restrição de bola, não basta somente aumentar o raio da região de confiança, pois assim estaria assumindo-se pontos distantes da solução e prejudicando o desenvolvimento do algoritmo.

Dentre os métodos que resolvem o subproblema (4.11) (ver em [15]), destaca-se o que encontra a solução p^k do subproblema (4.11) em duas fases: cada iteração desse método trabalha em uma região centrada na iteração atual x^k , e é composta por um passo de viabilidade (passo normal), e um passo de otimalidade (passo tangencial). O passo tangencial deve seguir uma direção no espaço nulo das restrições no ponto x^k .

Na primeira fase que é a do passo da viabilidade procura-se satisfazer as restrições lineares dentro da região de confiança, para isso ignora-se a função objetivo e resolve-se o subproblema normal:

$$\begin{aligned} \text{minimizar} \quad & \|A^k \nu + h(x^k)\|_2 \\ \text{sujeito a} \quad & \|\nu\|_2 \leq \xi \Delta^k \end{aligned} \tag{4.12}$$

onde ξ pode ser qualquer valor no intervalo $(0, 1)$.

Observe que o subproblema normal (4.12) é a minimização de uma quadrática na bola. Para isso pode-se usar os métodos de Steihaug ou Dogleg, que podem ser encontrados, por exemplo, em [15].

Na literatura chama-se de passo normal a solução ν^k deste subproblema.

Para a segunda fase retorna-se a função objetivo e exige-se que o “passo total” p^k (que será visto na próxima seção) reduza o seu valor, satisfazendo as restrições lineares, assim com o passo normal ν^k resulta um novo subproblema:

$$\begin{aligned} &\text{minimizar} && \frac{1}{2}p^t W^k p + \nabla^t f^k p \\ &\text{sujeito a} && A^k p = A^k \nu^k \\ &&& \|p\|_2 \leq \Delta^k \end{aligned} \tag{4.13}$$

Como o ponto $p = \nu^k$ satisfaz as restrições do subproblema (4.13), pode-se dizer que elas são consistentes.

O subproblema (4.13) difere do subproblema do método de região de confiança da seção anterior somente pela restrição de igualdade. Portanto deseja-se eliminar esta restrição, o que resulta num subproblema do tipo (4.12), que consiste em minimizar uma quadrática na bola e, para o qual conhecemos vários métodos de resolução, como os que já foram citados para encontrar o passo normal. Segundo [15], conseguimos isto fazendo

$$p^k = \nu^k + t,$$

mas como

$$A^k \nu^k + h(x^k) = 0$$

então

$$A^k p^k + h(x^k) = A^k \nu^k + h(x^k)$$

donde

$$\begin{aligned} A^k p^k &= A^k \nu^k \\ A^k (\nu^k + t) &= A^k \nu^k \\ A^k t &= 0, \end{aligned}$$

$t \in N(A^k)$ o que implica em $t = Z^k u^k$, onde as colunas da matriz Z^k formam uma base para o espaço nulo de A^k e $u^k \in \mathbb{R}^{n-m}$. Assim

$$p^k = \nu^k + Z^k u^k.$$

Substituindo $p^k = \nu^k + Z^k u^k$ na função objetivo do subproblema (4.13), obtém-se o seguinte subproblema “tangencial” na variável ν :

$$\begin{aligned} &\text{minimizar} && \frac{1}{2}(\nu^k + Z^k u)^t W^k (\nu^k + Z^k u) + \nabla^t f^k (\nu^k + Z^k u) \\ &\text{sujeito a} && A^k (\nu^k + Z^k u) = A^k \nu^k \\ &&& \|(\nu^k + Z^k u)\|_2 \leq \Delta^k \end{aligned} \tag{4.14}$$

Reescrevendo a função objetivo do subproblema (4.14), desprezando os termos constantes que são independentes de u , pois eles não alteram a solução, tem-se que

$$\frac{1}{2}(\nu^k + Z^k u)^t W^k (\nu^k + Z^k u) + \nabla^t f^k (\nu^k + Z^k u) =$$

$$\begin{aligned}
 &= \frac{1}{2}(\nu^k)^t W^k \nu^k + \frac{1}{2}(Z^k u)^t W^k \nu^k + \frac{1}{2}(\nu^k)^t W^k (Z^k u) + \frac{1}{2}(Z^k u)^t W^k (Z^k u) + \nabla^t f^k \nu^k + \nabla^t f^k Z^k u \\
 &\quad (\nu^k)^t W^k (Z^k u) + \frac{1}{2}(Z^k u)^t W^k Z^k u + \nabla^t f^k Z^k u = \\
 &\quad = \frac{1}{2}u^t ((Z^k)^t W^k Z^k) u + (\nabla f^k + W^k \nu^k)^t Z^k u,
 \end{aligned}$$

uma função quadrática.

Como $Z^k u \in N(A^k)$ e $\nu^k \in I(A^k)$, tem-se que $(Z^k u)^t \nu^k = 0$, isto é, $Z^k u$ e ν^k são ortogonais.

Elevando ao quadrado em ambos os lados da restrição de desigualdade em (4.14), tem-se:

$$\begin{aligned}
 \|\nu^k + Z^k u\|_2^2 &\leq (\Delta^k)^2 \\
 (\nu^k + Z^k u)^t (\nu^k + Z^k u) &\leq (\Delta^k)^2 \\
 (\nu^k)^t \nu^k + u^t Z^k \nu^k + (\nu^k)^t Z^k u + (Z^k u)^t Z^k u &\leq (\Delta^k)^2,
 \end{aligned}$$

como Z^k e ν^k são ortogonais,

$$\begin{aligned}
 \|\nu^k\|_2^2 + \|Z^k u\|_2^2 &\leq (\Delta^k)^2 \\
 \|Z^k u\|_2^2 &\leq \sqrt{(\Delta^k)^2 - \|\nu^k\|_2^2}.
 \end{aligned}$$

Com isto, o subproblema “tangencial” recai em minimizar uma quadrática com restrição em uma bola, e pode ser escrito como:

$$\begin{aligned}
 \text{minimizar} \quad & \frac{1}{2}u^t (Z^k W^k Z^k) u + (\nabla f^k + W^k \nu^k)^t Z^k u \\
 \text{sujeito a} \quad & \|Z^k u\|_2 \leq \sqrt{(\Delta^k)^2 - \|\nu^k\|_2^2}
 \end{aligned} \tag{4.15}$$

A solução u deste subproblema é o “passo tangencial”.

Dentre os métodos utilizados para resolver (4.15), destacam-se os já citados para resolver (4.12), ou seja, o Dogleg quando $(Z^k)^t W^k Z^k$ é definida positiva, e o Steihaug.

Uma vez calculado o passo normal e o tangencial, tem-se o passo total:

$$p^k = \nu^k + Z^k u^k.$$

Na figura (4.2) são ilustrados os passos normal, tangencial e o total para um problema com duas variáveis e uma restrição de igualdade não linear. Onde as linhas tracejadas são as curvas de nível da função objetivo, cujo minimizador é x^* . A restrição de igualdade está representada pela linha cheia e a circunferência com centro em x^k é a restrição da região de confiança. O espaço tangente das

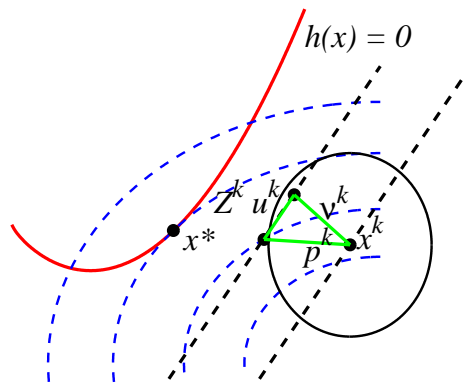


Figura 4.2: Passo do método Região de Confiança

restrições é representado pela linha pontilhada que passa pelo ponto x^k (nesse caso Z^k tem apenas uma coluna), um múltiplo do espaço tangente passando por $x^k + \nu^k$, define um conjunto, no qual deve estar o passo tangencial $Z^k u^k$, o passo normal é perpendicular ao espaço tangencial. Assim, o passo total $p^k = \nu^k + Z^k u^k$ satisfaz a restrição de região de confiança.

Capítulo 5

Implementação e Testes Numéricos

Neste capítulo será tratada a implementação dos algoritmos de Lagrangeano Aumentado com Penalidade Quadrática e a análise dos resultados obtidos através dos testes computacionais realizados.

Foi implementado em MATLAB, versão 6.0 R12 para Linux, o método de Lagrangeano Aumentado usando as penalidades tipo 1 (2.1) e a tipo 2 (2.2), associadas a penalidade quadrática (3.15).

Os problemas testados são da coleção CUTE: Constrained and Unconstrained Testing Enviromet [9], na versão small com interface para o MATLAB. Por ser um conjunto com milhares de problemas, foram escolhidos aqueles que satisfazem as condições desta dissertação, isto é, problemas com restrições de desigualdade e sem limites nas variáveis.

5.1 Implementação

Nesta seção será tratada a questão referente à implementação dos algoritmos de Lagrangeano Aumentado com Penalidade Quadrática.

Propõe-se aplicar o algoritmo de Lagrangeano Aumentado usando a penalidade tipo 1(2.1)

$$(y, \mu) \in \mathbb{R} \times \mathbb{R}_{++} \mapsto p(y, \mu) = \theta(\mu y) \in \mathbb{R}$$

associada a penalidade quadrática (3.15),

$$y \in \mathbb{R} \mapsto \theta(y) = \frac{1}{2} y^2 + y,$$

aos problemas quadráticos gerados pelo método de programação quadrática sequencial.

Considere o problema a ser resolvido

$$(P) \quad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & g_i(x) \leq 0, \quad i = 1, 2, \dots, m \\ & x \in \mathbb{R}^n \end{array}$$

onde as funções f e g_i para $i = 1, 2, \dots, m$ são contínuas e possuem derivadas, até ordem necessária.

Aplicando-se o método de Programação Quadrática Sequencial à (P) geram-se subproblemas quadráticos, como descrito no capítulo 4

$$\begin{array}{ll} \text{minimizar} & \frac{1}{2}d^t W^k d + \nabla^t f^k d \\ \text{sujeito a} & A^k d + g^k \leq 0, \end{array} \quad (5.1)$$

sendo, a função objetivo a aproximação quadrática da função lagrangeano associada ao problema (P), e as restrições são aproximações lineares das restrições do problema (P).

Para isso usa-se o algoritmo de programação quadrática sequencial [15].

Algoritmo 5.1 (*Programação Quadrática Sequencial*)

Escolha o par inicial (x^0, λ^0) ;

Para $k = 0, 1, 2, \dots$

Avaliar $f(x^k)$, $\nabla f(x^k)$, W^k , $g(x^k)$ e A^k ;

Resolver (5.1) para obter d^k e μ^k ;

$$x^{k+1} \leftarrow x^k + d^k;$$

$$\lambda^{k+1} \leftarrow \mu^k;$$

Se teste de convergência foi satisfeito

Pare com a solução aproximada (x^{k+1}, λ^{k+1}) ;

Fim

$k=k+1$

Fim

Para resolver (5.1) foi implementado o método de Lagrangeano Aumentado com penalidade p tipo 1 e 2 e θ quadrática.

Algoritmo 5.2 (*Lagrangeano Aumentado com Penalidade Quadrática para Problemas Quadráticos*)

Dados $\mu^0 > 0$, $\gamma > 1$, $\alpha > 1$ e $r^0 > 0$,

Faça $k = 0$

Repita

Encontre

$$x^+ \in \operatorname{argmin} \left\{ \frac{1}{2} d^t W^k d + \nabla^t f^k d + r^k \sum_{i=1}^m p \left(\frac{a_i^t d + g^k}{r^k}, \mu_i \right) : d \in \mathbb{R}^n \right\}$$

Atualize

$$\bar{\mu}_i = p' \left(\frac{a_i^t d + g^k}{r^k}, \mu_i^k \right) \text{ para } i = 1, 2, \dots, m$$

Se $\bar{\mu}_i \geq 0$

$$x^{k+1} = x^+$$

$$\mu^{k+1} = \bar{\mu}$$

$$r^{k+1} = \frac{r^k}{\alpha}$$

Se não

$$r^{k+1} = \gamma r^k$$

$k = k + 1$

Cada iteração do algoritmo (5.2) consiste em formar subproblemas irrestritos, que podem ser resolvidos por qualquer método de minimização irrestrita. Para resolver

$$x^+ \in \operatorname{argmin} \left\{ \frac{1}{2} d^t W^k d + \nabla^t f^k d + r^k \sum_{i=1}^m p \left(\frac{a_i^t d + g^k}{r^k}, \mu_i \right) : d \in \mathbb{R}^n \right\}$$

foi implementado o método Região de Confiança ver [10], e atualiza-se o parâmetro de penalidade da seguinte forma

- Se $\bar{\mu} \geq 0$, aceita-se o ponto corrente e o multiplicador de Lagrange para a próxima iteração, reduz-se o parâmetro de penalidade r (usando por exemplo $\alpha = 2$);
- Se não, mantêm-se o ponto, aumenta-se o valor do parâmetro de penalidade r , para tornar $\mu \geq 0$.

5.2 Resultados Numéricos

Testamos trinta e nove problemas que estão apresentados na tabela (5.1). A primeira coluna da tabela é formada pelos nomes dos problemas apresentados na coleção CUTE. As demais colunas são: número de variáveis, número de restrições, tipo das restrições, e tipo de função objetivo para cada problema.

Tabela 5.1: Problemas da coleção CUTE

Problema	# variáveis	# restrições	Tipo das restrições	tipo da objetivo
CB2	3	3	3 não lineares	linear
CB3	3	3	3 não lineares	linear
GIGOMEZ1	3	3	2 lineares 1 não linear	linear
HS10	2	1	não linear	linear
HS11	2	1	não linear	não linear
HS113	10	8	5 não lineares 3 lineares	não linear
HS12	2	1	não linear	não linear
HS22	2	2	1 não linear 1 linear	não linear
HS268	5	5	5 lineares	não linear
HS29	3	1	não linear	não linear
HS43	4	3	3 não linear	não linear
HS88	2	1	1 não linear	não linear
HS90	4	1	1 não linear	não linear
LISWET1	103	100	100 lineares	não linear
LISWET2	103	100	100 lineares	não linear
LISWET3	103	100	100 lineares	não linear
LISWET4	103	100	100 lineares	não linear
LISWET5	103	100	100 lineares	não linear
LISWET6	103	100	100 lineares	não linear
LISWET10	103	100	100 lineares	não linear
MINMAXRB	3	4	2 não lineares 2 lineares	linear
OET1	3	6	6 lineares	não linear
OET2	3	202	202 não lineares	linear
OET3	4	6	6 lineares	linear
OET3A	4	6	6 lineares	linear
POWELL20	10	10	10 lineares	não linear
PT	2	501	501 lineares	linear
PT	2	3	3 lineares	linear
PT	2	101	101 lineares	linear
SIAPOW1	2	20	20 lineares	linear
SIAPOW1	2	100	100 lineares	linear
SIAPOW1	2	500	500 lineares	linear
SIAPOW1M	2	20	20 lineares	linear
SIAPOW1M	2	100	100 lineares	linear
SIAPOW1M	2	500	500 lineares	linear
SIAPOW2	2	20	20 lineares	linear
SIAPOW2M	2	20	20 lineares	linear

continua ...

Tabela 5.1: Problemas da coleção CUTE (continuação)

Problema	# variáveis	# restrições	Tipo das restrições	tipo da objetivo
SIAPOW2M	2	100	100 lineares	linear
SIAPOW2M	2	500	500 lineares	linear

Os resultados dos testes computacionais estão apresentados nas tabelas (5.2), cujos dados estão dispostos na primeira linha, da seguinte maneira:

- PROB: Nome do problema definido pela coleção CUTE;
- PEN: Tipo da penalidade - **tipo 1** ou **tipo 2**;
- $W^0 > 0$: 0 se a matriz hessiana da função Lagrangeano no ponto x^0 for definida positiva e 1 caso contrário;
- $W^k > 0$: 0 se a matriz hessiana da função Lagrangeano em qualquer iteração for definida positiva e 1 caso contrário;
- Aum r : Número de vezes que o parâmetro de penalidade r aumentou
- it.ext : Quantidade de subproblemas quadráticos que foram gerados no algoritmo de Programação Quadrática Seqüencial;
- $\sum(it.int)$: Somatório de todas iterações necessárias para resolver cada problema quadrático gerado pelo algoritmo;
- max(it.int) : Número máximo de iterações necessárias para resolver cada problema quadrático, gerado pelo algoritmo;

Para exemplificar considere o problema "LISWET1". As matrizes hessiana da função Lagrangeano associado ao problema "LISWET1" no ponto x^0 e em todas as iterações k são definida positiva.

A penalidade tipo 1 não aumentou o parâmetro de penalidade, realizou uma iteração no algoritmo de programação quadrática seqüencial (it.ext = 1) e resolveu a função lagrangeano aumentado com penalidade quadrática em 8 iterações ($\sum(it.int) = 8$ e max(it.int) = 8), enquanto a penalidade tipo 2 aumentou 4 vezes o parâmetro de penalidade , realizou uma iteração no algoritmo de programação quadrática seqüencial (it.ext = 1) e resolveu a função lagrangeano aumentado com penalidade quadrática em 11 iterações ($\sum(it.int) = 11$ e max(it.int) = 11).

Tabela 5.2: Comparação entre as penalidades

PROB	PEN	$W^0 > 0$	$W^k > 0$	Aum r	it.ext	$\sum(\text{it.int})$	max(it.int)
CB2	tipo 1	1	1	0	13	125	11
	tipo 2			445	41	451	11
CB3	tipo 1	1	1	0	8	69	11
	tipo 2			445	41	451	11
GIGCOMIZ1	tipo 1	1	1	0	7	43	9
	tipo 2			6	7	37	11
HS10	tipo 1	0	0	0	11	34	5
	tipo 2			0	9	28	4
HS11	tipo 1	0	0	0	6	16	4
	tipo 2			0	6	17	4
HS113	tipo 1	0	0	3	6	44	11
	tipo 2			447	41	451	11
HS12	tipo 1	0	0	2	6	36	11
	tipo 2			6	6	28	11
HS22	tipo 1	0	0	0	5	20	6
	tipo 2			0	7	18	5
HS268	tipo 1	0	0	1	2	22	11
	tipo 2			441	41	451	11
HS29	tipo 1	0	1	8	7	22	11
	tipo 2			9	7	22	11
HS43	tipo 1	0	0	3	9	46	11
	tipo 2			45	13	88	11
HS88	tipo 1	0	0	0	20	56	5
	tipo 2			0	16	117	11
HS90	tipo 1	0	0	0	3	22	11
	tipo 2			14	3	17	11
LISWET1	tipo 1	0	0	0	1	8	8
	tipo 2			4	1	11	11
LISWET2	tipo 1	0	0	0	2	10	9
	tipo 2			4	1	11	11
LISWET3	tipo 1	0	0	0	1	11	11
	tipo 2			3	1	11	11
LISWET4	tipo 1	0	0	0	1	11	11
	tipo 2			3	1	11	11
LISWET5	tipo 1	0	0	0	2	10	9
	tipo 2			4	1	11	11
LISWET6	tipo 1	0	0	0	2	10	9
	tipo 2			4	1	11	11
LISWET10	tipo 1	0	0	0	2	10	9
	tipo 2			4	1	11	11

Tabela 5.2: Comparação entre as penalidades (continuação).

PROB	PEN	$W^0 > 0$	$W^k > 0$	Aum r	it.ext	$\sum(\text{it.int})$	max(it.int)
MINMAXRB	tipo 1	1	1	0	2	3	2
	tipo 2			0	2	3	2
OET1	tipo 1	1	1	2	1	11	11
	tipo 2			446	41	451	11
OET2	tipo 1	1	1	3	3	33	11
	tipo 2			60	15	165	11
OET3	tipo 1	1	1	0	1	2	2
	tipo 2			0	1	2	2
OET3A	tipo 1	1	1	1	1	11	11
	tipo 2			3	1	11	11
POWELL20	tipo 1	0	0	0	1	11	11
	tipo 2			6	3	28	11
PT	tipo 1	1	1	4	2	22	11
	tipo 2			4	1	11	11
PT1	tipo 1	1	1	0	1	10	10
	tipo 2			5	1	11	11
PT2	tipo 1	1	1	5	1	11	11
	tipo 2			5	1	11	11
SIPOW1	tipo 1	1	1	1	2	22	11
	tipo 2			445	41	451	11
SIPOW11	tipo 1	1	1	12	2	22	11
	tipo 2			445	41	451	11
SIPOW12	tipo 1	1	1	0	1	11	11
	tipo 2			445	41	451	11
SIPOW1M	tipo 1	1	1	1	1	11	11
	tipo 2			144	24	264	11
SIPOW1MA	tipo 1	1	1	0	2	22	11
	tipo 2			445	41	451	11
SIPOW1MB	tipo 1	1	1	3	1	11	11
	tipo 2			35	5	55	11
SIPOW2	tipo 1	1	1	2	1	11	11
	tipo 2			48	8	88	11
SIPOW2M	tipo 1	1	1	2	1	11	11
	tipo 2			108	18	198	11
SIPOW2MA	tipo 1	1	1	4	2	22	11
	tipo 2			447	41	451	11
SIPOW2MB	tipo 1	1	1	7	2	22	11
	tipo 2			112	16	176	11

O principal objetivo dos testes numéricos é avaliar o desempenho do algoritmo de Lagrangeano Aumentado com penalidade quadrática, em problemas de programação quadrática. Para simplificar a análise dos resultados, construímos outras tabelas menores fornecendo informações específicas sobre alguns dados dos testes numéricos.

Tabela 5.3: Resumo de Iterações

PENALIDADE	it.ext	$\sum(\text{it.int})$	max(it.int)
tipo 1	19	26	10
tipo 2	7	6	2
iguais	13	7	27

Começamos pela tabela (5.3), em que cada número apresentado é referente a quantidade de vezes que cada penalidade executou menos iterações, por exemplo, em 19 problemas a penalidade tipo 1 realizou menos iterações externas em relação a penalidade tipo 2, que em 7 problemas realizou menos iterações externas que a penalidade tipo 1. A linha chamada “iguais” é o número de vezes que a penalidade tipo 1 e a 2 realizou o mesmo número de iterações.

Tabela 5.4: Aumento do Parâmetro de Penalidade

PENALIDADE	Aum r
tipo 1	18
tipo 2	33

O aumento do parâmetro de penalidade, está relacionado com a diminuição do raio da Região de Confiança. Portanto, quando uma metodologia está aumentando o parâmetro de penalidade ela está gerando problemas internos mais simples, mas os passos serão curtos e a convergência será lenta [12].

A tabela (5.4), refere-se a quantidade de problemas que cada metodologia aumentou o parâmetro de penalidade. A tipo 2 em 33 problemas aumentou o parâmetro de penalidade, enquanto a tipo 1 aumentou o parâmetro de penalidade em 18 problemas. Isto da indicação que a metodologia utilizando a penalidade tipo 1 pode estar gerando subproblemas mais simples do que a metodologia utilizando a penalidade tipo 2.

Como podemos notar na terceira e quarta colunas da tabela (5.2), alguns problemas quadráticos gerados não são convexos no ponto inicial ou em algum subproblema gerado nas iterações posteriores (mesmo tendo como principal hipótese, o problema ser convexo, testamos alguns problemas não convexos), dentre esses, o algoritmo chegou na solução em 22 problemas não convexos. Como veremos na tabela 5.5) a seguir a penalidade tipo 1 realizou menos iterações do que a penalidade tipo 2 nos problemas não convexos.

Tabela 5.5: Resumo de Iterações em Problemas Não Convexos

PENALIDADE	it.ext	$\sum(\text{it.int})$	max(it.int)
tipo 1	14	15	4
tipo 2	7	2	0
iguais	1	5	18

Pelos testes que fizemos e os resultados mostrados nas tabelas desta seção, vimos que o algoritmo de Lagrangeano Aumentado com penalidade quadrática aplicado em problemas de programação quadrática apresentou um desempenho superior, em relação ao mesmo algoritmo com a penalidade tipo 2, a qual é a mais utilizada na literatura.

Também podemos notar que a penalidade tipo 1 gerou subproblemas ir-restritos mais simples, pois executou menos iterações internas conforme foi mostrado nas tabelas deste capítulo.

Conclusão

Neste trabalho propomos a aplicação do método de Lagrangeano Aumentado com penalidade quadrática para resolver subproblemas quadráticos gerados pelo método de programação quadrática seqüencial e mostramos que a função lagrangeano aumentado do subproblema possui boas propriedades. Assim, no caso em que a função Lagrangeano Aumentado associado ao problema original é estritamente convexa (convexa), a matriz hessiana da função Lagrangeano Aumentado é definida positiva (semidefinida positiva), logo satisfaz a condição de otimalidade suficiente de segunda ordem.

Com esse resultado, realizamos testes numéricos tendo como principal objetivo avaliar o desempenho do algoritmo de Lagrangeano Aumentado com penalidade quadrática, em problemas de programação quadrática convexos. Através da análise das tabelas dadas no Capítulo 5, notamos que a penalidade tipo 1 necessitou de menos subproblemas quadráticos para resolver o problema original, pois realizou menos iterações externas. Além disto, os subproblemas irrestritos gerados pelo algoritmo de Lagrangeano Aumentado com penalidade tipo 1 são mais simples de resolver, pois executou menos iterações internas em relação a penalidade tipo 2. Vimos nos testes que a metodologia proposta se sobressaiu em relação à mais utilizada na literatura.

Quando o problema é convexo, a proposta do método de Lagrangeano Aumentado com penalidade quadrática se mostrou bastante promissora. Seria conveniente uma continuidade da pesquisa para o caso da aplicação dessa metodologia para problemas não convexos. Deixamos está proposta como continuidade para trabalhos futuros. Além disso, outros testes computacionais podem ser realizados. Por exemplo, podem ser realizados testes com classes de problemas convexos conhecidos, como os problemas de finanças que possuem as características dos problemas quadráticos aqui trabalhados.

Referências Bibliográficas

- [1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming Theory and Algorithms*. John Wiley, New York, 2nd edition, 1993.
- [2] D. P. Bertsekas. *D. Constrained Optimization and Lagrange Multipliers*. Academic Press, New York, 1992.
- [3] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C.A. Sagastizábal. *Numerical Optimization: Theoretical and Practical Aspects*. Springer Verlag, Berlin, 2002.
- [4] R. Fletcher. *Practical Methods of Optimization*. John Wiley, New York, 2nd edition, 1990.
- [5] A. Friedlander. *Elementos de Programação Não-Linear*. Campinas, editora unicamp edition, 1994.
- [6] H. Fritzsche. *Programação Não-Linear Análise e Métodos*. São Paulo, edgard blücher edition, 1977.
- [7] C. C. Gonzaga. Um curso de programação não linear. Notas de aula, 2004.
- [8] J-B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms I*. Springer-Verlag, New York, 1993.
- [9] A. R. Conn I. Bongartz, N. I. M. Gould and Ph. L. Toint. CUTE: Constrained and unconstrained testing environment. *ACM Transactions on Mathematical Software*, 21:123–160, 1995.
- [10] L. F. Jussiani. Desempenho de método de lagrangeano aumentado com penalidade quadrática. Master's thesis, Universidade Federal do Paraná, 2004.
- [11] J. M. Martínez and S. A. Santos. Métodos computacionais de otimização. 20.⁰ Colóquio Brasileiro de Matemática - IMPA, julho 1995.
- [12] L. C. Matioli. *Uma Nova Metodologia Para Construção de Funções Penalização Para Algoritmos de Lagrangeano Aumentado*. PhD thesis, UFSC - Universidade Federal de Santa Catarina, Santa Catarina, Brasil, 2001.

- [13] M. Minoux. *Mathematical Programming*. John Wiley and Sons Ltd edition, 1986.
- [14] A. M. Mota. Convergência de algoritmos para programação não linear. Master's thesis, Universidade Federal do Paraná, 2005.
- [15] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag, 1999.
- [16] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control Optimization*, 14(5):877–898, 1976.