

Universidade Federal de Juiz de Fora

Programa de Pós-Graduação em Modelagem Computacional

**MÉTODOS SUBGRADIENTES EM OTIMIZAÇÃO CONVEXA
NÃO DIFERENCIÁVEL**

Por

Théssera Christine Araújo de Souza

Juiz de Fora, MG - BRASIL

Agosto de 2008

Souza, Théssea Christine Araújo de.

Métodos subgradientes em otimização convexa não diferenciável / Théssea Christine Araújo de Souza ; orientador: Professor Wilhelm Passarella Freire. – 2008.

104f.

Dissertação (Mestrado em Modelagem Computacional) – Faculdade de Engenharia, Universidade Federal de Juiz de Fora.

1. Matemática computacional. I. Freire, Wilhelm Passarella , orientador. II. Título.

CDU 519.67

MÉTODOS SUBGRADIENTES EM OTIMIZAÇÃO CONVEXA NÃO DIFERENCIÁVEL

Théssera Christine Araújo de Souza

DISSERTAÇÃO SUBMETIDA AO PROGRAMA DE PÓS-GRADUAÇÃO EM
MODELAGEM COMPUTACIONAL DA UNIVERSIDADE FEDERAL DE JUIZ DE
FORA COMO PARTE DOS REQUISITOS NECESSÁRIOS À OBTENÇÃO DO GRAU
DE MESTRE EM CIÊNCIAS (M.SC.) EM MODELAGEM COMPUTACIONAL.

Aprovada por:

Prof: Wilhelm Passarella Freire, D.Sc.
(Orientador)

Prof: José Herskovits Norman, D.Ing.

Prof: Sandro Rodrigues Mazorche, .Sc.

JUÍZ DE FORA, MG - BRASIL

AGOSTO DE 2008

AGRADECIMENTOS

Agradeço a Deus por essa dádiva e por sua infinita misericórdia que colocou em meu caminho uma infinidade de pessoas que tornaram possível essa conquista.

A toda minha família pelo carinho, amor e apoio incondicional sem o qual eu não teria chegado até aqui. Em especial, agradeço meu pai, minha mãe e minha irmã pelo muito que fizeram por mim sem jamais medir sacrifícios.

Ao Fernando, por seu incentivo e cumplicidade em todos os momentos;

Às Minhas amigas Andréia, Cristiane, Franciane, Isabel, Kelly, e Marcelle que me incentivaram, apoiaram e me ajudaram a conquistar esse sonho.

Um agradecimento especial ao meu orientador Wihelm Passarella Freire pelo incentivo, apoio, paciência e perseverança durante a tarefa de orientação.

Aos Professores Sandro Rodrigues Mazorche e José Herskovits Normam por terem participado da Banca e pelas valiosas sugestões.

Por fim, meus agradecimentos a todos aqueles que, direta ou indiretamente, contribuíram para o desenvolvimento desse trabalho.

Resumo da Dissertação apresentada à UFJF como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

**MÉTODOS SUBGRADIENTES EM OTIMIZAÇÃO CONVEXA NÃO
DIFERENCIÁVEL**

Théssera Christine Araújo de Souza

Agosto / 2008

Orientador: Wilhelm Passarella Freire

Este trabalho tem por finalidade descrever o Estado da Arte acerca de Métodos Subgradientes para otimização de funções convexas não diferenciáveis. Apresenta-se inicialmente um histórico desses métodos, conceitos básicos sobre otimização diferenciável, necessários para o entendimento de certas noções importantes referentes à problemas não diferenciáveis, bem como esses problemas e suas características próprias. Posteriormente, apresenta-se uma breve introdução aos métodos não diferenciáveis para, então dedicar-se ao objetivo principal do trabalho que são os Métodos Subgradientes, suas extensões e trabalhos recentes. Finaliza-se a Dissertação com a apresentação de algumas aplicações, seus resultados e conclusões.

Abstract of Dissertation presented to UFJF as a partial fulfillment of the requirements for the degree of Master of Sciences (M.Sc.)

**SUBGRADIENTS METHODS FOR OTIMIZATION OF NONDIFFERENTIABLE
CONVEX FUNCTION.**

Théssera Christine Araújo de Souza

August / 2008

Supervisor: Wilhelm Passarella Freire

The goal of this work is describe the State of the Art about Subgradients Methods for optimization of nondifferentiable convex functions. We initially present a historical of these methods, basic concepts on differentiable optimization, necessary to the comprehension of certain important notions about nondifferentiable problems, as well as these problems and its own characteristics. Subsequently, a short introduction about nondifferentiable methods is presented for, then, devote to Subgradients Methods, its extensions and recent works. The Dissertation is finished with the presentation of some applications, its results and conclusions.

Sumário

1. Introdução	1
1.1. Otimização	1
1.2. História dos Métodos de Otimização Não Diferenciáveis	2
1.3. Estrutura do Trabalho	4
2. Noções Preliminares	6
2.1. Definições Gerais	6
2.2. Caso Diferenciável	11
2.3. Convexidade	11
3. Otimização Não Linear Diferenciável	18
3.1. Otimização Irrestrita	18
3.2. Otimização com Restrições	19
3.2.1. Otimização com Restrições de Igualdade	19
3.2.2. Otimização com Restrição de Desigualdade	22
3.2.3. Otimização com Restrição de Igualdade e Desigualdade	24
3.3. Alguns Métodos de Otimização Não Linear Diferenciável	25
3.3.1. Métodos de Otimização para Problemas Irrestritos	26
3.3.1.1. Método de Newton	26
3.3.1.1.1. Algoritmo	27
3.3.1.1.2. Algumas limitações	28
3.3.1.2. Métodos Gradientes	28
3.3.1.2.1. Algoritmo	28
3.3.1.2.2. Interpretação Geométrica	29
3.3.1.2.3. Convergência do Método Gradiente	29
3.3.1.3. Métodos de Gradientes Conjugados	30
3.3.2. Métodos de Otimização para Problemas com Restrições	30
3.3.2.1. Método de Gradiente Projetado	30
3.3.2.1.1. Exemplo	31
3.3.2.2. Métodos de Penalidades	32
3.3.2.3. Métodos de Barreira	33
3.3.2.4. Método de Direções Viáveis	33
3.3.2.5. Métodos de Pontos Interiores	34

3.3.2.6. Métodos Afim Escala-----	34
3.3.3. Métodos Quase Newton-----	36
3.3.3.1. DFP-----	36
3.3.3.2. BFGS-----	37
4. Introdução aos Métodos Não Diferenciáveis -----	38
4.1. Alguns Métodos de Otimização Não Diferenciável.-----	38
4.1.1. Métodos Subgradientes -----	39
4.1.2. Métodos de Planos de Corte-----	39
4.1.2.1. Algoritmo -----	40
4.1.3. Métodos de Centros Analíticos -----	42
4.1.4. Métodos de Centros Analíticos e Planos de Corte -----	43
4.1.5. Métodos de Feixes -----	44
4.1.6. Método de direções viáveis -----	45
5. Métodos subgradientes e suas Extensões-----	47
5.1. Métodos Subgradientes -----	47
5.1.1. O Método Subgradiente -----	47
5.1.2. Dificuldades Encontradas -----	48
5.1.3. Escolha do Tamanho do Passo -----	50
5.1.4. Algoritmo-----	54
5.1.5. Exemplo -----	55
5.1.6. Uma Aplicação Importante dos Métodos Subgradientes -----	57
5.1.6.1. Relaxação Lagrangeana -----	57
5.2. Extensões dos Métodos Subgradientes -----	58
5.2.1. Métodos de Dilatação do Espaço-----	58
5.2.1.1. Métodos de Dilatação do Espaço na Direção	
do Subgradiente -----	60
5.2.1.2.1. Algoritmo-----	61
5.2.1.1.2. Convergência dos Métodos de Dilatação	
do Espaço na Direção do Subgradiente-----	63
5.2.1.2. Método de Elipsóides -----	63
5.2.1.2.1. Algoritmo -----	64
5.2.1.3. Métodos Subgradientes com Dilatação	
do Espaço na Direção da Diferença de Dois	

Subgradientes Sucessivos -----	65
5.2.2.3.1. r-algoritmos -----	66
5.2.1.3.2. Algoritmo-----	67
5.2.1.3.3. Exemplo -----	68
5.2.2. Métodos ε -Subgradientes ou ε -Máxima Descida -----	69
5.2.3. Método Subgradiente Conjugado-----	74
5.2.3.1. Algoritmo -----	75
5.2.3.2. Exemplo-----	76
5.2.4. Métodos Subgradientes Projetados -----	78
5.2.5. Métodos Subgradientes Afim Escalas -----	79
5.3. Trabalhos Recentes -----	81
6. Alguns Exemplos Numéricos-----	87
6.1. Problemas Teste-----	87
6.2. Resultados Obtidos-----	89
6.3 6.3 Um Exemplo dos Métodos de Dilatação do Espaço na Direção do Subgradiente-----	93
7. Conclusões -----	95

Nomenclatura

Vamos listar a simbologia comum a todos os capítulos da dissertação. Outras definições e notações serão apresentadas no capítulo 2. Deixaremos para cada capítulo as notações específicas utilizadas no capítulo. Usaremos letras maiúsculas para conjuntos e matrizes e minúsculas para vetores.

Nesta Dissertação,

$\mathbb{N} = \{1, 2, 3, \dots\}$ é o conjunto dos números naturais.

\mathbb{Z} representará o conjunto dos números inteiros.

$\mathbb{Z}^n = \{(x_1, x_2, \dots, x_n) / x_i \in \mathbb{Z}, i = 1, 2, \dots, n\}$.

$\mathbb{Z}_+^n = \{(x_1, x_2, \dots, x_n) / x_i \in \mathbb{Z}, x_i > 0, i = 1, 2, \dots, n\}$

\mathbb{R} representará o conjunto dos números reais.

$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) / x_i \in \mathbb{R}, i = 1, 2, \dots, n\}$ é o espaço euclidiano n -dimensional.

$\mathbb{R}_+^n = \{(x_1, x_2, \dots, x_n) / x_i \in \mathbb{R}, x_i > 0, i = 1, 2, \dots, n\}$

I_n representa a matriz identidade de ordem n , ou seja,

$$\begin{pmatrix} 1 & 0 & \Lambda & 0 \\ 0 & 1 & \Lambda & 0 \\ \text{M} & \text{M} & \text{O} & \text{M} \\ 0 & 0 & \Lambda & 1 \end{pmatrix}$$

$G(x)$ Representa a matriz diagonal abaixo:

$$G(x) = \text{diag}(g_1, g_2, \dots, g_m) = \begin{pmatrix} g_1 & 0 & \Lambda & 0 \\ 0 & g_2 & \Lambda & 0 \\ \text{M} & \text{M} & \text{O} & \text{M} \\ 0 & 0 & \Lambda & g_m \end{pmatrix}.$$

Em todo o texto usaremos vetores coluna. Assim, se $x \in \mathbb{R}^n$, então

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ e } x = (x_1, x_2, \dots, x_n)^t.$$

O produto interno de dois vetores $x, y \in \mathbb{R}^n$ ser indicado por

$$x \cdot y = \sum_{i=1}^n x_i y_i = x^t y.$$

Indicaremos a norma do vetor x por $\|x\| = (x \cdot x)^{\frac{1}{2}}$. Em particular, se $n=1$ escreveremos $\|x\| = |x|$.

Seja $A = (a_{ij})$ uma matriz de ordem $m \times n$. Então $A^t = (a_{ji})$ indicará a matriz transposta de A e

$$A_i = \begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{im} \end{pmatrix} \text{ é a } i\text{-ésima coluna de } A.$$

Seja $X \subset \mathbb{R}^n$.

Indicaremos por X^0 o interior de X e $X^c = \mathbb{R}^n - X$ o complementar de X .

O vetor gradiente da função $f: X \subset \mathbb{R}^n \rightarrow \mathbb{R}$ em $a \in X$ será denotado por

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \frac{\partial f}{\partial x_2}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right)^t$$

e a matriz hessiana por

$$Hf(a) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_2^2}(a) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(a) \\ \text{M} & \text{M} & \text{O} & \text{M} \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \frac{\partial^2 f}{\partial x_n \partial x_2}(a) & \text{K} & \frac{\partial^2 f}{\partial x_n^2}(a) \end{bmatrix}$$

A matriz jacobiana da função $g : X \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ no ponto $a \in X$ indicada por

$$Jg(a) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(a) & \frac{\partial g_1}{\partial x_2}(a) & \dots & \frac{\partial g_1}{\partial x_n}(a) \\ \frac{\partial g_2}{\partial x_1}(a) & \frac{\partial g_2}{\partial x_2}(a) & \dots & \frac{\partial g_2}{\partial x_n}(a) \\ \text{M} & \text{M} & \text{O} & \text{M} \\ \frac{\partial g_m}{\partial x_1}(a) & \frac{\partial g_m}{\partial x_2}(a) & \text{K} & \frac{\partial g_m}{\partial x_n}(a) \end{bmatrix}$$

Nos algoritmos que apresentaremos d indicará a direção de busca e t denotará o passo dado na direção d .

O símbolo x^k indicará um elemento da seqüência de vetores $\{x^k\} \subset \mathbb{R}^n$, x_k indicará um elemento da seqüência de números reais $\{x_k\} \subset \mathbb{R}$ e A^k indicará um elemento da seqüência de matrizes $\{A^k\}$.

A bola aberta de centro em a e raio $r > 0$ será denotada por $B(a, r)$.

Dizer que o vetor $v \in \mathbb{R}^n$ e tal que $v \geq 0$ significará $v_i \geq 0$ para todo $i = 1, 2, \dots, n$, valendo a mesma observação para as desigualdades $<$, \leq e $>$.

A notação $v < \infty$ estará indicando que o vetor $v \in \mathbb{R}^n$ é limitado superiormente.

Capítulo 1

INTRODUÇÃO

1.1 Otimização

A Otimização é a área da Programação Matemática que trata de problemas cujo interesse consiste em encontrar pontos de máximo ou de mínimo de funções.

A Otimização pode ser dividida em diferenciável e não diferenciável. A Otimização Diferenciável trabalha com problemas cujas funções possuem derivadas em todos os pontos do seu domínio, enquanto que a Otimização Não Diferenciável trabalha com problemas cujas funções não possuem derivadas em alguns pontos do seu domínio. Este trabalho será focado nesta classe de problemas.

Em geral, os métodos de otimização são métodos iterativos, os quais a partir de um ponto inicial, constroem uma seqüência de pontos que se aproximam da solução do problema.

Trabalharemos com Otimização Convexa cujas funções são convexas não necessariamente diferenciáveis. Para resolver esta classe de problemas são necessárias técnicas que substituem o cálculo diferencial clássico. Tais técnicas são derivadas de uma área da Matemática chamada Análise Convexa. No caso de funções não convexas, estendem-se os conceitos da Análise Convexa para funções localmente lipschitzianas.

Portanto, estamos interessados em problemas de otimização não diferenciáveis, com a hipótese adicional de que f é convexa. Nas três últimas décadas, este problema tem recebido uma grande atenção e vários novos métodos têm sido propostos.

Como maximizar uma função f equivale a minimizar $-f$, consideraremos neste trabalho apenas problemas de minimização.

1.2 História dos Métodos de Otimização Não Diferenciável

Os Métodos de otimização, na maioria das vezes, utilizam a seguinte fórmula para atualizar o ponto: $x^{k+1} = x^k + t_k d^k$, cujo t_k é o tamanho do passo e d^k a direção de busca.

Na otimização diferenciável clássica, geralmente, os métodos de minimização usados são baseados na aproximação de primeira e segunda ordem envolvendo, respectivamente, o gradiente (∇f) e a hessiana (Hf) da função objetivo f . Os métodos mais conhecidos em otimização diferenciável são os Métodos Gradientes e o Método de Newton.

Na otimização não diferenciável, existem pontos do domínio onde a função não possui gradiente e, obviamente, nem matriz hessiana. Desta forma, aplicar métodos clássicos para resolver problemas diferenciáveis em problemas não diferenciáveis fica inviável. O que se pode fazer é adaptar os métodos clássicos para funções não diferenciáveis. Por exemplo, os Métodos Gradientes podem ser modificados para funções não diferenciáveis utilizando a direção oposta a um subgradiente como a direção de busca. Esses métodos, denominados Métodos Subgradientes, elaborados por Shor na União Soviética na década de 60, foram os primeiros métodos de otimização não diferenciável. Eles possuem uma estrutura extremamente simples, mas em geral, não apresentam bons resultados numéricos.

Nos Métodos Subgradientes a direção oposta ao subgradiente não é necessariamente uma direção de descida, e mesmo se for, não se tem a garantia de convergência do método. Para contornar essa dificuldade, uma opção é escolher o tamanho do passo t_k conveniente.

Tal como acontece com a direção do gradiente para o caso diferenciável, a direção oposta ao subgradiente no caso não diferenciável pode resultar em um fenômeno zig-zag que possivelmente manifesta-se em alguma fase do Algoritmo Subgradiente, causando problemas nas condições de otimalidade. Para superar este problema, foi elaborado um

método, denominado Método Subgradiente Conjugado, cuja direção de busca é calculada através da combinação da direção oposta ao subgradiente atual com a direção anterior.

A maioria dos algoritmos para otimização não diferenciável utilizam subgradientes com algumas modificações. Estas modificações, normalmente, estão relacionadas à forma de encontrar uma direção de busca e de se utilizar um tamanho do passo adequado para adotar ao longo desta direção. Vêm sendo propostos, ao longo dos últimos anos, vários métodos de otimização não diferenciáveis.

Um problema encontrado nos Métodos Subgradientes é a necessidade de se conhecer o subdiferencial $\partial f(x^k)$ da função f no ponto corrente x^k . Para vencer tal dificuldade, os Métodos ε -Subgradientes utilizam não apenas um subgradiente no ponto corrente, mas subgradientes de pontos de uma vizinhança do ponto corrente.

Shor em [63] e [64] apresenta um Método de Dilatação do Espaço que utiliza um processo análogo de transformação do espaço ao longo da direção do subgradiente. De notável valor devido ao seu desempenho computacional, Shor [64], propõem um Algoritmo de Dilatação do Espaço que dilata o espaço ao longo da diferença de dois subgradiente sucessivos, denominado r-algoritmo.

O método clássico de Planos de Corte, apresentado por Kelley em [35], utiliza subgradientes para construir uma aproximação em um ponto de uma função convexa não linear por uma função linear, ou seja, utiliza subgradientes para aproximar a função f por modelo linear por partes.

Os Métodos de Feixes, propostos por Mifflin [46] e Lemarechal [39], são atualmente os que apresentam melhores resultados. A direção de busca é obtida através da combinação convexa de um conjunto de subgradientes gerados em iterações anteriores. Ao contrário dos Métodos Subgradientes, nos Métodos de Feixes a escolha do passo envolve uma busca linear inexata que produz uma solução melhor (passo sério) ou uma solução que é rejeitada (passo nulo). Em ambos os casos, um novo subgradiente é calculado e acrescentado ao feixe atual para encontrar uma direção de busca modificada.

Uma diferença importante entre os Métodos de Feixes e os Métodos Subgradientes é que, contrariamente a este último, os Métodos Tipo Feixes geram uma seqüência de iterações para que os valores da função objetivo sejam monótonos decrescentes. Por esta razão, esses métodos são classificados como "Métodos de Descida". No entanto, uma dificuldade encontrada nos Métodos de Feixes é que eles exigem a solução de um subproblema quadrático em cada iteração para encontrar a direção de busca, e isto pode tornar-se bastante caro, em particular para problemas maiores.

Um método que vem recebendo atenção nos últimos anos é o Método de Centros Analíticos. Em geral, os métodos baseados na noção de centro realizam um corte em uma região convexa, limitada e que contenha solução ótima. Cada corte utiliza um hiperplano, que separa esta região em duas partes. Na parte onde a solução do problema está contida são realizadas operações, de acordo com o método que está sendo utilizado, para determinar o próximo centro.

1.3 Estrutura do Trabalho

Este trabalho está organizado em sete capítulos, sendo este o primeiro. No capítulo 2, relembramos algumas definições e resultados que serão utilizados nos capítulos subseqüentes. No capítulo 3 introduziremos o problema de otimização não linear diferenciável, as condições de otimalidade para esses problemas e apresentaremos alguns métodos para problemas diferenciáveis. No capítulo 4 faremos uma rápida introdução aos métodos para problemas não diferenciáveis.

O objetivo deste trabalho é estudar os Métodos Subgradiente e suas extensões, o que faremos no capítulo 5. O que motivou o nosso estudo é que mesmo não apresentando bons resultados numéricos, os Métodos Subgradientes e suas extensões foram utilizados em trabalhos recentes, como por exemplo, em [6], [13], [25], [27], [47], [48], [49], [51], [54], [56], e [57].

No capítulo 6 são apresentados os problemas teste e os resultados obtidos utilizando os Métodos Subgradientes e um exemplo dos Métodos de Dilatação do Espaço na Direção do Subgradiente.

Para finalizar apresentamos as conclusões e as referências bibliográficas.

Capítulo 2

NOÇÕES PRELIMINARES

Neste capítulo apresentaremos algumas definições e resultados necessários para o desenvolvimento dos demais capítulos.

2.1 Definições Gerais

Definição 2.1.1

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Uma direção $d \in \mathbb{R}^n$ é dita direção de descida para f em $a \in \mathbb{R}^n$, se existe $\lambda > 0$ tal que $f(a + td) < f(a)$, $\forall t \in (0, \lambda)$.

Observe na figura 2.1 que a direção oposta a d não é direção de descida. Note também, que se $t > \lambda$ então $f(a+td) > f(a)$.

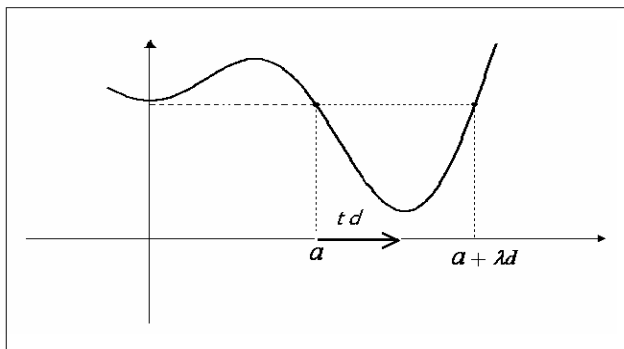


Figura 2.1

Definição 2.1.2

Um vetor $d \in \mathbb{R}^n$ é uma direção viável em $a \in X \subset \mathbb{R}^n$ se existe $\lambda > 0$ tal que $a + td \in X$ para todo $t \in [0, \lambda]$.

Obviamente, se a pertence ao interior de X , então toda direção $d \in \mathbb{R}^n$ é viável em a .

As figuras 2.2 e 2.3 ilustram uma direção viável e uma direção não viável.

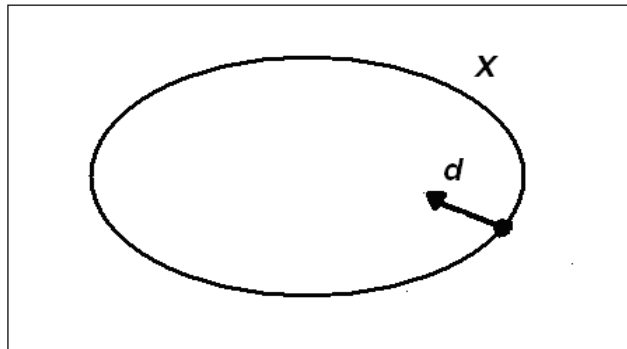


figura 2.2: direção viável

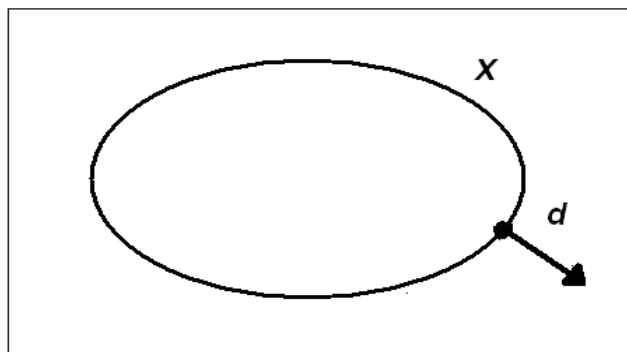


figura 2.3: direção não viável

Definição 2.1.3

O epígrafo de uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é o conjunto:

$$\text{epi}(f) = \{(x, r) \in \mathbb{R}^n \times \mathbb{R}; f(x) \leq r\}.$$

Veja a figura 2.4 cuja parte colorida ilustra o epígrafo de f .

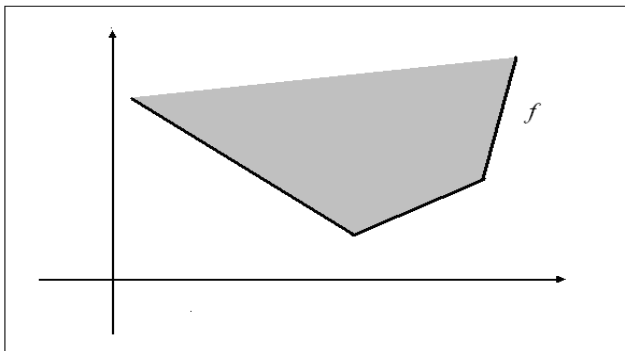


Figura 2.4

Definição 2.1.4

Uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é localmente lipschitziana com constante K em $a \in \mathbb{R}^n$ se existe $r > 0$ tal que:

$$|f(y) - f(x)| \leq K \|y - x\|, \quad \forall x, y \in B(a, r)$$

Definição 2.1.5

Uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é dita

1. Positivamente homogênea quando

$$f(\lambda x) = \lambda f(x), \quad \forall \lambda \geq 0, \quad \forall x \in \mathbb{R}^n.$$

2. Subaditiva quando

$$f(x + y) \leq f(x) + f(y), \quad \forall x, y \in \mathbb{R}^n.$$

Definição 2.1.6

Uma matriz P $n \times n$ é chamada de matriz projeção se $P = P^t$ e $PP = P$.

Lema 2.1.1: Seja P a matriz $n \times n$, então:

1. Se P é uma matriz projeção, então P é semidefinida positiva.
2. P é uma matriz projeção se, e somente se, $I - P$ é uma matriz projeção.

Definição 2.1.7

Sejam $d \in \mathbb{R}^n$ tal que $\|d\|=1$ e $\alpha \geq 0$. Um operador $R_\alpha(d)$ com um vetor transformação x da forma $R_\alpha(d)x = (\alpha - 1)(x \cdot d)d + x$ é chamado operador de dilatação do espaço ao longo da direção d com coeficiente de dilatação α .

Para a melhor visualização deste operador observe na figura 2.5 o seu efeito quando $\alpha < 1$ e quando $\alpha > 1$. Note que se $\alpha < 1$ o vetor é contraído e se $\alpha > 1$ o vetor é dilatado.

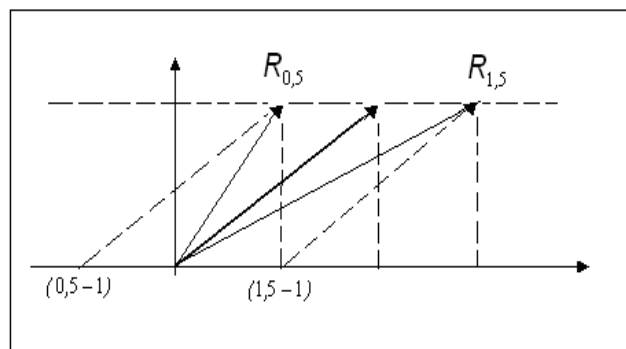


Figura 2.5

Propriedades:

1. $R_\alpha(d) = (\alpha - 1) d \cdot d + I$;
2. $R_\alpha(d)^{-1} x = R_{\frac{1}{\alpha}}(d) = \left(\frac{1}{\alpha} - 1\right)(x \cdot d)d + x = \left[\left(\frac{1}{\alpha} - 1\right) d \cdot d + I\right] x$;
3. $R_{\alpha\beta}(d) = R_\alpha(d) R_\beta(d)$;
4. $R_\alpha(d) R_\alpha(d)^{-1} = I$;
5. $x = R_\alpha(d) y$ no qual $y = R_\alpha(d)^{-1} x$, ou seja, y é obtido pela dilatação do vetor x

ao longo da direção d dependendo do valor de α .

Definição 2.1.8

Considere o problema de minimização:

$$\begin{cases} \text{Minimizar} & f(x) \\ \text{sujeito a} & h(x) = 0 \\ & g(x) \leq 0 \end{cases}$$

Com $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h: \mathbb{R}^n \rightarrow \mathbb{R}^p$, no qual, $h(x) = (h_1(x), h_2(x), \dots, h_p(x))$ e $g(x) = (g_1(x), g_2(x), \dots, g_m(x))$.

1. A função f é denominada função objetivo.
2. As funções $h_i: \mathbb{R}^n \rightarrow \mathbb{R}$, $i=1, \dots, p$, são denominadas restrições de igualdade do problema.
3. As funções $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$, $i=1, \dots, m$, são denominadas restrições de desigualdade do problema.
4. Uma restrição $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$ será dita ativa no ponto $a \in \mathbb{R}^n$ se $g_i(a) = 0$.
Denotaremos por $I(a) = \{i; g_i(a) = 0\}$ o conjunto de restrições ativas no ponto a .
5. O ponto a do problema de otimização com restrição é dito ponto regular, se o conjunto $\{\nabla g_i(a); i \in I(a)\}$ é linearmente independente.

Definição 2.1.9

Seja a um ponto regular do problema

$$\begin{cases} \text{Minimizar} & f(x) \\ \text{sujeito a} & h(x) = 0 \end{cases} \quad (3.2)$$

Com $f: \mathbb{R}^n \rightarrow \mathbb{R}$ e $h: \mathbb{R}^n \rightarrow \mathbb{R}^p$, no qual, $h(x) = (h_1(x), h_2(x), \dots, h_p(x))$.

O espaço tangente em a é definido como:

$$T = \{x / \nabla h(a) \cdot x = 0\}$$

2.2 Caso Diferenciável

Definição 2.2.1

A função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é diferenciável no ponto $a \in \mathbb{R}^n$ quando existirem as derivadas parciais $\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)$ e, além disso, para todo vetor $v = (v_1, \dots, v_n)^t \in \mathbb{R}^n$ tivermos

$$f(a+v) = f(a) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) v_i + r(v) \quad \text{com} \quad \lim_{v \rightarrow 0} \frac{r(v)}{\|v\|} = 0.$$

Teorema 2.2.1:

Sejam $d \in \mathbb{R}^n$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável em a . Se $d \cdot \nabla f(a) < 0$ então d é uma direção de descida para f em a .

Teorema 2.2.2:

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável em $a \in \mathbb{R}^n$. Se a é um ponto de mínimo local de f então $\nabla f(a) = 0$.

2.3 Convexidade

Existem muitos problemas em otimização nos quais não se dispõe da diferenciabilidade das funções envolvidas. Entretanto, se essas funções possuem certas propriedades como, por exemplo, a convexidade, pode-se contar com resultados que facilitam na resolução de problemas de otimização. Veremos nesta seção alguns conceitos básicos sobre convexidade de conjuntos e de funções. Um estudo mais detalhado sobre convexidade pode ser encontrado em [5], [44] e [53]. O texto clássico sobre esse assunto é [58].

Definição 2.3.1

Seja $C \subset \mathbb{R}^n$. Dizemos que C é um conjunto convexo quando

$\lambda x + (1 - \lambda)y \in C$, para todo $x, y \in C$ e para todo $\lambda \in [0, 1]$.

Geometricamente, significa que ao unirmos dois pontos quaisquer de C por um segmento de reta, esse estará contido em C . Caso contrário, será não-convexo. As figuras 2.6. e 2.7 ilustram esse conceito.

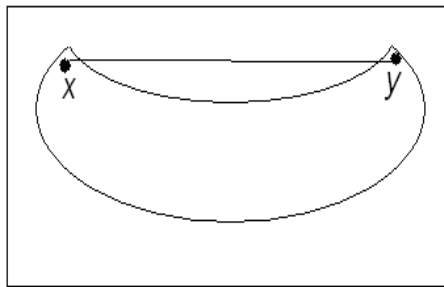


figura 2.6: Não convexo

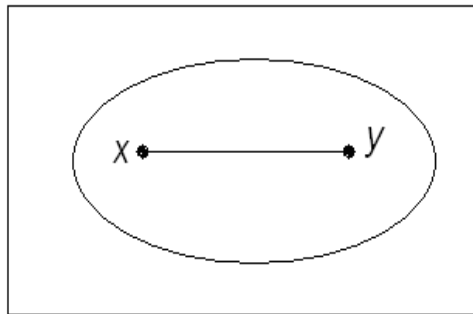


figura 2.7: convexo

Definição 2.3.2:

Uma combinação linear $\sum_{j=1}^k \lambda_j x^j$ é chamada combinação convexa dos pontos

x_1, \dots, x_k em \mathbb{R}^n , se $\lambda_j \geq 0$, $j=1, \dots, k$ e $\sum_{j=1}^k \lambda_j = 1$.

Definição 2.3.3:

A envoltória convexa de $C \subset \mathbb{R}^n$, denotada por $\text{conv } C$, é o conjunto de todas as combinações convexas de pontos em C , ou seja, $\text{conv } C$ é o menor conjunto convexo que contém C .

Tem-se que C é convexo se, e somente se, $C = \text{conv } C$.

Definição 2.3.4:

Seja $f : C \rightarrow \mathbb{R}^n$ uma função definida no conjunto convexo $C \subset \mathbb{R}^n$.

1. f é convexa se $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$ para todo $x, y \in C$ e para todo $\lambda \in [0, 1]$.
2. f é estritamente convexa se $f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$ para todo $x, y \in C$ e para todo $\lambda \in [0, 1]$.

No exemplo ilustrado pela figura 2.8, a função não é convexa pois existem pontos para os quais $f(\lambda x + (1 - \lambda)y) > \lambda f(x) + (1 - \lambda)f(y)$. Por outro lado, no exemplo ilustrado na figura 2.9, a função é convexa.

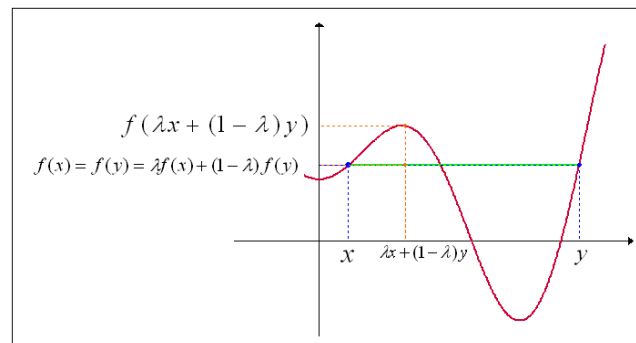


Figura 2.8 (função não convexa)

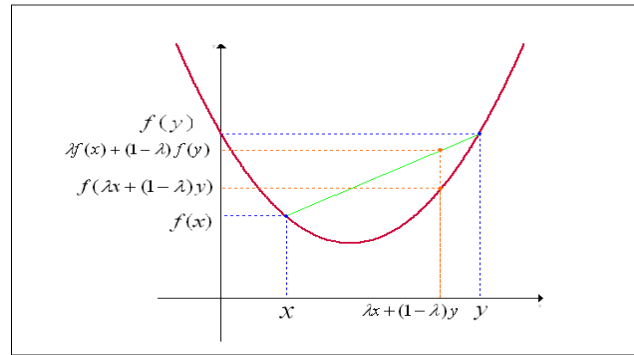


figura 2.9 (função convexa)

O teorema 2.3.1 mostra a importância da convexidade das funções nos problemas de otimização.

Teorema 2.3.1

Sejam $C \subset \mathbb{R}^n$ convexo e $f : C \rightarrow \mathbb{R}$ uma função convexa.

Todo mínimo local de f é também um mínimo global. Se f for estritamente convexa, então todo mínimo local é o único mínimo global de f em C .

Definição 2.3.5

A norma mínima de S , denotada por $\| \cdot \|_S$, é a menor norma de S . Isto é, para todo conjunto $S \subseteq \mathbb{R}^n$ existe um único ponto v no fecho da envoltória convexa de S tendo norma mínima. Algebricamente, o ponto é caracterizado pela relação:

$$\|s\| \geq \|v\| \text{ para todo } s \in S.$$

Definição 2.3.6

O subdiferencial da função convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no ponto $a \in \mathbb{R}^n$ é o conjunto

$$\partial f(a) = \{g \in \mathbb{R}^n; f(x) \geq f(a) + g \cdot (x - a), \forall x \in \mathbb{R}^n\}.$$

Os elementos $g \in \partial f(a)$ são chamados de subgradientes.

Geometricamente, um subgradiente g define um hiperplano suporte não vertical ao epígrafo de f no ponto $(a, f(a))$ como mostra a figura 2.10. Assim, g é um subgradiente de f no ponto a se, e somente se, o hiperplano em \mathbb{R}^{n+1} com normal $(-g, 1)$ que passa através de $(a, f(a))$, suporta o epígrafo de f .

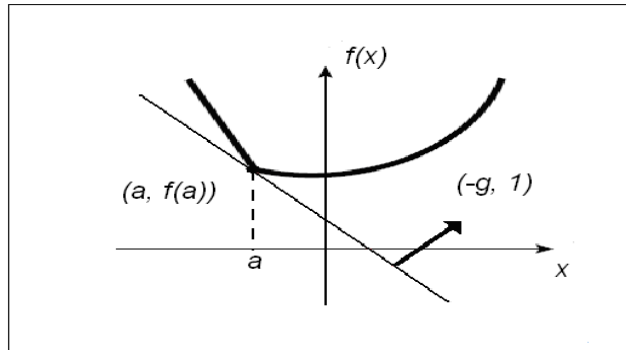


Figura 2.10.

Definição 2.3.7:

Seja $\varepsilon \geq 0$, então o ε -subdiferencial de uma função convexa $f: \mathbb{R}^n \rightarrow \mathbb{R}$ em $a \in \mathbb{R}^n$ é o conjunto

$$\partial f_\varepsilon(a) = \{g \in \mathbb{R}^n; f(x) \geq f(a) + g \cdot (x - a) - \varepsilon, \forall x \in \mathbb{R}^n\}.$$

Os elementos $g \in \partial f_\varepsilon(a)$ são chamados de ε -subgradientes.

A figura 2.11 ilustra um ε -subgradiente de uma função convexa f . g é um ε -subgradiente em a , pois, o epígrafo de f está contido no semi-espaco acima do plano em \mathbb{R}^2 de normal $(-g, 1)$ que passa através de $(a, f(a) - \varepsilon)$.

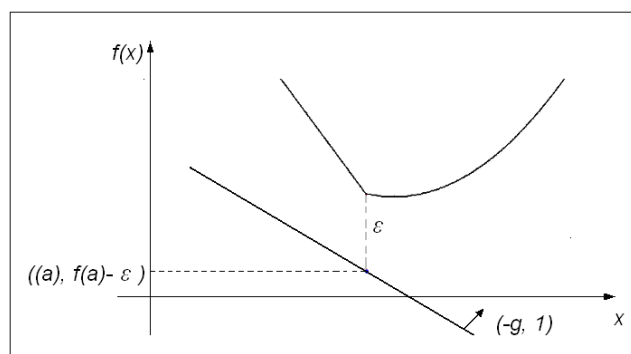


figura 2.11

Teorema 2.3.2

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e diferenciável em $x \in \mathbb{R}^n$, então:

$$\partial f(x) = \{\nabla f(x)\}$$

Teorema 2.3.3

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Então, para todo $x \in \mathbb{R}^n$, f é *localmente lipschitziana* em x .

Uma consequência imediata do Teorema 2.3.4 é que convexidade implica em continuidade.

Teorema 2.3.4

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa com constante de Lipschitz K em $x \in \mathbb{R}^n$.

Então:

1. $\partial_0 f(x) = \partial f(x)$;
2. Se $\varepsilon_1 \leq \varepsilon_2$ então $\partial_{\varepsilon_1} f(x) \subset \partial_{\varepsilon_2} f(x)$;
3. $\partial_{\varepsilon} f(x)$ é um conjunto não vazio, convexo e compacto tal que $\|g\| \leq K$, para

todo $g \in \partial_{\varepsilon} f(x)$;

4. $f_{\varepsilon}'(x;d) = \max \{g \cdot d; g \in \partial_{\varepsilon} f(x), \forall d \in \mathbb{R}^n\}$;

5. $\partial_{\varepsilon} f(x) = \{g \in \mathbb{R}^n; f_{\varepsilon}'(x;d) \geq g \cdot d, \forall d \in \mathbb{R}^n\}$.

Teorema 2.3.5

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. As seguintes condições são equivalentes:

- (1) f atinge seu mínimo global em a ;
- (2) $0 \in \partial f(a)$.

Teorema 2.3.6 (Teorema MinMax)

Sejam $A, B \subset \mathbb{R}^n$ conjuntos convexos e compactos. Então

$$\min_{x \in A} \max_{y \in B} x \cdot y = \max_{y \in B} \min_{x \in A} x \cdot y.$$

Capítulo 3

OTIMIZAÇÃO NÃO LINEAR DIFERENCIÁVEL

Nesse capítulo introduziremos o problema de otimização não linear diferenciável. Apresentaremos as condições de otimalidade e caracterizaremos suas soluções. Em seguida, apresentaremos alguns dos métodos para resolvê-los.

Um problema de otimização é não linear quando a função objetivo e/ou restrições são funções não lineares. Faremos uma breve exposição sobre otimização irrestrita, e também sobre otimização com restrições de igualdade e/ou desigualdades.

3.1 Otimização Irrestrita

O problema de programação não linear irrestrito é apresentado da seguinte forma:

$$(PNL_1) \begin{cases} \text{Minimizar } f(x) \\ x \in \mathbb{R}^n \end{cases} \quad (3.1)$$

com $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável.

Teorema 3.1: (Condições necessárias de primeira ordem)

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^1$. Se x^* é um mínimo local de f em \mathbb{R}^n então $\nabla f(x^*) = 0$.

Teorema 3.2: (Condições necessárias de segunda ordem)

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$. Se x^* é um mínimo local de f em \mathbb{R}^n então:

- $\nabla f(x^*) = 0$;
- $Hf(x^*)$ é semidefinida positiva.

Teorema 3.3: (Condições suficientes de segunda ordem)

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$. Se $x^* \in \mathbb{R}^n$, $\nabla f(x^*) = 0$ e $Hf(x^*) > 0$ (hessiana definida positiva) então x^* é um mínimo local estrito de f em \mathbb{R}^n .

As provas dos teoremas acima estão em [26].

3.2. Otimização com Restrições

3.2.1. Otimização com Restrições de Igualdade

Consideraremos o problema de otimização não linear com restrições de igualdade representado da seguinte forma:

$$(\text{PNL}_2) \begin{cases} \text{Minimizar} & f(x) \\ \text{sujeito a} & h(x) = 0 \end{cases} \quad (3.3)$$

onde a função objetivo $f: \mathbb{R}^n \rightarrow \mathbb{R}$ e as restrições $h: \mathbb{R}^n \rightarrow \mathbb{R}^p$ são funções diferenciáveis e $h(x) = (h_1(x), h_2(x), \dots, h_p(x))$.

Determinaremos de forma intuitiva as condições de otimalidade de primeira ordem para uma função de duas variáveis sujeita a uma restrição de igualdade:

$$\begin{cases} \text{Minimizar} & f(x,y) \\ \text{sujeito a} & h(x,y) - c = 0 \end{cases}$$

Na figura 3.1 estão representadas as curva de nível $f(x,y) = c_1$, $f(x,y) = c_2$, $f(x,y) = c_3$ e $f(x,y) = c_4$ onde, supõe-se $c_1 < c_2 < c_3 < c_4$, e a curva $h(x,y) = c$.

Note que o ponto q não pode ser extremo local de f , pois existem curvas de nível f que interceptam $h(x,y) = c$ em pontos arbitrariamente próximos, tanto ao lado direito de q cujo valor da função é menor do que o valor no ponto q , quanto ao lado esquerdo de q cujo valor

da função de f é maior do que o valor no ponto q . Portanto, q não pode ser nem máximo e nem mínimo local de f .

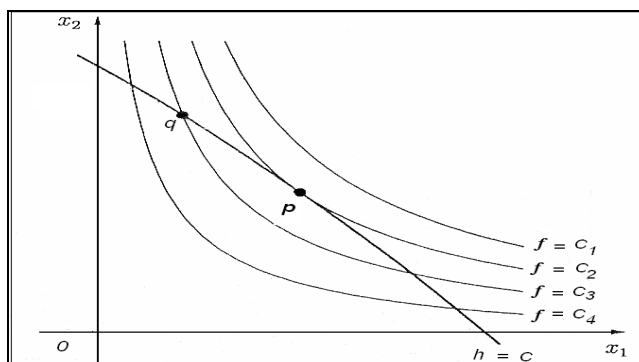


figura 3.1

Vejamos agora, a curva de nível de f que passa pelo ponto p e é tangente à curva $h(x,y)=c$. As curvas de nível interceptam $h(x,y)=c$ em pontos arbitrariamente próximos, tanto ao lado direito quanto ao lado esquerdo do ponto p . Podemos observar que o valor da função nesses pontos é maior do que no ponto p . Logo, o valor de f em qualquer ponto diferente de p é sempre maior do que o valor da função no ponto p . Portanto, p é um ponto de mínimo local de f .

Se olharmos a curva $h(x,y)=c$ como uma curva de nível da função de duas variáveis h no nível c e se o vetor gradiente de h no ponto p é diferente de zero ($\nabla h(p) \neq 0$), então ele é perpendicular à curva $h(x,y)=c$. Analogamente, se vetor gradiente de f no ponto p é diferente de zero ($\nabla f(p) \neq 0$), então ele é perpendicular à curva de nível $f(x,y) = c_2$. Como as curvas $h(x,y)=c$ e $f(x,y)=c_2$ são tangentes em p , tem-se que $\nabla f(p)$ e $\nabla h(p)$ devem ser paralelos como pode ser observado nas figuras 3.2 e 3.3,. Logo, existe $\lambda \in \mathbb{R}$ tal que $\nabla f(p) + \lambda \cdot \nabla h(p) = 0$.

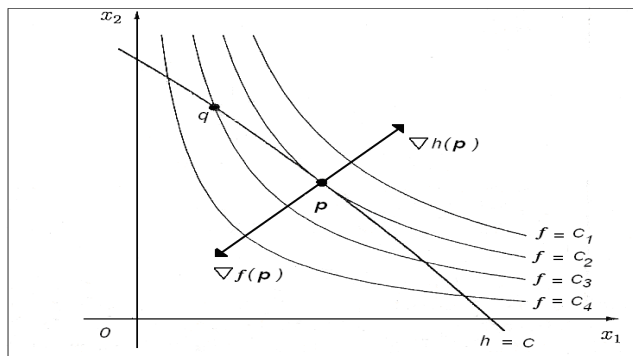


Figura 3.2

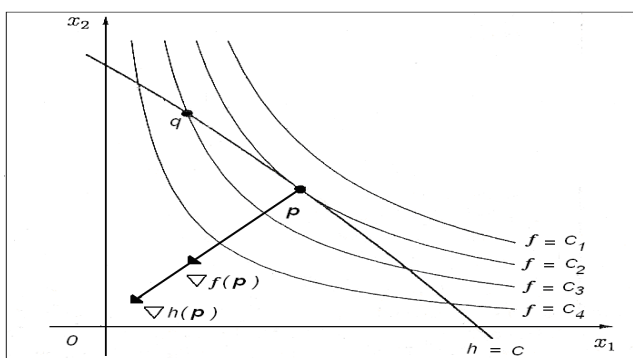


Figura 3.3

Generalizando obtemos o teorema 3.4.

Teorema 3.4: (Condição necessária de primeira ordem)

Seja x^* um ponto regular de $h(x)=0$ e solução do problema (3.3). Então existe um vetor $\lambda^* \in \mathbb{R}^p$ tal que:

- $\nabla f(x^*) + \lambda^* \cdot Jh(x^*) = 0$;
- $h(x^*) = 0$.

Teorema 3.5: (Condição necessária de segunda ordem)

Seja x^* um ponto regular de $h(x)=0$ e um mínimo local do problema 3.3. Então existe um vetor λ^* tal que $\nabla f(x^*) + \lambda^* \cdot Jh(x^*) = 0$ e $h(x^*)=0$ e a matriz

$$H(x^*, \lambda^*) = Hf(x^*) + \sum_{i=1}^p \lambda_i^* Jh_i(x^*)$$

é semidefinida positiva sobre o espaço tangente, isto é, $y^t \cdot H(x^*, \lambda^*) y \geq 0$ para todo $y \in T$.

3.2.2. Otimização com Restrição de Desigualdade

Consideraremos agora o problema de otimização não linear com restrição de desigualdade que é representado da seguinte forma:

$$(PNL_3) \begin{cases} \text{Minimizar } f(x) \\ \text{sujeito a } g(x) \leq 0 \end{cases} \quad (3.4)$$

com $f: \mathbb{R}^n \rightarrow \mathbb{R}$ e $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ no qual f e g são funções diferenciáveis e $g(x) = (g_1(x), g_2(x), \dots, g_m(x))$.

Obteremos, intuitivamente, as condições de otimalidade de primeira ordem para uma função de duas variáveis sujeita a uma restrições de desigualdade:

$$\begin{cases} \text{Maximizar } f(x, y) \\ \text{sujeito a } g(x, y) - b \leq 0 \end{cases}$$

Suponha que x^* seja uma solução do problema acima.

Temos que $g(x^*) = b$ ou $g(x^*) < b$.

Se $g(x^*) = b$, ou seja, a restrição g está ativa no ponto x^* , temos um caso semelhante ao caso de uma única restrição de igualdade.

Como x^* é um ponto de mínimo, então $\nabla f(x^*)$ e $\nabla g(x^*)$ devem ser paralelos no ponto x^* . Podemos observar na figura 3.4 no qual $c_1 > c_2 > c_3 > c_4$, que $\nabla f(x^*) = -\mu^* \nabla h(x^*)$ com $\mu^* \geq 0$. Logo, para $\mu^* \geq 0$ temos $\nabla f(x^*) + \mu^* \cdot \nabla h(x^*) = 0$.

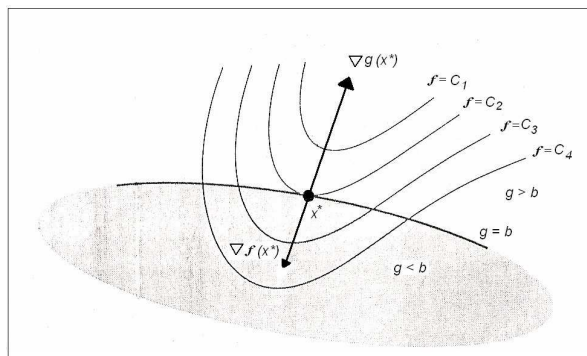


Figura 3.4.

Se $g(x^*) < b$, temos que x^* está no interior do conjunto $\{x \in \mathbb{R}^2; g(x) \leq b\}$. Como x^* é um ponto de mínimo local, então deve satisfazer a regra de Fermat, ou seja, $\nabla f(x^*) = 0$.

Logo temos:

$$1^{\circ} \text{ caso: } \begin{cases} g(x^*) = b \\ \mu^* \geq 0 \\ \nabla f(x^*) + \mu^* \nabla g(x^*) = 0 \end{cases}$$

ou

$$2^{\circ} \text{ caso: } \begin{cases} g(x^*) < b \\ \nabla f(x^*) = 0 \end{cases}$$

Podemos unificar os dois casos considerando o seguinte sistema:

$$\begin{cases} \mu^* [g(x^*) - b] = 0 & (3.3.2.1) \\ \nabla f(x^*) + \mu^* \nabla g(x^*) = 0 & (3.3.2.1) \\ \mu^* \geq 0 \\ g(x^*) \leq b \end{cases}$$

De fato, da condição 3.3.2.1 temos que $\mu^* = 0$ ou $g(x^*) - b = 0$.

Se $\mu^* = 0$ então de 3.3.2.2 temos que $\nabla f(x^*) = 0$, ou seja, x^* é um ponto crítico, logo estamos no segundo caso. Se $g(x^*) - b = 0$ então $g(x^*) = b$, ou seja, estamos no primeiro caso. Logo, $\nabla f(x^*) + \mu^* \nabla g(x^*) = 0$ para $\mu^* \geq 0$.

Generalizando obtemos o teorema 3.6.

Teorema 3.6: (Condição necessária de primeira ordem)

Seja x^* um ponto regular de $g(x) \leq 0$ e um mínimo local do problema 3.4. Então existe um vetor $\mu^* \in \mathbb{R}^m$ tal que:

- $\nabla f(x^*) + \mu^* \cdot Jg(x^*) = 0$;
- $G(x^*) \mu^* = 0$;
- $\mu^* \geq 0$;
- $g(x^*) \leq 0$;

com $G(x) = \text{diag}(g_1, g_2, \dots, g_m)$.

3.2.3. Otimização com Restrição de Igualdade e Desigualdade

Consideraremos agora o problema de otimização não linear com restrição de igualdade e de desigualdade que é representado da seguinte forma:

$$(\text{PNL}_4) \begin{cases} \text{Minimizar } f(x) \\ h(x) = 0 \\ g(x) \leq 0 \end{cases} \quad (3.5)$$

com $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h: \mathbb{R}^n \rightarrow \mathbb{R}^p$, no qual, f , g e h são funções diferenciáveis, $h(x) = (h_1(x), h_2(x), \dots, h_p(x))$ e $g(x) = (g_1(x), g_2(x), \dots, g_m(x))$.

Teorema 3.7: (Condição necessária de primeira ordem Karush-Khun-Tucker)

Seja x^* um ponto regular de $g(x) \leq 0$ e $h(x) = 0$ e um mínimo local do problema 3.5.

Então existem vetores $\lambda^* \in \mathbb{R}^p$ e $\mu^* \in \mathbb{R}^m$ tal que:

- $\nabla f(x^*) + \mu^* \cdot Jg(x^*) + \lambda^* \cdot Jh(x^*) = 0$;
- $G(x^*) \mu^* = 0$;
- $h(x^*) = 0$;

- $g(x^*) \leq 0$;
- $\mu^* \geq 0$;

com $G(x) = \text{diag}(g_1, g_2, \dots, g_m)$.

Teorema 3.8: (Condição necessária de segunda ordem)

Seja x^* um ponto regular de $h(x)=0$ e $g(x) \leq 0$ e um mínimo local do problema 3.5.

Então existem $\lambda^* \in \mathbb{R}^p$ e $\mu^* \in \mathbb{R}^m$ tal que o teorema 3.6 é satisfeito e a matriz

$$H(x^*, \lambda^*, \mu^*) = Hf(x^*) + \sum_{i=1}^p \lambda_i^* Jh_i(x^*) + \sum_{i=1}^m \mu_i^* Jg_i(x_i^*)$$

é semidefinida positiva sobre o espaço tangente, isto é, $y^t \cdot H(x^*, \lambda^*) y \geq 0$ para todo $y \in T$.

Teorema 3.9 (Condições suficientes de segunda ordem)

Seja $x^* \in \mathbb{R}^n$ tal que $h(x^*)=0$ e $g(x^*) > 0$. Então existem $\lambda^* \in \mathbb{R}^p$ e $\mu^* \in \mathbb{R}^m$

com $\mu^* \geq 0$ tal que $\nabla f(x^*) + \mu^* \cdot Jg(x^*) + \lambda^* \cdot Jh(x^*) = 0$ e $H(x^*, \lambda^*, \mu^*)$ é definida positiva sobre

o espaço tangente. Então x^* é um mínimo estrito local do problema 3.5.

3.3. Alguns Métodos de Otimização Não Linear Diferenciável

Foram desenvolvidos vários métodos para resolver problemas de otimização não linear diferenciável. Faremos uma breve exposição de alguns desses métodos que são importantes para um melhor entendimento dos métodos que estudaremos mais adiante.

3.3.1. Métodos de otimização para Problemas Irrestritos

Os métodos mais conhecidos para resolver problemas diferenciáveis irrestritos são os Métodos de Newton e o Método Gradiente. Porém, devido às dificuldades encontradas

nos Métodos Gradientes, como por exemplo, baixa velocidade de convergência, foram propostas algumas extensões desses Métodos.

3.3.1.1. Método de Newton

De maneira geral, o Método de Newton é usado para resolver sistemas de equações do tipo

$$\varphi(x) = 0 \quad (3.6)$$

no qual $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\varphi(x) = (\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x))$ é diferenciável.

Dada uma estimativa x^k , o Método de Newton aproxima φ pelo hiperplano tangente no ponto $\varphi(x^k)$. As soluções do sistema formado pelo hiperplano fornecem uma nova estimativa para x^k . Veja na figura 3.7 uma ilustração desse método.

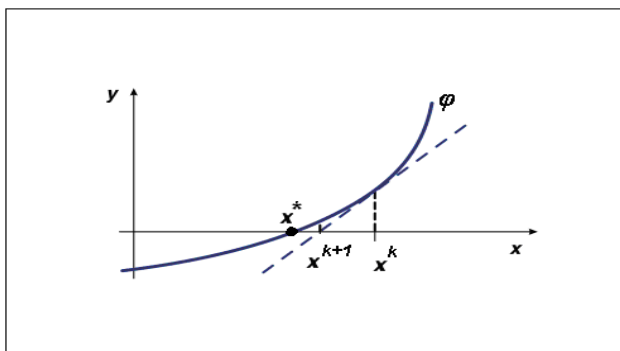


Figura 3.7

Para derivar as fórmulas utilizadas pelo método, vamos expandir $\varphi(x)$ por uma série de Taylor (truncada) ao redor do ponto x^k :

$$\varphi(x) = \varphi(x^k) + \nabla \varphi(x^k)(x - x^k) + \frac{1}{2}(x - x^k)H\varphi(x^k)(x - x^k)$$

O hiperplano tangente é dado pelos dois primeiros termos da série, ou seja,

$$y = \varphi(x^k) + \nabla \varphi(x^k)(x - x^k)$$

Fazendo $y = 0$, temos:

$$\varphi(x^k) = -\nabla \varphi(x^k)(x - x^k)$$

Considerando $d^k = x - x^k$, obtemos

$$\varphi(x^k) = -\nabla \varphi(x^k) \cdot d^k \quad (3.7)$$

Assim, a fórmula para encontrar a direção de busca é dada por (3.7).

Se substituirmos $\varphi(x)$ por $\nabla f(x)$ na equação (3.6) temos

$$\nabla f(x) = 0 \quad (3.8)$$

Podemos aplicar o Método de Newton em (3.8) para encontrar os pontos críticos de f , logo, podemos resolver problemas de minimização de funções utilizando este método.

Assim, substituindo $\varphi(x)$ por $\nabla f(x)$ na equação (3.7) obtemos

$$\nabla f(x^k) = -Hf(x)d_k \quad (3.9)$$

3.3.1.1.1 Algoritmo

Seja $x^k \in \mathbb{R}^n$ o ponto na k -ésima iteração, t_k o tamanho do passo na k -ésima iteração, $\nabla f(x^k)$ o gradiente de f no ponto x^k , $Hf(x^k)$ a hessiana de f no ponto x^k .

Considere k tal que $\nabla f(x^k) \neq 0$.

- **Passo 1:** Calcular d_k sendo que $\nabla f(x^k) = -Hf(x)d_k$;
- **Passo 2** (busca linear): Considerar $t_1 = 1$ e determinar t_k de modo que $f(x^k + t d_k) < f(x^k) + t \alpha \nabla f(x^k) \cdot d_k$, onde o parâmetro $\alpha \in (0,1)$.
- **Passo 3:** fazer $x^{k+1} = x^k + t_k d_k$ e $k = k+1$;

Critério de parada: $\|\nabla f(x^k)\| \leq \varepsilon$;

3.3.1.1.2. Algumas limitações

Existem algumas dificuldades encontradas nesses métodos. Primeiramente, a convergência global não é assegurada e a direção de busca pode não ser uma direção viável.

Outra dificuldade é encontrar a hessiana da função, pois, na maioria das vezes é um processo computacionalmente caro. Se o número de variáveis for muito grande, a memória necessária para armazenar esta informação pode ser insuficiente e este processo torna-se inviável.

Para maiores detalhes dos Métodos de Newton, consulte [5],[9], [26], [31] e [43].

3.3.1.2. Métodos Gradientes

Os Métodos Gradientes, também chamados de Métodos de Máxima Descida, são utilizados para resolver problemas de minimização sem restrições, cuja direção de busca em cada iteração é oposta à direção do gradiente, ou seja, $d^k = -\nabla f(x^k)$. Essa direção é uma direção de descida.

3.3.1.2.1. Algoritmo

Seja $x^k \in \mathbb{R}^n$, $\varepsilon > 0$ tal que $\nabla f(x^k) \neq 0$.

- **Passo 1:** Calcular $d_k = -\nabla f(x^k)$;
- **Passo 2:** (busca linear exata). Determinar t_k , que minimiza $f(x^k + td_k)$ sujeito a $t > 0$, ou seja, determinar t_k de modo que $f(x^k + td_k) < f(x^k)$;
- **Passo 3:** fazer $x^{k+1} = x^k + t_k d_k$ e $k=k+1$;

Critério de parada: $\|\nabla f(x^k)\| \leq \varepsilon$.

3.3.1.2.2. Interpretação Geométrica

O Método Gradiente move-se em passos perpendiculares, como ilustrado na figura

3.5.

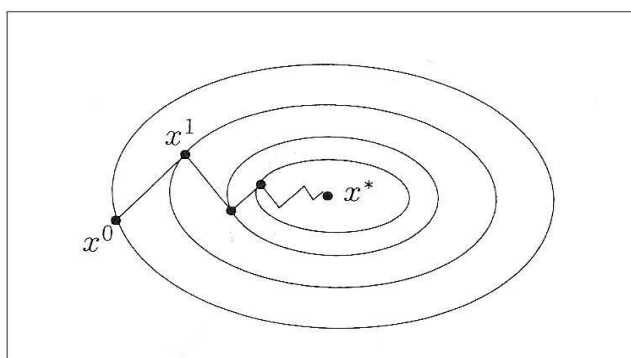


Figura 3.5

3.3.1.2.3. Convergência do Método Gradiente

O algoritmo é globalmente convergente, ou seja, converge para um mínimo local de f a partir de qualquer ponto inicial dado. Porém, a taxa de convergência é muito baixa.

Com o intuito de melhorar a taxa de convergência, foram elaborados outros métodos baseados nos Métodos Gradientes. Dentre eles, destacam-se os Métodos de Gradientes Conjugados, que veremos a seguir.

Em [9], [21], [26], e [43] encontra-se um estudo mais detalhado sobre os Métodos Gradientes.

3.3.1.3. Métodos de Gradientes Conjugados.

Os Métodos de Gradiente Conjugados são baseados nos Métodos Gradientes. A direção de busca é obtida pela soma da direção oposta ao gradiente com a direção anterior. Assim, a nova direção será:

$$d_k = -\nabla f(x^k) + t_k d_{k-1}$$

no qual, t_k é definido como, por exemplo, por

$$t_1 = 0 \text{ e } t_k = \frac{\|f(x^k)\|^2}{\|f(x^{k-1})\|^2} \text{ para } k > 0.$$

O Método de Gradiente Conjugado é um método bastante eficiente e satisfatório especialmente para problemas de grande escala.

Para um estudo mais aprofundado ver [5], [9], [21] e [43].

3.3.2. Métodos de otimização para Problemas com Restrições

Apresentaremos agora, alguns métodos para resolver problemas com restrições.

3.3.2.1. Método de Gradiente Projetado

O Método de Gradiente Projetado foi elaborado para resolver problemas com restrições, no qual, projeta-se o oposto do gradiente de tal modo que melhora-se a função objetivo e mantém-se a viabilidade. Assim, para obter a direção projeta-se $-\nabla f(x)$, ou seja, $d = -P \cdot \nabla f(x)$, onde P é uma matriz projeção adequada. Em seguida é feita uma busca linear ao longo da direção encontrada para determinar o tamanho do passo.

Para melhor entendimento, veja o exemplo 3.3.2.1.1.

3.3.2.1.1. Exemplo

Figura 3.6. ilustra o Método Gradiente Projetado para o seguinte problema:

$$\begin{cases} \min f(x) \\ \text{sa : } Ax \leq b \end{cases}$$

no qual $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função diferenciável, A é uma matriz $m \times n$, $b \in \mathbb{R}^m$ é um vetor .

Seja x o ponto viável tal que $A_1 x \leq b_1$ e $A_2 x \leq b_2$ onde $A^t = (A_1^t, A_2^t)$ e $b^t = (b_1^t, b_2^t)$.

Utilizamos a projeção $P = I - A_1^t (A_1 A_1^t)^{-1} A_1$.

Bazaraa, em [5], prova que se a projeção é dada pela fórmula descrita anteriormente, então d é uma direção viável.

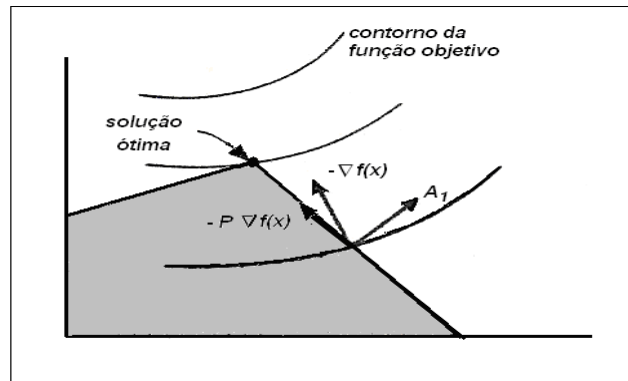


Figura 3.6

Para melhores esclarecimentos sobre o Método Gradiente Projetado, consulte [5], [9], [26] e [43].

3.3.2.2. Métodos de Penalidades

Os Métodos de Penalidades, utilizados para resolver problemas com restrições, são métodos iterativos que a cada iteração transformam o problema original em outro problema que é irrestrito. Ou seja, inicialmente tem-se um problema com restrição, no qual, modificam-se as restrições e as incorpora na função objetivo, obtendo-se assim, um problema sem restrições.

Consideramos, para facilitar a exemplificação, o seguinte problema:

$$(PNL_2) \begin{cases} \text{Minimizar } f(x) \\ \text{s.a. } h(x) = 0 \end{cases} \quad (3.3)$$

Com $f: IR^n \rightarrow IR$, $h: IR^n \rightarrow IR^p$ e, no qual, f e h são funções diferenciáveis e $h(x) = (h_1(x), h_2(x), \dots, h_p(x))$.

Transformaremos (PNL_2) em um problema auxiliar irrestrito formado pela função objetivo e um múltiplo das restrições do problema original. Isto é,

$$\varphi(x, \mu) = f(x) + \mu \sum_{i=1}^m (h_i(x))^2$$

no qual o parâmetro $\mu > 0$.

A idéia central desses métodos é que se μ cresce indefinidamente, a solução de $\varphi(x, \mu)$ será cada vez mais próxima da solução de (PNL_2) . Mas, segundo [26], na prática, quando μ é muito grande, pode-se obter resultados não confiáveis na resolução do problema irrestrito. Para contornar esta dificuldade, foram introduzidas modificações nos Métodos de Penalidade. Assim, surgiu o Método Lagrangeano Aumentado, no qual, a função objetivo é dada da seguinte forma:

$$\varphi(x, \lambda, \mu) = f(x) + \mu \sum_{i=1}^m \lambda_i h_i(x) + \mu \sum_{i=1}^m (h_i(x))^2$$

3.3.2.3. Métodos de Barreira

Os Métodos de Barreira, em geral, são utilizados para resolver problemas com restrições de desigualdade. Ou seja, problemas do tipo:

$$(\text{PNL}_3) \begin{cases} \text{Minimizar } f(x) \\ \text{s.a. } g(x) \leq 0 \end{cases} \quad (3.4)$$

com $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$, no qual, f e g são funções diferenciáveis e $g(x) = (g_1(x), g_2(x), \dots, g_m(x))$. A região viável deve ter interior não vazio.

A principal diferença entre o Método de Penalidade e o Método de Barreira, é que no primeiro método, as aproximações sucessivas das soluções podem não ser viáveis, já no segundo, elas sempre serão viáveis [26]. Assim, a barreira faz com que as iterações permaneçam no interior da região viável.

Nesse método, uma barreira muito utilizada é barreira logarítmica, no qual a função objetivo é representada da seguinte forma:

$$f_\mu(x) = f(x) - \mu \sum_{i=1}^n \log x^i$$

cujo f_μ é estritamente convexa em \mathbb{R}_+^n e o parâmetro $\mu > 0$.

Em [5] e [43] encontra-se mais detalhes sobre esses métodos.

3.3.2.4. Métodos de Direções Viáveis

Nos Métodos de Direções Viáveis, em cada iteração, a direção de busca é uma direção viável. Dado um ponto viável x^k , determina-se uma direção de busca d^k , de descida, e um tamanho de passo $t^k > 0$, de modo que o próximo ponto x^{k+1} também seja viável. Assim, obtém-se uma seqüência de pontos viáveis.

Para um estudo mais detalhado, consulte [31] e [5]

3.3.2.5. Métodos de Pontos Interiores

Os Métodos de Pontos Interiores têm sido usados com sucesso para resolver Problemas de Programação Linear. Estes métodos são eficientes porque eles convergem, relativamente, em poucas iterações. Entretanto, uma iteração de um Método de Pontos Interiores é mais cara do que a iteração do Método Simplex.

Nos Método de Pontos Interiores, dado um ponto viável interior, gera-se novos pontos interiores em uma vizinhança de uma trajetória central até atingir certa tolerância para uma solução ótima.

Seu bom desempenho na prática e suas propriedades teóricas têm motivado a implementação de códigos sofisticados para resolver problemas de grande porte.

Herskovits [17] propôs um Método de Pontos Interiores e Direções Viáveis para solução de problemas com restrição, onde, em cada iteração uma direção de descida é calculada resolvendo-se dois sistemas de equações lineares com a mesma matriz . O algoritmo converge globalmente, é de fácil implementação e não necessita de solução de problemas quadráticos.

3.3.2.6. Método Afim Escala

O Algoritmo Afim Escala foi desenvolvido inicialmente para resolver problemas de programação linear, ou seja,

$$(PPL) \left\{ \begin{array}{l} \text{Minimizar } f(x) \\ sa : Ax = b \\ x \geq 0 \end{array} \right.$$

onde $A \in \mathbb{R}^{m \times n}$ tem posto m .

Considere o conjunto viável $P = \{x; Ax = b, x > 0\}$. Como minimizar f em um poliedro P pode ser um trabalho difícil, utiliza-se minimizar f em um elipsóide.

Geometricamente, podemos descrevê-lo da seguinte forma:

1. Considera-se o ponto inicial $x^0 \in P^0$;
2. Forma-se um elipsóide S_k centrado em x^k , com $S_k \subset P^0$;
3. Encontra-se $x^{k+1} \in \arg \min_{x \in S_k} f(x)$;

Se x^{k+1} é a solução ótima, pare. Senão, vá para 2.

Assim, cada iteração inicia com um ponto viável x^k , formando um elipsóide S_k centrado em x^k com $S_k \subset P^0$ e minimiza a função objetivo em uma região viável, gerando uma direção de descida d , seguido de uma busca unidirecional ao longo de t , a partir de x^k , para determinar o tamanho do passo.

A figura 5.5 mostra o processo do algoritmo Afim Escala.

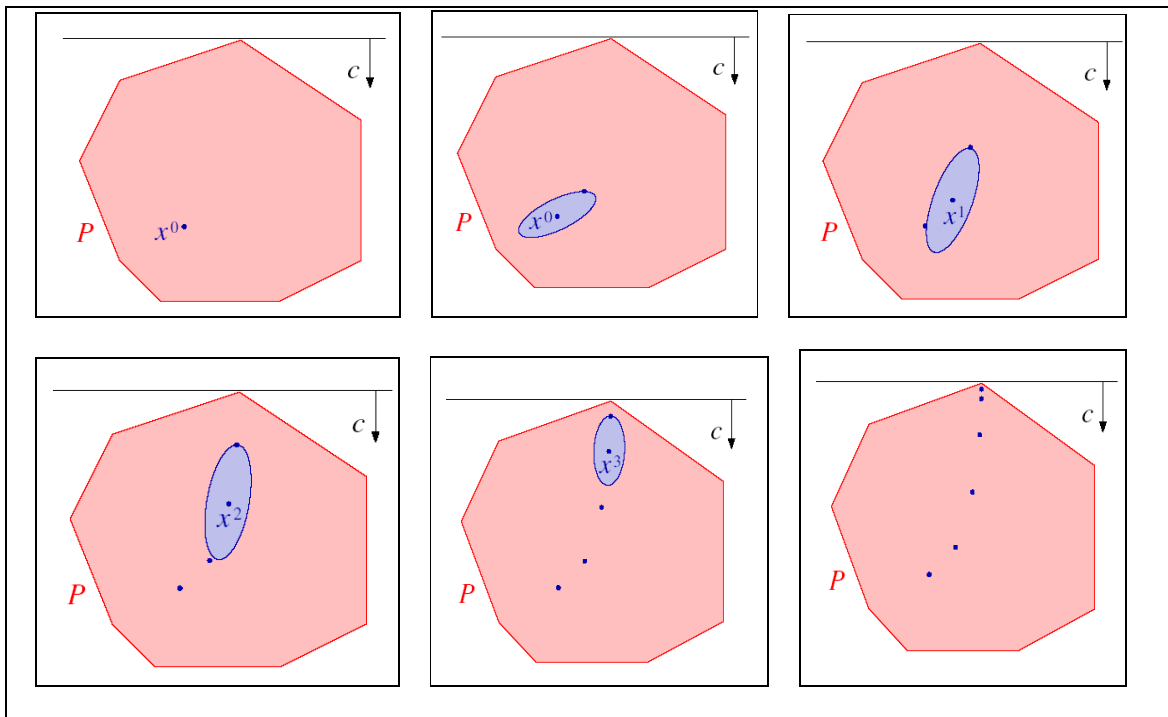


Figura 5.5

O maior trabalho desse método é o do cálculo do novo ponto x^k , que consiste em minimizar, em um hiperplano, uma função linear num elipsóide.

Segundo Jussiani em [34], o algoritmo resolve uma seqüência de problemas mais simples cujas soluções devem convergir para uma solução ótima do problema

3.3.3. Métodos Quase-Newton

Os Métodos Quase-Newton utilizam uma aproximação da matriz hessiana de modo que o trabalho computacional resultante é menor que o do Método de Newton. Dentre eles destacam-se o Método DFP e o Método BFGS. Nesses métodos, é dada uma matriz simétrica inicial como, por exemplo, a matriz identidade, e a cada iteração a matriz é atualizada através de uma fórmula.

3.3.3.1. DFP

Este método de atualização da matriz hessiana foi proposto por Davidon, Fletcher e Powell [3].

Considere $d_k = -B \nabla f(x^k)$, a atualização é feita da seguinte forma:

$$B^{k+1} = B^k + \frac{\gamma \gamma^t}{\delta^t \gamma} - \frac{B^k \delta \delta^t B^k}{\delta^t B^k \delta}$$

no qual $\delta = x^{k+1} - x^k$ e $\gamma = \nabla f(x^{k+1}) - \nabla f(x^k)$.

A fórmula para atualizar B é recursiva, e como pode ser observado, é obtida através de uma correção de B^k .

3.3.3.2. BFGS

Outra fórmula bastante utilizada devido ao seu bom desempenho é a atualização de Broyden-Fletcher-Goldfarb-Shanno. Para atualizar a hessiana a partir de B é calculada a inversa, ou seja, em vez de encontrarmos B , podemos encontrar a inversa de B . Considere

$d_k = -B^{-1} \nabla f(x^k)$, a atualização de B é feita de acordo com a seguinte formula:

$$B^{k+1} = B^k + \left(\frac{1 + \delta^t B^k \delta}{\delta^t \gamma} \right) \frac{\gamma \lambda^t}{\gamma^t \delta} + \frac{\gamma \delta B^k + B^k \delta \gamma^t}{\delta^t \gamma}$$

no qual $\delta = x^{k+1} - x^k$ e $\gamma = \nabla f(x^{k+1}) - \nabla f(x^k)$.

A vantagem em termos de trabalho computacional dos métodos DFP e BFGS é que o número de operações para determinar d é menor do que no método de Newton.

Para um estudo mais detalhado dos Métodos Quase-Newton, consulte [3], [9], [19], [31] e [43].

Capítulo 4

Introdução aos Métodos Não Diferenciáveis

No capítulo anterior vimos as condições de otimalidade para problemas de otimização não linear diferenciável e alguns métodos para resolver tais problemas. A não diferenciabilidade gera uma série de dificuldades que acarretam na inviabilidade de aplicação dos algoritmos para problemas diferenciáveis em problemas não diferenciáveis.

A primeira dificuldade aparece na etapa que determina a direção de busca, pois a direção obtida não é necessariamente uma direção de descida, e conseqüentemente, a busca linear não pode ser aplicada. Outra dificuldade é estabelecer um critério de parada que, na maioria das vezes, não fica bem definido. Por exemplo, o critério de parada que utiliza $\|\nabla f(x^k)\| \leq \varepsilon$, para algum $\varepsilon > 0$ não pode ser aplicado em problemas não diferenciáveis. Veja o exemplo: considere $f: \mathbb{R} \rightarrow \mathbb{R}$ definida por $f(x) = |x|$. Tem-se $\|\nabla f(x^k)\| = 1$, $\forall x^k \neq 0$, não importando o quanto x^k esteja próximo da solução $x^* = 0$. Portanto, não é possível gerar uma seqüência de gradientes que tende para zero.

4.1 Métodos de Otimização Não Diferenciável.

Entre os métodos para resolver problemas de otimização não diferenciáveis se destacam os Métodos Subgradientes, os Métodos de Planos de Corte e os Métodos de Feixes.

Neste capítulo, faremos uma breve exposição acerca dos principais métodos para resolver problemas não diferenciáveis. Consideraremos a hipótese de que para todo $x \in \mathbb{R}^n$, pode-se calcular o valor da função f em x e um subgradiente $g \in \partial f(x)$.

Nosso foco será no seguinte problema:

$$(P) \begin{cases} \text{Minimizar } f(x) \\ x \in \mathbb{R}^n \end{cases}$$

com $f: \mathbb{R}^n \rightarrow \mathbb{R}$ convexa e não diferenciável.

4.1.1. Métodos Subgradientes

Os Métodos Subgradientes, que surgiram na década de 60 na antiga União Soviética, foram os primeiros métodos criados para resolver problemas de otimização não diferenciáveis.

Este método, baseado no Método Gradiente para problemas diferenciáveis, utiliza a direção oposta a um subgradiente como a direção de busca.

Os Métodos Subgradientes possuem uma estrutura extremamente simples mas, não apresentam bons resultados numéricos.

Recentemente foram publicados vários trabalhos que utilizam os Métodos Subgradientes [6], [13], [25], [27], [41], [47], [48], [49] e [51], o que nos motivou a estudar esses métodos. O objetivo principal deste trabalho é estudar com mais detalhes os Métodos Subgradientes, o que faremos no capítulo 5.

4.1.2. Métodos de Planos de Corte

Os Métodos de Planos de Corte utilizam subgradientes em um dado ponto x^i para construir uma aproximação linear para uma função convexa não linear f , isto é:

$$f(x) \cong \max\{f(x^i) + g^i \cdot (x - x^i)\}, \text{ onde } g^i \in \partial f(x^i) \text{ e } i = 1, 2, \dots$$

Assim, o problema (P) é substituído por:

$$(P_1) \begin{cases} \min \max\{f(x^i) + g^i \cdot (x - x^i), i = 1, 2, \dots\} \\ x \in \mathbb{R}^n \end{cases}$$

que é equivalente a:

$$(P_2) \begin{cases} \text{Minimizar } z \\ \text{sa: } f(x^i) + g^i \cdot (x - x^i) \leq z, i = 1, 2, \dots \end{cases}$$

Os Métodos de Planos de Corte são baseados em progressivos refinamentos de aproximações poliedrais do epigrafo de f . Portanto, o problema (P_2) é linear cuja solução (caso exista) é apenas uma aproximação da solução de (P) . Entretanto, quanto mais planos de corte são acrescentados, mais precisa fica a aproximação da solução.

4.1.2.1 Algoritmo

Seja x^1 o ponto inicial, x^k o ponto na k -ésima iteração, t_k o tamanho do passo nessa iteração, g^k um subgradiente da função no ponto x^k , e $\partial f(x^k)$ o conjunto de todos os g^k . Considere uma dada tolerância $\text{tol} \geq 0$, um conjunto S não vazio compacto convexo que contém o ponto mínimo de f e a função $\tilde{f}(x) \equiv \max\{f(x^i) + g^i \cdot (x - x^i), i = 1, 2, \dots\}$.

- **1º passo:** Escolher $x^1 \in S$ e tomar $k=1$. Defina $\tilde{f}_1 \equiv -\infty$.
- **2º passo:** Obter: $g^k \in \partial f(x^k)$ e $f(x^k)$ e computar o decréscimo nominal

$$\alpha_k \equiv f(x^k) - \tilde{f}_{k-1}(x^k)$$

Critério de parada: $\alpha_k \leq \text{tol}$

- **3º passo:** Escolher t_k apropriado, por exemplo, $t_k = 1$.

- **4º passo:** Encontrar $d^k \in \arg \min_{x^k + d \in S} \tilde{f}_k(x^k + d)$.
- **5º passo:** Definir $x^{k+1} = x^k - t_k \cdot d^k$ e $k=k+1$. Retornar para o 2º passo.

Note que o algoritmo de planos de corte utiliza além das informações atuais, as informações obtidas em iterações anteriores.

A figura 4.1 ilustra o processo do algoritmo de planos de corte em um exemplo simples.

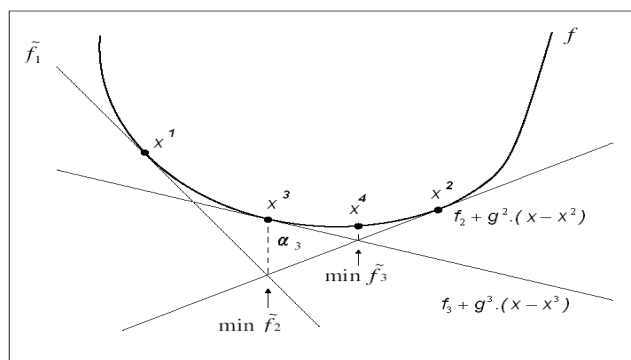


figura 4.1

Um problema encontrado nesses métodos é que um número crescente de funções afins que definem o modelo se acumula, dificultando a resolução do problema.

Outro problema encontrado é que o método não garante o decréscimo da função objetivo. Tal fato pode ser observado na figura 4.2, quando introduzimos um modelo f_3 de uma função afim quase “horizontal”. Note que no ponto x^3 foi encontrado um $\alpha_3 > tol$, portanto, o processo continua obtendo $f(x^4) < f(x^3)$.

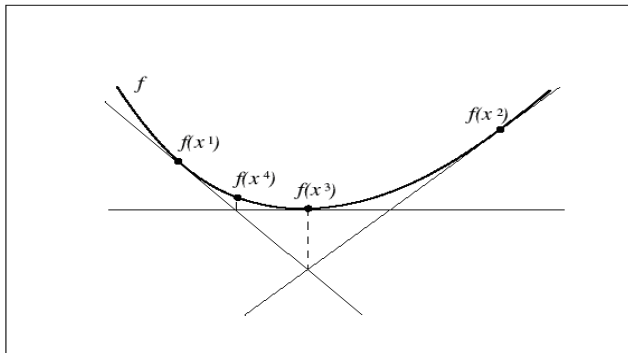


figura 4.2

Nos últimos anos foram propostos vários métodos de Plano de Corte, como em [9] e [3]. Tais métodos se distinguem pela maneira de escolher os pontos para gerar os planos de corte. O método clássico proposto por Kelley [19], utiliza as soluções do problema (P_2) para gerar os planos de corte.

Os Métodos de Feixes e de Centros Analíticos também utilizam planos de corte.

4.1.3. Métodos de Centros Analíticos

Vários algoritmos, para resolver problemas de otimização, utilizam em cada iteração um ponto considerado como centro, ou no sentido geométrico ou no sentido analítico.

Em geral, os métodos baseados na noção de centro realizam um corte em uma região convexa, limitada e que contenha solução ótima. Cada corte utiliza um hiperplano, que separa esta região em duas partes. Na parte onde a solução do problema está contida são realizadas operações, de acordo com o método que está sendo utilizado, para determinar o próximo centro.

Os vários métodos existentes se diferenciam na forma como determinam os centros. A velocidade e a convergência do método dependem do processo adotado para determinar o centro.

Um bom estudo desse assunto pode ser encontrado em [4], [28], [29] e [61].

Dos métodos baseados em centros analíticos, um que mais se destaca é o Método de Centros Analíticos e Planos de Cortes, também conhecido como ACCPM.

4.1.4. Métodos de Centros Analíticos e Planos de Corte

Considere o problema (P). Seja $S = \{x \in \mathbb{R}^n; a_i \cdot x \leq b_i, i = 1, \dots, m\}$ no qual $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, $i = 1, \dots, m$, um conjunto convexo, compacto, com interior não vazio e limitado, contendo pelo menos uma solução de (P).

Aplicando o Método de Planos de Corte no problema (P) para aproximarmos a função f , como vimos na seção 4.1.2, consideramos $\tilde{f}^i(x) \equiv \max\{f(x^i) + g \cdot (x - x^i)\}$, $i = 1, \dots, k$.

Após a k -ésima iteração, encontraremos o ponto x^{k+1} resolvendo o problema:

$$(P_3) \begin{cases} \text{Minimizar } \tilde{f}^k(x) \\ x \in S \end{cases}$$

que é equivalente a:

$$(P_4) \begin{cases} \text{Minimizar } r \\ r \geq \tilde{f}^k(x) \\ x \in S, r \in \mathbb{R} \end{cases}$$

Definindo $S^i = \{(x, r) \in \mathbb{R}^{n+1}; x \in S, f(x^i) + g^i \cdot (x - x^i) \leq r, i = 1, \dots, k\}$, podemos reescrever o problema (P_4) da seguinte forma:

$$(P_5) \begin{cases} \text{Minimizar } r \\ (x, r) \in S^k \end{cases}$$

O método de centro analítico associa ao problema (P_5) um problema não linear definido por:

$$(P_6) \begin{cases} \text{Minimizar } f_{\rho^k}(x, r) \\ (x, r) \in (S_{\rho^k})^0 \end{cases}$$

$$f_{\rho^k}(x, r) = -\ln(\rho^k - r) - \sum_{j=1}^m \ln(b_j - a_j \cdot x) - \sum_{j=1}^k \ln(r - f(x^j) - s^j \cdot (x - x^j))$$

$$S_{\rho^k} = S^k \cap \{(x, r) \in \mathbb{R}^{n+1}; r \leq \rho^k\}$$

e ρ^k é uma cota superior para o valor corrente de r .

A solução do problema (P_6) define o próximo ponto x^{k+1} . A cada iteração, encontra-se o centro analítico da região atual e o limitante inferior ρ é acrescido. As restrições relacionadas à função objetivo são removidas e outras restrições associadas ao novo limitante são incorporadas, definindo uma nova região. Este procedimento é repetido até que seja alcançado um ponto próximo a trajetória central e suficientemente perto da solução ótima. Assim, obtém-se uma seqüência que converge à solução do problema (P_5) .

Existem várias versões dos Métodos de Centros Analíticos e Planos de Cortes que podem ser encontradas com mais detalhes em [4], [27], [28] e [57]

4.1.5. Métodos de Feixes

Os Métodos de Feixes, propostos por Mifflin [46] e Lemarechal [39], são atualmente os que apresentam melhores resultados.

Esses métodos exploram as informações obtidas nas iterações anteriores acumulando-as ao chamado “feixe de informações”. Eles determinam os planos de corte a partir das soluções de problemas quadráticos construídos considerando-se alguns subgradientes calculados em iterações anteriores. A direção de busca é obtida através da envoltória convexa de um conjunto de subgradiente gerados anteriormente.

Uma característica dos Métodos de Feixes é que eles utilizam o conceito de passo sério e passo nulo. Ou seja, calculada a direção d^k , sejam $\varepsilon > 0$ e $y^{k+1} = x^k + t_k d^k$ para

algum $t_k > 0$. Se $f(y^{k+1}) \leq f(x^k) - \varepsilon_k$ então $x^{k+1} = y^{k+1}$ (passo sério), senão $x^{k+1} = x^k$ (passo nulo). Em qualquer caso calcula-se o $g^{k+1} \in \partial f(y^{k+1})$ e incorpora-se esse subgradiente ao feixe.

Os Métodos de Feixes geram uma seqüência de iterações para que os valores da função objetivos sejam monótonos decrescentes. Por esta razão, esses métodos são classificados como "métodos de descida". No entanto, uma dificuldade encontrada nesses métodos é que eles exigem a solução de um subproblema quadrático em cada iteração para encontrar a direção de busca, e isto pode tornar-se bastante caro, em particular para problemas maiores.

Para um estudo mais detalhado, consulte [38], [39], [44] e [46].

4.1.6 . Método de Direções Viáveis

Recentemente, Passarella, em sua Tese de Doutorado [23], propôs um algoritmo que utiliza pontos interiores e direções viáveis para resolver problemas convexos não diferenciáveis irrestritos. Este algoritmo, denominado Algoritmo de Direções Viáveis, determina o mínimo de funções convexas não necessariamente diferenciáveis , para o problema irrestrito:

$$(PND) \begin{cases} \min F(x) \\ x \in \mathbb{R}^n \end{cases}$$

onde $F : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa e não necessariamente diferenciável.

Primeiramente, substitui-se o problema original (PND) pelo problema equivalente $\min z \in \mathbb{R}$ tal que $F(x) \leq z$. Assim, substitui-se (PND) pela minimização de uma função linear com uma restrição não diferenciável. Em seguida, constroe-se uma seqüência de problemas auxiliares, onde a restrição do problema equivalente é aproximada por hiperplanos de apoio ao epígrafo da função objetivo F . Em cada iteração uma direção de busca para o problema auxiliar é calculada resolvendo-se dois sistemas de equações lineares com a mesma matriz,

obtidos do problema auxiliar. A busca linear é substituída por uma estratégia equivalente à utilizada nos Métodos de Feixes. Constroe-se uma seqüência de pontos interiores ao epígrafo de F cujos pontos de acumulação são soluções do problema original.

Este método possui uma estrutura simples, é de fácil implementação computacional e não requer a solução de problemas quadráticos.

Capítulo 5

Métodos subgradientes, suas Extensões e Trabalhos Recentes

5.1. Métodos Subgradientes

Os Métodos Subgradientes são pioneiros em otimização não diferenciável. Foram originalmente desenvolvidos por Shor [63], na União Soviética, nas décadas de 60 e 70.

Esses métodos, também chamados de Métodos Gradientes Generalizados [40], são uma generalização dos Métodos Gradientes no qual o gradiente da função é substituído por um subgradiente para obter uma nova direção de busca. Possuem uma estrutura muito simples que não utiliza busca linear. O tamanho do passo pode ser fixado ou pode mudar com as iterações, mas dependendo do passo escolhido não se tem a garantia de convergência global.

5.1.1. O Método Subgradiente

Considere novamente o problema:

$$(P) \begin{cases} \text{Minimizar } f(x) \\ x \in \mathbb{R}^n \end{cases} \quad (5.1)$$

Com $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convexa e não diferenciável.

O Método Subgradiente é um algoritmo que gera uma seqüência de pontos $\{x^k\}_{k=1}^{\infty}$ da forma:

$$x^{k+1} = x^k - t_k \frac{g^k}{\|g^k\|} \quad (5.2)$$

No qual, x^k é o ponto na k -ésima iteração, t_k representa o tamanho do passo nessa iteração e $g^k \in \partial f(x^k)$ é um subgradiente da função f no ponto x^k .

5.1.2. Dificuldades Encontradas

Uma grande dificuldade encontrada nesses métodos é que a direção oposta ao subgradiente pode não ser uma direção de descida. De fato,

Considere $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ tal que $f(x,y) = |x| + 2|y|$.

Seja $(x^k, y^k) = (1,0)$ um ponto em uma dada iteração k . Primeiramente mostraremos que $g = (1,2) \in \partial f(1,0)$.

Temos

$$f(x,y) = |x| + 2|y| \geq x + 2y = 1 + x - 1 + 2y = f(1,0) + (1,2) \cdot (x-1, y-0)$$

Portanto,

$$g = (1,2) \in \partial f(1,0)$$

Veremos agora que $-g = (-1,-2)$ não é uma direção de descida em $(1,0)$.

De fato,

Seja

$$(x^{k+1}, y^{k+1}) = (1,0) - t_k(1,2)$$

conseqüentemente

$$(x^{k+1}, y^{k+1}) = (1 - t_k, -2t_k).$$

Logo,

$$f(x^{k+1}, y^{k+1}) = |1 - t_k| + 2|-2t_k| = |1 - t_k| + 4|t_k|$$

mas,

$$|1 - t_k| \geq 1 - |t_k| > 1 - 4|t_k| \Rightarrow |1 - t_k| + 4|t_k| > 1$$

logo,

$$f(x^{k+1}, y^{k+1}) = |1 - t_k| + 4|t_k| > 1 = f(1, 0)$$

O que mostra que $-g$ não é uma direção de descida.

Também podemos observar este fato através da figura 5.1 que mostra curvas de nível da função f , cujo ponto de mínimo é $0 \in \mathbb{R}^2$. As direções de descida estão representadas nessa figura em cinza claro. Podemos observar que a direção g é um extremo do cone associado a $\partial f(x)$ e que $-g$ não é uma direção de descida.

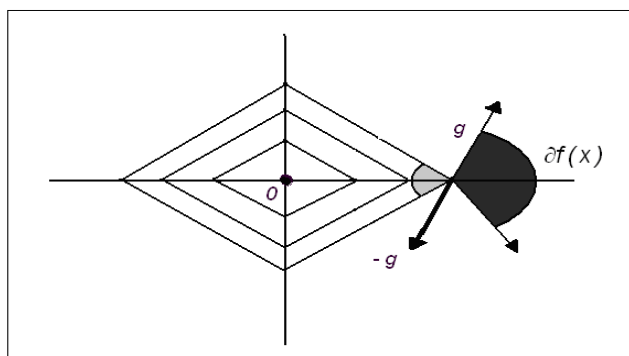


Figura 5.1

Outra dificuldade dos Métodos Subgradientes é estabelecer um critério de parada uma vez que os subgradientes são encontrados arbitrariamente e por isso, não contém informação sobre a condição de otimalidade. De fato, descobrir se $0 \in \partial f(x)$ pode ser inviável pois, na maioria das vezes, é muito difícil encontrar o subdiferencial $\partial f(x)$. Outros testes práticos devem ser aplicados observando a especificidade do problema. Segundo Lima, em [40], um critério de parada para os Métodos Subgradientes pode ser dado pelo número máximo de iterações atingido ou pela condição $x^{K+1} \approx x^k$, ou ainda se $f(x^{K+1}) \approx f(x^k)$.

Um outro grande inconveniente desses métodos é sua baixa velocidade de convergência, o que os torna bastante lentos. Este fato poderá ser observado no capítulo 6.

5.1.3- Escolha do Tamanho do Passo

É fundamental uma escolha adequada do tamanho do passo $t_k > 0$ em cada iteração.

Se os passos forem muito pequenos, x^k se aproximará muito lentamente do ponto ótimo.

Por outro lado, segundo [40], se forem excessivamente largos, o método poderá oscilar desnecessariamente em torno da solução.

Estabeleceremos agora uma condição sobre o passo que garante a convergência global dos Métodos Subgradientes.

Seja x^* solução de (P). Se $g^k \in \partial f(x^k)$ então

$$f(x) - f(x^k) \geq g^k \cdot (x - x^k), \forall x \in \mathbb{R}^n$$

daí

$$f(x^k) - f(x) \leq -g^k \cdot (x - x^k), \forall x \in \mathbb{R}^n$$

segue-se que

$$-g^k \cdot (x^* - x^k) \geq f(x^k) - f(x^*).$$

Como x^* é solução de (P) segue que

$$f(x^k) \geq f(x^*) \text{ ou } f(x^k) - f(x^*) \geq 0.$$

Portanto, $-g^k \cdot (x^* - x^k) \geq 0$, conseqüentemente, o ângulo entre $-g^k$ e $x^* - x^k$ é agudo.

Veja agora a figura 5.2.

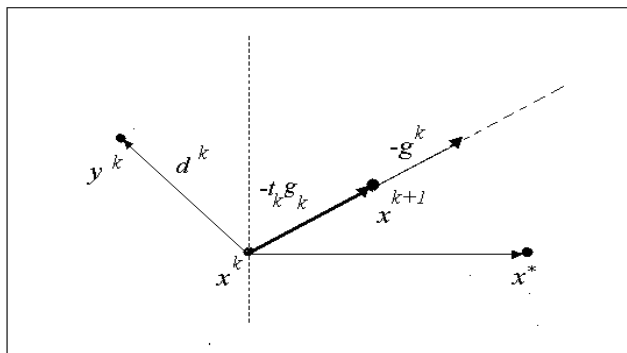


Figura 5.2

Para t_k pequeno é fácil ver que

$$\|x^{k+1} - x^*\| < \|x^k - x^*\|$$

Por outro lado, se o ângulo entre d^k e $x^k - x^*$ for obtuso tem-se

$$\|y^k - x^*\| > \|x^k - x^*\|$$

independentemente da escolha de t_k . Assim, para $t_k > 0$ suficientemente pequeno, tem-se que x^{k+1} está mais próximo da solução x^* do que x^k . Isto motiva escolher a seqüência $\{t_k\}$ satisfazendo $\lim_{k \rightarrow \infty} t_k = 0$. Mais precisamente, vale o resultado abaixo cuja demonstração encontra-se em [38].

Teorema 5.1

Seja x^* uma solução de (P). Então $\|x^{k+1} - x^*\| < \|x^k - x^*\|$ sempre que

$$0 < t_k < \frac{2[f(x_k) - f(x^*)]}{\|g_k\|}.$$

Mas somente essa condição para a escolha de t_k não garante a convergência do método. De fato, se $\sum_{k=1}^{\infty} t_k$ é convergente, ou seja, $\sum_{k=1}^{\infty} t_k = r$, a seqüência de $\{x^k\}_{k=1}^{\infty}$ tem limite, mas esse limite pode não pertencer ao conjunto X^* das soluções de (P).

Com efeito,

$$\begin{aligned} \|x^k - x^1\| &= \|x^k - x^{k-1} + x^{k-1} - x^1\| \leq \|x^k - x^{k-1}\| + \|x^{k-1} - x^1\| = \\ &= \|x^k - x^{k-1}\| + \|x^{k-1} - x^{k-2} + x^{k-2} - x^1\| \leq \\ &\leq \|x^k - x^{k-1}\| + \|x^{k-1} - x^{k-2}\| + \|x^{k-2} - x^1\| \leq \\ &\leq \|x^k - x^{k-1}\| + \|x^{k-1} - x^{k-2}\| + \dots + \|x^3 - x^2\| + \|x^2 - x^1\| \end{aligned}$$

como

$$x^{k+1} = x^k - t_k \frac{g^k}{\|g^k\|} \Rightarrow \|x^{k+1} - x^k\| = \|t_k\| \cdot \left\| \frac{g^k}{\|g^k\|} \right\| \Rightarrow \|x^{k+1} - x^k\| = t_k$$

então

$$\|x^k - x^1\| \leq t_{k-1} + t_{k-2} \dots + t_2 + t_1 = \sum_{j=1}^{k-1} t_j$$

conseqüentemente

$$\|x^k - x^1\| \leq r,$$

Assim, o método encontrará solução x^* se, e somente se, $\forall k \in \mathbb{N}$ tem-se $x^* \in B(x^1, r)$, no qual x^1 representa o ponto inicial. Para contornar este problema, impõe-se

que $\sum_{k=1}^{\infty} t_k = \infty$.

Teorema 5.2

Seja X^* o conjunto de todas as soluções de (P), limitado e não vazio. Então para qualquer ponto inicial x^1 , a seqüência $\{x^k\}_{k \in \mathbb{N}}$ definida por $x^{k+1} = x^k - t_k \frac{g^k}{\|g^k\|}$ com t_k

satisfazendo $\lim_{k \rightarrow \infty} t_k = 0$ e $\sum_{k=1}^{\infty} t_k = +\infty$ é limitada e todos os seus pontos de acumulação

pertencem a X^* .

A demonstração do teorema acima se encontra em [38].

Assim, para garantir a convergência global do método, pode-se escolher o tamanho do passo t_k , chamado de Passo da Série Divergente [59], de forma que $\lim_{k \rightarrow \infty} t_k = 0$ e

$\sum_{k=1}^{\infty} t_k = +\infty$. Uma sugestão apresentada em [13] é utilizar $t_k = \frac{a}{\sqrt{k}}$ com $a \geq 0$.

Apesar do Passo da Série Divergente ser um dos mais utilizados, existe uma série de variantes dos Métodos Subgradientes que diferem entre si essencialmente na forma de calcular o tamanho do passo ao longo do processo.

Shor, em [63], apresenta o tamanho do passo que não utiliza o subgradiente normalizado, ou seja, tem-se $x^{k+1} = x^k - t_k g^k$. Shor [63] também mostra que se t_k satisfaz as condições do teorema 5.2, a seqüência $\{x^k\}_{k=1}^{\infty}$ é limitada e algoritmo é convergente, ou a seqüência $\{x^k\}_{k=1}^{\infty}$ é ilimitada e o algoritmo diverge, e que se a função f é linear por partes com um número finito de partes, então o algoritmo é globalmente convergente.

Nedic e Bertsekas em [49], apresentam alguns tamanhos de passos, entre eles, o *tamanho de passo constante* definido por $t_k = t$ para todo $k \in \mathbb{N}$ e *passos longos*

determinados por $t = \frac{\|x^{k+1} - x^k\|}{\|g^k\|}$. A convergência, nos dois casos, é garantida dentro de

alguma escala do valor ótimo, isto é, o Método Subgradiente encontra um ponto sub-ótimo com um número finito de passos.

Shor, partindo da idéia de B.T.Polyak, sugere em [63], uma escolha do tamanho do passo t quando se conhece a solução f^* . Propõe $t = \frac{\gamma[f(x^k) - f^*]}{\|g^k\|^2}$, no qual $0 < \gamma < 2$.

Mostra que o método converge com velocidade de uma progressão geométrica. Quando

$f(x^*)$ for desconhecido, Shor [63] sugere $t = \frac{\gamma[f(x^k) - \bar{f}]}{\|g^k\|^2}$, de forma que $0 < \gamma < 2$ e \bar{f} seja

a melhor estimativa de $f(x^*)$.

Um grupo de cientistas italianos [14], baseado na idéia de Shor acima, apresenta uma modificação dos métodos subgradientes que utiliza a melhor estimativa do valor mínimo da função e calcula a direção como uma combinação linear dos subgradientes e das

direções usadas em passos anteriores, ou seja, considera-se $t_k = \frac{\gamma_k [f(x^k) - f^*]}{\|s^k\|^2}$, tal que,

$s^k = g^k + \beta_k \cdot s_k$, $0 < \gamma_k \leq 1$ e $\beta_k \geq 0$ para todo $k = 1, 2, \dots$. Para mais detalhes, ver [63] e [64].

5.1.4. Algoritmo

Seja x^1 o ponto inicial, x^k o ponto na k -ésima iteração, t_k o tamanho do passo nessa iteração, g^k um subgradiente da função no ponto x^k .

- **1º passo:** Escolher $x^1 \in \mathbb{R}^n$ e tomar $k=1$.
- **2º passo:** Obter: $g^k \in \partial f(x^k)$ e $f(x^k)$.

Critério de parada: $0 \in \partial f(x^k)$

- **3º passo:** Escolher t_k apropriado.

Uma opção é escolher t_k de modo que $\lim_{k \rightarrow \infty} t_k = 0$ e $\sum_{k=1}^{\infty} t_k = +\infty$.

- **4º passo:** Definir $d^k = -\frac{g^k}{\|g^k\|}$.
- **5º passo:** Definir $x^{k+1} = x^k + t_k \cdot d^k$ e $k=k+1$. Retornar para o 2º passo.

Makela sugere em [44], generalizar os métodos que são mais eficientes do que o Método Gradiente para aumentar a velocidade de convergência. Cita como exemplo os Métodos Quase Newton, porém, substitui o gradiente pelo subgradiente normalizado obtendo a direção $d^k = -\frac{H^k g^k}{g^{kT} H^k g^{k/2}}$, no qual H^k representa a Hessiana da função. No

entanto, não aconselha empregar uma generalização direta dos métodos DFP e BFGS para atualizar a inversa da Hessiana, pois gera resultados numéricos pobres.

5.1.5. Exemplo

Considere o problema não diferenciável: Minimizar $f(x,y) = |x| + |y|$.

A tabela 5.1 apresenta os resultados obtidos através do algoritmo 5.1.4 no qual se utiliza a série harmônica como tamanho do passo, ou seja, $t_k = \frac{1}{k}$. Inicia-se com o ponto $x^1 = (2,0)$.

Iteração	Subgradiente e $g^k \in \partial f(x^k)$	Tamanho do Passo t_k	Direção d^k	Ponto x^{k+1}	Valor de $f(x^{k+1})$
1	(1, 0)	1.0000	(1.0000, 0.0000)	(1.0000, 0.0000)	1.0000
2	(1, 1)	0.5000	(0.7071, 0.7071)	(0.6464, -0.3536)	1.0000
3	(1, -1)	0.3333	(0.7071, -0.7071)	(0.4107, -0.1179)	0.5286
4	(1, -1)	0.2500	(0.7071, -0.7071)	(0.2340, 0.0589)	0.2929
5	(1, 1)	0.2000	(0.7071, 0.7071)	(0.0925, -0.0825)	0.1750
6	(1, -1)	0.1667	(0.7071, -0.7071)	(-0.0253, 0.0354)	0.0607
7	(-1, 1)	0.1429	(-0.7071, 0.7071)	(0.0757, -0.0657)	0.1414
8	(1, -1)	0.1250	(0.7071, -0.7071)	(-0.0127, 0.0227)	0.0354
9	(-1, 1)	0.1111	(-0.7071, 0.7071)	(0.0659, -0.0558)	0.1217
10	(1, -1)	0.1000	(0.7071, -0.7071)	(-0.0048, 0.0149)	0.0197
11	(-1, 1)	0.0909	(-0.7071, 0.7071)	(0.0595, -0.0494)	0.1089
12	(1, -1)	0.0833	(0.7071, -0.7071)	(0.0005, 0.0095)	0.0101
13	(1, 1)	0.0769	(0.7071, 0.7071)	(-0.0539, -0.0449)	0.0987

Tabela 5.1

Note que nas iterações 7, 9, 11 e 13 o valor da função aumentou, portanto, os pontos nas iterações anteriores estavam mais próximos da solução. Isso mostra que a direção pode não ser necessariamente uma direção de descida, o que é um grande problema nos Métodos Subgradientes.

Stephen Boyd e outros no trabalho [13], apresentaram a proposta de armazenar somente os melhores valores da função em cada iteração, ou seja, encontra-se em cada iteração $f_{best}^k = \min\{f_{best}^{k-1}, f(x^k)\} \Leftrightarrow f_{best}^k = \min\{f(x^1), f(x^2), \dots, f(x^k)\}$. Se $f(x^k) = f_{best}^k$ então x^k é o melhor ponto encontrado e $i_{best}^k = k$ é a melhor iteração encontrada. Portanto, as

iterações 7, 9, 11 e 13 são descartadas e a solução seria $x=(0.0005, 0.0095)$. Observe a tabela 5.2.

Iteração (k)	Melhor Iteração i_{best}^k	Ponto x^{k+1}	Ponto x_{best}^{k+1}	Valor de $f(x^{k+1})$	Valor de f_{best}^k
1	1	(1.0000, 0.0000)	(1.0000, 0.0000)	1.0000	1.0000
2	2	(0.6464, -0.3536)	(0.6464, -0.3536)	1.0000	1.0000
3	3	(0.4107, -0.1179)	(0.4107, -0.1179)	0.5286	0.5286
4	4	(0.2340, 0.0589)	(0.2340, 0.0589)	0.2929	0.2929
5	5	(0.0925, -0.0825)	(0.0925, -0.0825)	0.1750	0.1750
6	6	(-0.0253, 0.0354)	(-0.0253, 0.0354)	0.0607	0.0607
7	6	(0.0757, -0.0657)	(-0.0253, 0.0354)	0.1414	0.0607
8	8	(-0.0127, 0.0227)	(-0.0127, 0.0227)	0.0354	0.0354
9	8	(0.0659, -0.0558)	(-0.0127, 0.0227)	0.1217	0.0354
10	10	(-0.0048, 0.0149)	(-0.0048, 0.0149)	0.0197	0.0197
11	10	(0.0595, -0.0494)	(-0.0048, 0.0149)	0.1089	0.0197
12	12	(0.0005, 0.0095)	(0.0005, 0.0095)	0.0101	0.0101
13	12	(-0.0539, -0.449)	(0.0005, 0.0095)	0.0987	0.0101

Tabela 5.2

A figura 5.3 ilustra o processo do Método Subgradiente de acordo com a tabela 5.1.

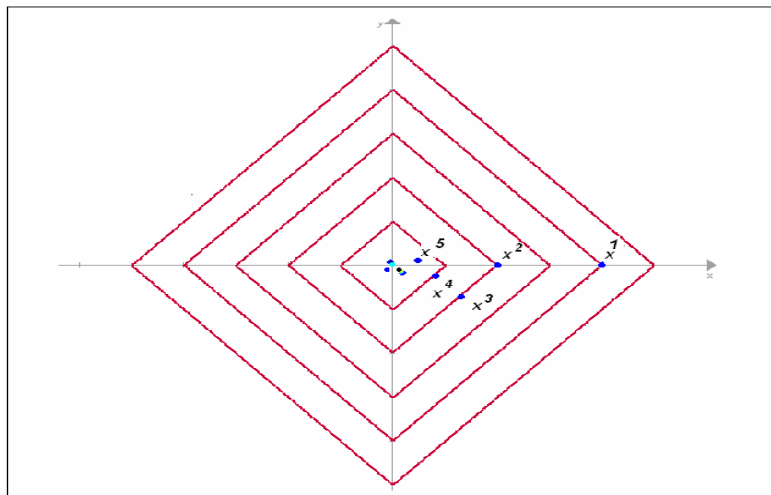


Figura 5.3

Como pode ser observado tanto pelas tabelas, quanto pela figura 5.3, o algoritmo converge para a solução $x^* = (0,0)$.

5.1.6. Uma Aplicação Importante dos Métodos Subgradientes.

Apesar de, em geral, possuir baixa velocidade de convergência, o Método Subgradiente é bastante empregado. Segundo [57], esse método é um dos mais utilizados para resolver Problema Lagrangeano Dual (que será apresentado a seguir).

5.1.6.1. Relaxação Lagrangeana

A Relaxação Lagrangeana é baseada na idéia dos Métodos de Penalidade, no qual, transformam-se problemas mais complicados devido às restrições, em problemas irrestritos, considerados mais simples. Acrescentam-se as restrições à função objetivo através de um vetor de multiplicadores, denominados de Multiplicadores de Lagrange, obtendo-se um problema irrestrito, cujo valor da solução ótima é um limite inferior (em problemas de minimização) para o valor ótimo do problema original.

Considere o seguinte problema:

$$(PP) \left\{ \begin{array}{l} \text{Minimizar } f(x) \\ \text{sa : } g(x) \leq 0 \\ x \in X \\ x = (x_1, \dots, x_n)^t \in Z_+^n \end{array} \right. \quad (5.3)$$

com $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$, no qual, f e g são funções diferenciáveis com $g(x) = (g_1(x), g_2(x), \dots, g_m(x))$, e X é um conjunto discreto finito.

Para cada restrição, associa-se um número real $\mu > 0$, denominado Multiplicador de Lagrange, e define-se a função Lagrangeana como :

$$F(x, \mu) = f(x) + \mu g(x)$$

onde $\mu = (\mu_1, \dots, \mu_m)$ e $g(x) = (g_1(x), g_2(x), \dots, g_m(x))^t$.

Assume-se que a função Lagrangeana acima satisfaz a seguinte condição.

$$L(\mu) = \min_{x \in S} \{F(x, \mu)\}, \quad \forall \mu \in \mathbb{R}_+^m$$

Passando (PP) para a forma dual temos:

$$(PD) \left\{ \begin{array}{l} \text{Max} \\ \mu \in \mathbb{R}_+^m \end{array} L(\mu) \right.$$

ou equivalentemente,

$$(PD) \left\{ \begin{array}{l} \text{Max} \\ \mu \in \mathbb{R}_+^m \end{array} \min_{x \in S} F(x, \mu) \right.$$

O Problema Lagrangeano Dual possui propriedades importantes que são destacadas em [34] e [40].

Lima [40] apresenta um algoritmo para encontrar o μ 'melhorado'. Este mesmo trabalho, sugere a utilização simultânea do procedimento do Subgradiente com heurísticas, para encontrar soluções viáveis 'melhoradas'.

A principal vantagem da utilização dos Métodos Subgradientes para resolver Problema Lagrangeano Dual, em relação aos Métodos Simplex e Branch-and-Bound, é o baixo custo computacional.

5.2. Extensões dos Métodos Subgradientes

5.2.1. Métodos de Dilatação do Espaço

Como já citado anteriormente, a convergência dos algoritmos baseados nos Métodos Subgradientes é geralmente muito lenta e depende do passo escolhido. Com intuito de melhorar a taxa de convergência, Shor [64] sugere um método baseado em uma classe de algoritmos elaborados para minimização de funções convexas em que o movimento na direção do subgradiente é combinado com uma dilatação do espaço ao longo desta direção.

Observe na figura 5.4, que após aplicarmos o operador dilatação B , a distância entre x^{k+1} e x^* é menor do que a distância entre y^k e x^* .

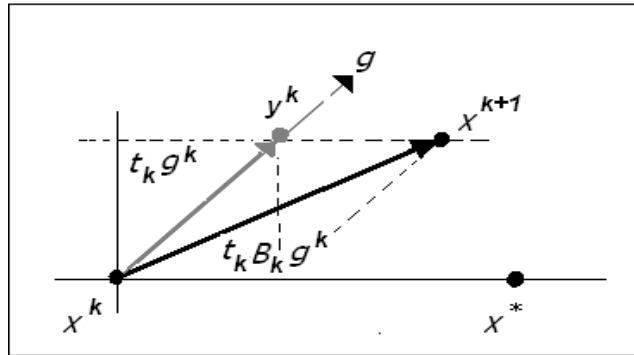


Figura 5.4

Tal como acontece com a direção de máxima descida para o caso diferenciável, a direção oposta ao subgradiente para o caso não diferenciável, possivelmente resulta em um fenômeno de zig-zag que pode manifestar-se em qualquer fase do algoritmo subgradiente, podendo até divergir do ponto ótimo. Veja a figura 5.5 onde $c_1 > c_2 > c_3$.

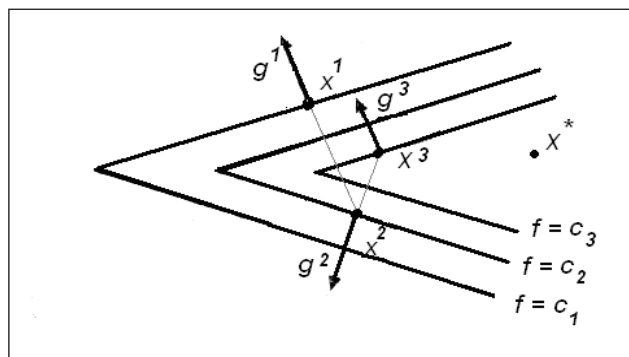


Figura 5.5

Observe que o ângulo entre g^k e g^{k+1} é muito aberto embora os pontos estejam próximos, e que as direções formam zig-zag.

Para contornar estes problemas, pode-se dilatar o espaço na direção g^k , veja a figura 5.6.

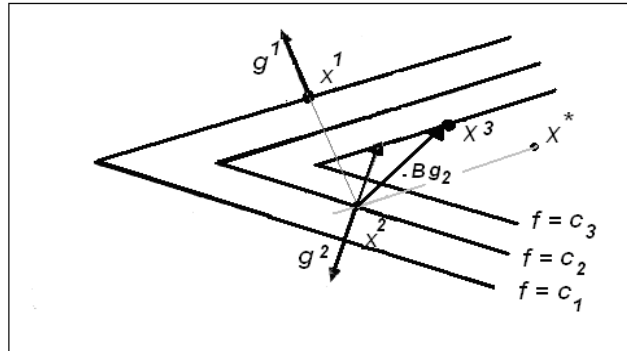


Figura 5.6

5.2.1.1. Métodos de Dilatação do Espaço na Direção do Subgradiente

Suponha que seja feita uma mudança de variável $x = B_k y$ após k iterações dos Métodos Subgradientes, de forma que B_k seja uma matriz não-singular com dimensão apropriada. Em seguida, aplica-se o Método Subgradiente à função $\varphi(y) = f(B_k y)$. Temos que $\tilde{g}^k = g_{\varphi_k}(y_k) = B_k g^k$ é um subgradiente de $\varphi(y)$. Com efeito, como g^k é subgradiente de f em x^k então

$$f(x^k) - f(x) \leq -g^k \cdot (x - x^k) \text{ para todo } x.$$

Como $x = B_k y$ tem-se

$$f(B_k y^k) - f(B_k y) \leq -g^k \cdot (B_k y - B_k y^k) \text{ para todo } y.$$

Como $\varphi(y) = f(B_k y)$ segue que,

$$\varphi(y^k) - \varphi(y) \leq -g^k \cdot B_k (y - y^k)$$

Portanto, $\tilde{g}^k = B_k g^k$ é um subgradiente de $\varphi(y)$.

Seja $t_k > 0$ o tamanho do passo do algoritmo subgradiente aplicado à função

$\varphi(y) = f(B_k y)$, assim

$$y^{k+1} = y^k - t_k \tilde{d}^k$$

como $y = B_k^{-1}x$ temos

$$B_k^{-1}x^{k+1} = B_k^{-1}x^k + t_k d^k$$

multiplicando B_k em ambos os membros obtém-se

$$x^{k+1} = x^k + t_k B_k^t d^k$$

onde $d^k = -\frac{\tilde{g}^k}{\|\tilde{g}^k\|}$.

Nos Métodos de Dilatação do Espaço, as matrizes B_k são atualizadas da seguinte forma:

$$B_{k+1} = B_k R_{\alpha_k}(d_k)^{-1}$$

no qual R é o operador de dilatação do espaço ao longo da direção d com coeficiente de dilatação α , de modo que $\|d\| = 1$ e $\alpha \geq 0$. Portanto,

$$R_\alpha(d) = (\alpha - 1) d^t \cdot d + I, \text{ onde } I \text{ é a matriz identidade.}$$

assim

$$B_{k+1} = B_k R_{1/\alpha_{k+1}}(d_k)$$

Conseqüentemente

$$B_{k+1} = B_k \left(\frac{1}{\alpha} - 1\right) d^t \cdot d + I.$$

5.2.1.1. 1. Algoritmo

Seja x^1 o ponto inicial e B_1 uma matriz não singular (por exemplo a matriz identidade), x^k o ponto na k -ésima iteração, t_k o tamanho do passo nessa iteração, α_k o coeficiente de dilatação do espaço, g^k um subgradiente da função no ponto x^k . Considere a matriz B_k uma matriz quadrada de ordem n .

- **1º passo:** Escolher $x^1 \in \mathbb{R}^n$ e tomar $k=1$;

Na k-ésima iteração

- **2º passo:** Obter: $g^k \in \partial f(x^k)$;

Critério de parada: $0 \in \partial f(x^k)$;

- **3º passo:** Calcular $\tilde{g}^k = B_k g^k$;

4º passo: Calcular $d^k = -\frac{\tilde{g}^k}{\|\tilde{g}^k\|}$;

- **5º passo:** Escolher t_k apropriado;
- **6º passo:** Escolher α_k apropriado;
- **7º passo:** Definir $x^{k+1} = x^k + t_k B_k d^k$;
- **8º passo:** Calcular $B_{k+1} = B_k R_\alpha(d^k)^{-1} = B_k \left(\frac{1}{\alpha_k} - 1 \right) d^{k^t} \cdot d^k + I$, no qual I é a matriz identidade;
- **9º passo:** Atualizar: $k=k+1$.

Se considerarmos $B_1 = I$ e $\alpha_k = 1$ obteremos exatamente o método subgradiente. De fato,

$$g_{\varphi_k}(y^k) = \tilde{g}^k = B_k g^k = I g^k = g^k$$

logo,

$$d^k = -\frac{I_k g^k}{\|I_k g^k\|} = -\frac{g^k}{\|g^k\|}, \quad \text{e}$$

$$B^{k+1} = B^k \left(\frac{1}{\alpha_k} - 1 \right) d^{k^t} \cdot d^k + I = I.$$

Segundo Shor [64], os Métodos de Dilatação no Espaço ao Longo do Subgradiente são mais flexíveis do que os Métodos Subgradientes pois não dependem somente da seqüência de tamanho dos passos $\{t_k\}$, mas também da seqüência de coeficientes de dilatação $\{\alpha_k\}$.

Para uma melhor compreensão sobre os Métodos de Dilatação do Espaço, veja [59], [63] e [64].

Podemos ter várias versões eficientes desse método, basta escolhermos várias regras para determinar as seqüências $\{t_k\}$ e $\{\alpha_k\}$. Uma versão bastante conhecida é o Método de Elipsóides que veremos mais adiante.

5.2.1.1.2. Convergência dos Métodos de Dilatação do Espaço na Direção do Subgradiente

Shor [64] mostra que para certos algoritmos de dilatação do espaço ao longo do subgradiente, o valor da função melhora com uma velocidade de uma progressão geométrica, no qual a razão depende das propriedades da função objetivo que são invariantes com relação à transformação do espaço.

5.2.1.2. Método de Elipsóides

Vamos considerar novamente o problema (P), ou seja,

$$(P) \begin{cases} \text{Minimizar } f(x) \\ x \in \mathbb{R}^n \end{cases}$$

Com $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convexa e não diferenciável. Seja x^* um mínimo de f e suponha que existe R tal que $x^* \in \{x; \|x - x^*\| \leq R\}$ para algum $x^1 \in \mathbb{R}^n$.

A seqüência de tamanho dos passos $\{t_k\}$ é encontrada através da fórmula

$$t_{k+1} = \delta t_k, \text{ no qual } \delta = \frac{n}{\sqrt{n^2 - 1}}. \text{ Tem-se que } t_1 = -\frac{1}{n+1}.$$

A seqüência de coeficientes de dilatação $\{\alpha_k\}$ é definida por:

$$\alpha_k = \sqrt{\frac{n+1}{n-1}}, \forall k \in \mathbb{N}.$$

Podemos descrever o algoritmo geometricamente da seguinte forma:

Considere a bola de centro x^1 e raio R , isto é, $H_1 = S_1 = \{x; \|x - x^1\| \leq R\}$. Encontramos um subgradiente no ponto x^1 , se $g(x^1) = 0$ então x^1 é um mínimo de f . Caso contrário, forma-se um hiperplano $g^1 \cdot (x - x^1) = 0$, assim, $x^* \in \bar{S} \equiv S_1 \cap \{x; g^1 \cdot (x - x^1) \leq 0\}$. Determinamos o volume mínimo do elipsóide de \bar{S} . O centro do novo elipsóide é $x^2 = x^1 - t_2 d^1$ no qual, $d^1 = \frac{g^1}{\|g^1\|}$. Dilatamos o espaço na direção de d^1 . No novo espaço a imagem de H_2 é $\{y; \|y - y^2\| \leq R\delta\}$, no qual, $y^2 = R_\alpha(g^1)x^2$, isto é, temos uma esfera de raio $R\delta$ e centro y^2 . Em seguida, aplicamos o algoritmo no novo espaço, e assim sucessivamente, até o critério de parada ser satisfeito.

5.2.1.2.1. Algoritmo

Seja x^1 o ponto inicial e B_1 uma matriz não singular (por exemplo a matriz identidade), x^k o ponto na k -ésima iteração, t_k o tamanho do passo nessa iteração, α_k o coeficiente de dilatação do espaço, g^k um subgradiente da função no ponto x^k . Considere a matriz B_k uma matriz quadrada de ordem n e $t_1 = -\frac{1}{n+1}$.

- **1º passo:** Escolher $x^1 \in \mathbb{R}^n$ e tomar $k=1$;

Na k -ésima iteração

- **2º passo:** Obter: $g^k \in \partial f(x^k)$;
Critério de parada: $0 \in \partial f(x^k)$;
- **3º passo:** Calcular $\tilde{g}^k = B_k g^k$;
- **4º passo:** Calcular $d^k = -\frac{\tilde{g}^k}{\|\tilde{g}^k\|}$;
- **5º passo:** Definir $x^k = x^{k-1} + t_k B_k d^k$;
- **6º passo:** Calcular $B_{k+1} = B_k R_\alpha (d^k)^{-1}$, sendo $\alpha = \sqrt{\frac{n+1}{n-1}}$;
- **7º passo:** Calcular $t_{k+1} = \delta \cdot t_k$, sendo $\delta = \frac{n}{\sqrt{n^2-1}}$ e atualizar $k=k+1$.

Os trabalhos [1] e [64] abordam os Métodos de Elipsóides com maiores detalhes.

5.2.1.3. Métodos Subgradientes com Dilatação do Espaço na Direção da Diferença de Dois Subgradientes Sucessivos.

Considere a figura 5.7.

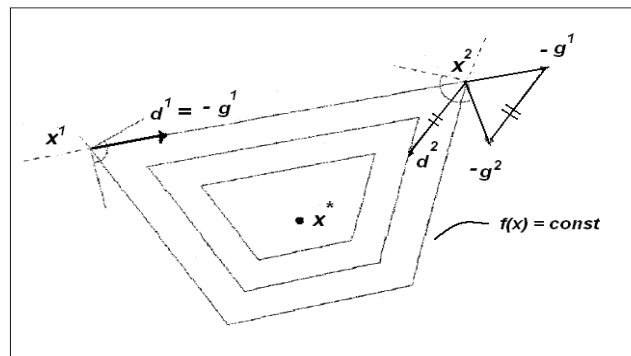


Figura 5.7

Observe que a direção usada em x^1 é $d^1 = -g^1$, e que se utilizarmos em x^2 a direção $d^2 = -g^2$ não obteremos uma direção de descida. Assim, Shor [64] sugere a dilatação do espaço ao longo da diferença entre os dois subgradientes sucessivos, ou seja, $d^2 = -g^2 - (-g^1)$.

Estes métodos, também denominados r-algoritmos, são baseados na aplicação de operações de dilatação do espaço na direção da diferença de dois subgradientes sucessivos, os quais, segundo Sandra [59], são considerados um dos métodos mais eficazes para resolver problemas de otimização não diferenciável.

5.2.1.3.1. r-algoritmos

Suponha que após a k-ésima iteração dos Métodos Subgradientes seja feita uma mudança de variável $x = B_k y$, de forma que B_k seja uma matriz não-singular com dimensão apropriada. Em seguida, aplica-se o método subgradiente à função $\varphi(y) = f(B_k y)$.

Neste procedimento, em cada iteração k, tem-se um subgradiente g^k , o subgradiente anterior g^{k-1} e o operador transformação B_k . Considere os pontos $y^k = B_k^{-1} x^k$ e $\hat{y}^k = B_k^{-1} x^{k-1}$, e seus respectivos subgradientes $\tilde{g}^k = B_k g^k$ e $\hat{g}^k = B_k g^{k-1}$.

A direção r^k é obtida pela diferença dos dois subgradientes da função $\varphi^k(y)$ calculada nos pontos y^k e \hat{y}^k , isto é, $r^k = \tilde{g}^k - \hat{g}^k$, logo, $r^k = B_k g^k - B_k g^{k-1}$. Assim,

$$d^k = -\frac{r^k}{\|r^k\|} = -\frac{B_k^t (g^k - g^{k-1})}{\|B_k^t (g^k - g^{k-1})\|}.$$

A atualização de B_k é dada por $B_k = B_k R_{\alpha_k}(d_k)$.

Aplica-se o tamanho do passo $t_k > 0$ à $\varphi(y)$ obtendo $y^{k+1} = y^k - t_k d^k$. Analogamente

ao Método de Dilatação na Direção do Subgradiente, encontramos $x^{k+1} = x^k + t_k B_k \frac{B_k g^k}{\|B_k g^k\|}$;

Se o critério de parada não for satisfeito, são armazenados x^{k+1} , g^k e B_{k+1} para a próxima iteração.

5.2.1.3.2. Algoritmo

Seja x^1 o ponto inicial e $B_1 = I$ (I é a matriz identidade), x^k o ponto na k -ésima iteração, t_k o tamanho do passo nessa iteração, α_k o coeficiente da dilatação do espaço, g^k um subgradiente da função no ponto x^k .

1ª iteração:

- **1º passo:** Escolher $x^1 \in \mathbb{R}^n$ e tomar $k=1$.
- **2º passo:** Obter: $g^1 \in \partial f(x^1)$ e $\tilde{g}^1 = B_1 g^1$,

Critério de parada: $0 \in \partial f(x^k)$;

- **3º passo:** Escolher $t_1 > 0$ apropriado;

- **4º passo:** Calcular $d^1 = -\frac{\tilde{g}^1}{\|\tilde{g}^1\|}$

- **5º passo:** Calcular $x^2 = x^1 + t_1 d^1$.

Na k -ésima iteração:

- **1º passo:** Obter: $g^k \in \partial f(x^k)$;
Critério de parada: $0 \in \partial f(x^k)$;
- **2º passo:** Calcular $r^k = B_k [g^k - g^{k-1}]$;
- **3º passo:** Calcular $d^k = -\frac{r^k}{\|r^k\|}$;
- **4º passo:** Obter α_k de forma que $0 < \frac{1}{\alpha_k} < 1$;
- **5º passo:** Atualizar $B_{k+1} = B_k \left(\frac{1}{\alpha_k} - 1 \right) d^k + I$;
- **6º passo:** Escolher t_k apropriado;
- **7º passo:** Definir $x^{k+1} = x^k + t_k B_k \frac{B_k g^k}{\|B_k g^k\|}$;
- **8º passo:** Atualizar $k=k+1$

A vantagem deste algoritmo é que a seqüência de valores $f(x^k)$ gerados é monotonicamente decrescente. Segundo [64], a convergência para o ótimo pode ser provada, mas requer algumas hipóteses particulares.

5.2.1.3.3. Exemplo:

Considere o problema de minimizar uma função cujas curvas de níveis estão representadas na figura 5.8, no qual, $c_1 > c_2$.

Sejam $B_1 = I_2$, $0 < \frac{1}{\alpha_k} < 1$, $x^1 \in \mathbb{R}^2$ e $g^1 \in \partial f(x^1) = \tilde{g}^1$.

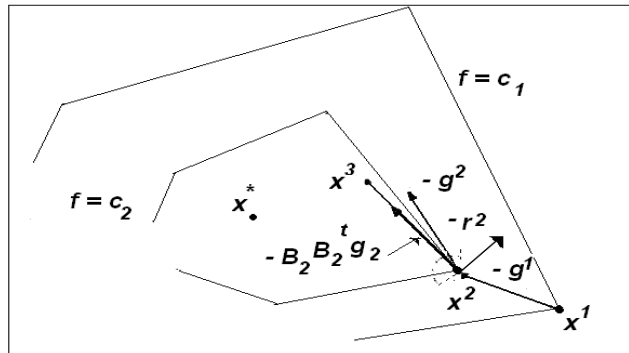


Figura 5.8

Suponha que na segunda iteração encontramos um subgradiente $g^2 \in \partial f(x^2) = \tilde{g}^2$.

Note que g^2 não é uma direção de descida.

Calculamos $r^2 = l.(g^2 - g^1)$, a direção d^1 e a nova matriz dilatação dada por:

$$B_2 = B_1 \left(\frac{1}{\alpha_1} - 1 \right) d^{1 \ t} . d^1 + I .$$

Em seguida, determinamos a nova direção $B_2 B_2^t g^2$ que é uma direção de descida, para encontrar o próximo ponto x^3 .

Para mais detalhes sobre r-algoritmos, consulte [38], [63] e [64].

5.2.2. Métodos ε -Subgradientes ou ε -Máxima Descida

Nos Métodos ε -Subgradientes a direção de busca é encontrada com base nas informações fornecidas pelos subgradientes calculados em pontos numa certa vizinhança do ponto atual.

Para encontrar essa nova direção d^k , resolve-se o problema (P), de modo que $f(x^k + d^k) < f(x^k) - \varepsilon$, para $\varepsilon > 0$ dado. Equivalentemente:

$$(P_{\varepsilon_1}) \begin{cases} \text{Minimizar } f'_\varepsilon(x^k, d) \\ \text{sa } \|d\| \leq 1 \end{cases}$$

De acordo com Makela [44], como $f'(x^k, d)$ é positivamente homogênea, a restrição $\|d\| \leq 1$ é necessária.

De fato, suponhamos que $\|d\| > 1$ e seja \bar{d} o mínimo da função f . Considere

$$\bar{f} = f(\bar{d}) \text{ e } d^* = \frac{\bar{d}}{\|d\|} \text{ então } f(d^*) = f\left(\frac{\bar{d}}{\|d\|}\right). \text{ Como } f \text{ é positivamente homogênea, segue}$$

$$f(d^*) = \frac{1}{\|d\|} f(\bar{d}). \text{ Assim } f(d^*) = \frac{1}{\|d\|} \bar{f}.$$

Por outro lado, como $\|d\| > 1$ segue-se $\frac{1}{\|d\|} < 1$, logo, $\bar{f} \frac{1}{\|d\|} < \bar{f}$. Deste modo, tem-se

$f(d^*) < \bar{f}$ que é uma contradição. Assim $\|d\| \leq 1$.

Tem-se pelo teorema (2.3.4)

$$f'_\varepsilon(x; d) = \max \{g \cdot d; g \in \partial_\varepsilon f(x), \forall d \in \mathbb{R}^n\}.$$

Conseqüentemente, o problema (P_{ε_1}) equivale a:

$$\min_{\|d\| \leq 1} \max_{g \in \partial_\varepsilon f(x^k)} g \cdot d.$$

Pelo teorema (2.3.6) podemos reescrever o problema acima da seguinte forma:

$$\max_{g \in \partial_\varepsilon f(x^k)} \min_{\|d\| \leq 1} g \cdot d.$$

A solução de $\min_{\|d\| \leq 1} g \cdot d$, para $g \neq 0$ é $d = -\frac{g}{\|g\|}$.

Com efeito, temos

$$g \cdot d = \|g\| \|d\| \cos \alpha.$$

O mínimo ocorre quando $\|d\| = 1$ e $\cos \alpha = -1$,

logo

$$g \cdot d = -\|g\| \quad (5.7.1)$$

Por outro lado,

$$g \cdot g = \|g\|^2 \Rightarrow -\frac{1}{\|g\|}(g \cdot g) = -\|g\| \Rightarrow -g \cdot \frac{g}{\|g\|} = \|g\|$$

assim,

$$g \cdot \frac{g}{\|g\|} = -\|g\| \quad (5.7.2)$$

de (5.7.1) e (5.7.2) obtemos:

$$d = -\frac{g}{\|g\|}.$$

Portanto,

$$\max_{g \in \partial_\varepsilon f(x^k)} \min_{\|d\| \leq 1} g \cdot d = \max_{g \in \partial_\varepsilon f(x^k)} g \cdot \left(-\frac{g}{\|g\|} \right) = \max_{g \in \partial_\varepsilon f(x^k)} -\|g\| = \min_{g \in \partial_\varepsilon f(x)} \|g\|.$$

Conseqüentemente, podemos reescrever o problema (P_{ε_1}) da seguinte forma:

$$(P_{\varepsilon_2}) \begin{cases} \text{Minimizar} & \|g\| \\ \text{sa} & g \in \partial_\varepsilon f(x^k) \end{cases}$$

Mas, para resolver (P_{ε_2}), é necessário encontrar todo o ε -subdiferencial $\partial f_\varepsilon(x^k)$, que é extremamente trabalhoso ou mesmo impossível.

Para contornar esse problema, substitui-se o ε -subdiferencial $\partial f_\varepsilon(x)$ por um conjunto que é uma aproximação do $\partial f_\varepsilon(x)$. Com o intuito de encontrar esse conjunto, definiremos o erro de linearização para cada $g^k \in \partial f(x^k)$

$$\alpha(x, x^k) = f(x) - [f(x^k) + g \cdot (x - x^k)] \text{ para todo } x \in \mathbb{R}^n$$

A função α , denominada erro de linearização, mede o erro cometido ao aproximar a função f pelo hiperplano que passa por $(x^k, f(x^k))$ e tem como vetor normal $g^k \in \partial f(x^k)$.

Observe a figura 5.9.

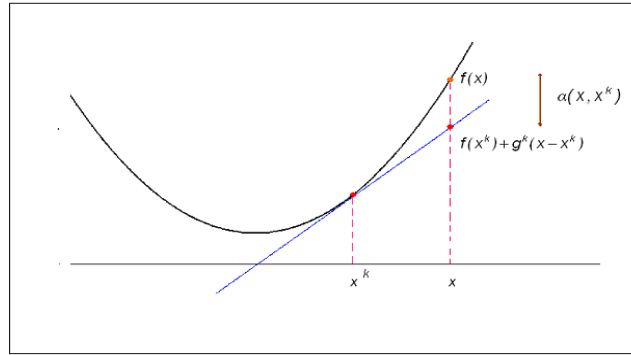


Figura 5.9

Como f é convexa tem-se $\alpha(x, x^k) \geq 0$, para todo $x, x^k \in \mathbb{R}^n$.

Considere os pontos os pontos $x^j \in \mathbb{R}^n$ calculados em iterações anteriores, os subgradientes $g^j \in \partial f(x^j)$ e os erros $\alpha(x, x^j)$, com $j \in J_k$ no qual $J_k \neq \emptyset$ e $J_k \subset \{1, 2, \dots, k\}$.

Define-se o conjunto G_k , para todo $\varepsilon > 0$, como:

$$G_k = \left\{ g \in \mathbb{R}^n; g = \sum_{j \in J_k} \lambda_j g^j, \sum_{j \in J_k} \lambda_j \alpha(x^k, x^j) \leq \varepsilon_k, \lambda_j \geq 0, \sum_{j \in J_k} \lambda_j = 1 \right\}.$$

Considere o exemplo representado na figura 5.10 abaixo,

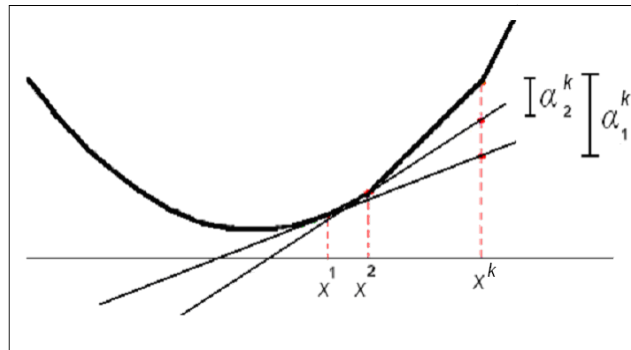


Figura 5.10

Temos que $\alpha_1^k = \alpha(x^k, x^1)$ e $\alpha_2^k = \alpha(x^k, x^2)$. Logo, $J_k = \{1, 2\}$ e $g = \lambda_1 g^1 + \lambda_2 g^2$ de modo que,

$g^1 \in \partial f(x^1)$ e $g^2 \in \partial f(x^2)$, $\lambda_1 \alpha_1^k + \lambda_2 \alpha_2^k \leq \varepsilon_k$, $\lambda_1 + \lambda_2 = 1$ e $\lambda_1, \lambda_2 \geq 0$.

O conjunto G_k é convexo, compacto e tal que $G_k \subset \partial_{\varepsilon_k} f(x^k)$. De fato, por construção, G_k é convexo e compacto. Considere agora $g \in G_k$, então $g = \sum_{j \in J_k} \lambda_j g^j$ no qual,

$g^j \in \partial f(x^j)$. Para $x \in \mathbb{R}^n$ tem-se:

$$\begin{aligned} f(x) &= \sum_{j \in J_k} \lambda_j f(x) \geq \sum_{j \in J_k} \lambda_j [f(x^j) + g^j \cdot (x - x^j)] = \\ &= \sum_{j \in J_k} \lambda_j \{f(x^k) + g^j \cdot (x - x^k) - [f(x^k) - f(x^j) - g^j \cdot (x^k - x^j)]\} = \\ &= \sum_{j \in J_k} \lambda_j [f(x^k) + g^j \cdot (x - x^k) - \alpha(x^k, x^j)] = \\ &= \sum_{j \in J_k} \lambda_j f(x^k) + \sum_{j \in J_k} \lambda_j g^j \cdot (x - x^k) - \sum_{j \in J_k} \lambda_j \alpha(x^k, x^j) \geq \\ &\geq f(x^k) + g \cdot (x - x^k) - \varepsilon_k \end{aligned}$$

logo, $g \in \partial_{\varepsilon_k} f(x)$.

Portanto, pode-se substituir um ε -subgradiente por um elemento de G_k .

Conseqüentemente, pode-se substituir o $\partial_{\varepsilon} f(x)$ por G_k em (P_{ε_2}) . Assim obtemos :

$$(P_{\varepsilon_3}) \begin{cases} \text{Minimizar} & \frac{1}{2} \|g\|^2 \\ \text{sa} & g \in G_k \end{cases}$$

ou equivalentemente,

$$(P_{\varepsilon_k}) \begin{cases} \text{Minimizar} & \frac{1}{2} \left\| \sum_{j \in J_k} \lambda_j g_j \right\|^2 \\ \text{sa} & \sum_{j \in J_k} \lambda_j \alpha(x^k, x^j) \leq \varepsilon_k \\ & \lambda_j \geq 0 \\ & \sum_{j \in J_k} \lambda_j = 1 \end{cases}$$

A partir da solução de (P_{ε_k}) define-se a direção :

$$d^k = - \sum_{j \in J_k} \lambda_j^k g^j.$$

A principal dificuldade encontrada nesse método é a escolha do ε_k adequado. Se ε_k for grande, G_k pode não ser uma boa aproximação de $\partial f_\varepsilon(x)$. Por outro lado, se ε_k for pequeno, o decréscimo esperado no valor da função f será pequeno.

De um modo geral, é muito difícil obter uma regra para a escolha de ε_k .

Em [44], [63] e [64] encontram-se maiores detalhes sobre os Métodos ε -Subgradientes.

5.2.3. Método Subgradiente Conjugado

O Método Subgradiente Conjugado, proposto por Philip Wolfe em [66], é uma extensão do Método Gradiente Conjugado citado no capítulo 3. Foi elaborado com intuito de ser razoavelmente eficaz para as funções convexas.

Como nos Métodos Subgradientes, limita-se a capacidade de obter informações sobre a função objetivo f a ser minimizada, supondo que, dado $x \in \mathbb{R}^n$, tem-se um processo finito, que irá determinar $f(x)$, e um processo finito que irá encontrar um subgradiente, em que não se podem impor quaisquer outras condições.

A idéia principal do algoritmo é gerar uma seqüência de pontos $\{x^k\}$, de direções $\{d^k\}$, de subgradientes $\{g^k\}$ e de escalares $t_k > 0$, tal que, $x^{k+1} = x^k + t_k d^k$ e $f(x^{k+1}) \leq f(x^k)$.

Na iteração k será formado o feixe $G_k = \{g^k, -d^{k-1}\}$, no qual, $d^k = -\text{Nr } G_k$. Podemos observar que $G_k \subseteq \text{conv} \{g^1, \dots, g^k\}$ para todo k .

O passo t_k será determinado por uma minimização aproximada de $f(x^k + t_k d^k)$ para $t \geq 0$. A idéia é que em algum passo os últimos $p+1$ pontos $(x^{k-p}, \dots, x^{k-1}, x^k)$ estejam todos muito próximos, de modo que (g^{k-p}, \dots, g^k) estejam aproximadamente contidos em $\partial f(x^k)$, e também que $d^k = -\text{Nr } G_k$, isto é, $d^k = -\text{Nr} \{g^{k-p}, \dots, g^k\}$ seja suficientemente

pequeno. Então, se a função for diferenciável, tem-se que $\nabla f(x^k) = -Nr \partial f(x^k)$ será muito pequeno e x^k será uma solução aproximada.

5.2.3.1. Algoritmo

Considere os parâmetros positivos $\varepsilon, \delta, b, m_2 < m_1 < 0,5$. Sejam x^1 o ponto inicial, $g^1 \in \partial f(x^1)$, $G^1 = \{g^1\}$ e $a_1 = 0$. Temos que x^k é o ponto na k -ésima iteração, t_k o tamanho do passo nessa iteração, g^k uma aproximação do subgradiente da função no ponto x^k .

- **1º passo:** Escolher $x^1 \in \mathbb{R}^n$ e tomar $k=1$;

Na k -ésima iteração

- **2º passo:** Obter: $d^k = -Nr G$;
se $|d| < \varepsilon$ então vá para o passo 3; senão, vá para o passo 5;
- **3º passo:** Se $a \leq \delta$ então pare; senão, vá para o passo 4.
- **4º passo: (reset):** Considere $x_+ = x$, $g_+ = g$, $G_+ = \{g\}$, e $a_+ = 0$
(o passo k é finalizado);
- **5º passo: (busca linear):** Obter $t > 0$ e $g_+ \in \partial f(x + td)$ tal que:

$$g_+ \cdot d \geq -m_1 |d|^2$$

CASO A: $f(x + td) - f(x) \leq -m_2 t |d|^2$; atualizar $x_+ = x + td$;

OU

CASO B: $t|d| \leq b$; atualizar $x_+ = x$;

- **6º passo:** Escolha $H \subseteq G$ e $G_+ = H \cup \{-d, g_+\}$ e $a_+ = a + |x_+ - x|$
(o passo k é finalizado).

Este algoritmo ou gera uma seqüência de pontos em que f tende para menos infinito ($f \rightarrow -\infty$), ou termina em um ponto x^k que é a solução aproximada do problema.

A busca linear pode ser feita de varias maneiras. Para maiores detalhes, veja [66], o qual, também apresenta uma versão do algoritmo para funções quadráticas e mostra que se a função objetivo for diferenciável e fortemente convexa, normalmente possui uma taxa de convergência linear.

5.2.3.2. Exemplo:

Considere novamente o problema:

$$(P) \begin{cases} \text{Minimizar } f(x,y) \\ (x,y) \in \mathbb{R}^2 \end{cases} \quad (5.1)$$

onde $f(x,y) = \max\{f_1, f_2, f_3\}$ com $f_1 = -x$, $f_2 = x + y$, $f_3 = x - 2y$.

A diferenciabilidade falha nos pontos $(x,0)$, $(-x,-x)$ e $(-x,2x)$ para todo $x > 0$. Considerando que será escolhido $g = \nabla f_i$ no qual i é o menor índice com $f(x,y) = f_i$. Assim, temos que $g = (1,1)$, quando $x > 0$, $y \geq 0$ e $g = (1,-2)$, quando $x > 0$, $y < 0$.

1ª iteração: Seja $x^1 = (6,3)$, logo, $g^1 = (1,1)$. Para $t_1 = 3$, obtemos, $x^2 = (3,0)$.

2ª iteração: Seja $x^2 = (3,0)$, logo $g^2 = (1,1)$. Para um passo suficientemente pequeno, obtemos $x^3 = (2,99999, -0,00001)$, ou seja, $x^3 \cong (3,0)$.

3ª iteração: $x^3 = (2,99999, -0,00001)$, ou seja, $x^3 \cong (3,0)$, porém $g^3 = (1,-2)$.

Assim, $G = \{(1,1), (1,-2)\}$, então, como $d = -\text{Nr } G$, temos que $d = -(1,0) = (-1,0)$. Pela figura 5.11, vemos claramente que $\text{Nr } G = (1,0)$.

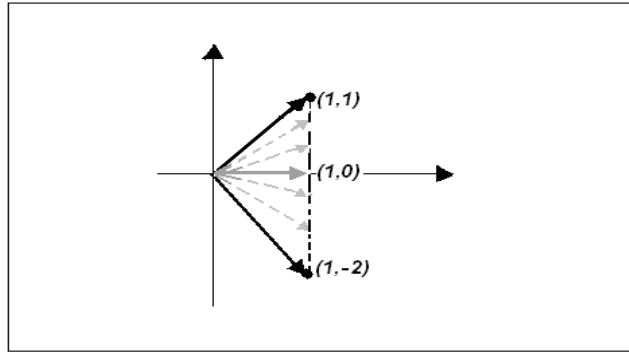


figura 5.11

Temos $|d|=1 > \varepsilon$. Como d é uma direção de descida, ou seja, satisfaz a condição $f(x+td) - f(x) \leq -m_2 t |d|^2$, assim, $x_+ = x^3 + td = (-0,00001, -0,00001)$ e $g_+ = (-1,0)$. Temos que $G_+ = \{(-1,0), (1,0)\}$ e $a_+ \cong 0$, conseqüentemente $d = -\text{Nr } G \Rightarrow d = (0,0)$, assim, $|d|=0 < \varepsilon$ e $a \leq \delta$.

Portanto a solução aproximada é $x^* = (-0,00001, -0,00001)$.

Geometricamente podemos observar o processo pela figura 5.12.

Para maiores detalhes sobre os Métodos Subgradientes Conjugados, veja [66].

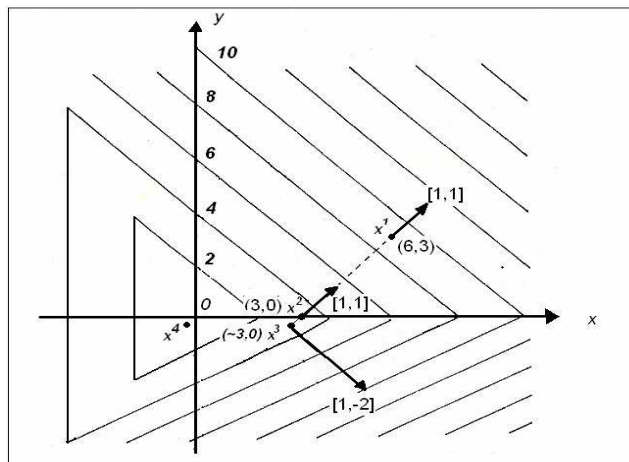


figura 5.12

5.2.4. Métodos Subgradientes Projetados

O Método Subgradiente Projetado é uma extensão dos Métodos Subgradientes, porém, baseados nos Métodos Gradientes Projetados vistos no capítulo 3. Esse método foi elaborado, principalmente, para resolver problemas com restrições, no qual, projeta-se o oposto de um subgradiente, de tal modo que melhora a função objetivo e mantém a viabilidade.

Considere o problema:

$$(PCR) \begin{cases} \text{Minimizar } f(x) \\ \text{sujeito a } g(x) \leq 0 \\ x \in X \end{cases}$$

no qual X é um conjunto convexo simples o bastante para se calcular projeções em X , e $g(x) = (g_1(x), g_2(x), \dots, g_m(x))^t$ são as restrições .

A atualização do ponto é dada pela seguinte fórmula:

$$x^{k+1} = P_X(x^k - t_k g^k)$$

Onde P_X é a projeção sobre X e g^k é um subgradiente de f em x^k .

Para resolver o problema (PCR) utilizando os Métodos de Subgradientes Projetados, pode-se utilizar a relaxação lagrangeana dual, ou seja,

$$F(x, \mu) = f(x) + \mu g(x)$$

cujo $\mu = (\mu_1, \dots, \mu_m)$ e $g(x) = (g_1(x), g_2(x), \dots, g_m(x))^t$.

Para um estudo mais detalhado consulte [9], [13] e [24].

5.2.5. Métodos Subgradientes Afim Escalas

Frenk, Sturm e Zhang em [24] propõem um método denominado Método Subgradiente Afim Escala, que como pode ser observado pelo próprio nome, é uma combinação dos Métodos Subgradientes com os Métodos Afim Escala.

Este método é uma generalização dos Métodos Afim Escala e Dual para problemas Min-Max.

O Método Subgradiente Afim Escala foi elaborado para resolver problemas convexos não diferenciáveis, com restrições lineares. Ou seja,

$$(PRL) \begin{cases} \text{Minimizar } f(x) \\ sa : Ax = b \\ x \geq 0 \end{cases} \quad (5.4)$$

no qual $f(x)$ é uma função convexa, não diferenciável e $A \in \mathbb{R}^{m \times n}$ tem posto m . Considera que existe x^1 tal que $Ax^1 = b$ e $x^1 > 0$.

O Método Subgradiente Afim Escala assume que a região viável $F = \{x; Ax=b \text{ com } x \geq 0\}$ é limitada e que a função objetivo é continuamente lipzchitziana com constante K de Lipzchitz. O método utiliza uma barreira logarítmica com parâmetro $\mu > 0$, isto é,

$$f_\mu(x) = f(x) - \mu \sum_{i=1}^n \log x^i$$

cujo f_μ é estritamente convexa em \mathbb{R}_+^n .

O conjunto de subgradientes é dado por:

$$\partial f_\mu(x) = \partial f(x) - \mu X^{-1}u$$

no qual $u=(1,1,\dots)$, e X é a matriz diagonal com elementos de x .

Devido à barreira logarítmica, só existirá um ponto de mínimo de f_μ sobre F .

A direção de busca é obtida projetando, num espaço escalar, um subgradiente da função objetivo com uma barreira logarítmica. Assim,

$$d = -X \frac{P_{AX} X g_\mu}{\|P_{AX} X g_\mu\|}$$

no qual $g_\mu \in \partial f_\mu(x)$ e $P_{AX} g_\mu$ é a projeção de g_μ sobre o espaço nulo AX , dado por:

$$P_{AX} g_\mu = (I - XA^t (AX^2 A^t)^{-1} AX) g_\mu$$

onde $x > 0$ e I é a matriz identidade.

A escolha do tamanho do passo é análoga à dos Métodos Subgradientes, e é essencial para a velocidade de convergência.

O artigo [24] prova que devido a barreira logarítmica, as iterações no interior de IR_+^n estão contidas num subconjunto compacto de IR_+^n . Prova também que pelo menos uma subsequência das iterações converge para o ponto ótimo.

Supondo que x^k está próximo de x^* e que o ângulo entre a direção do subgradiente d^k e $(x^k - x^*)$ é menor que 90° , tem-se que $\lim_{k \rightarrow \infty} x^k = x^*$, ou seja, o método é globalmente convergente.

A principal vantagem desse método é que todas as iterações permanecem no interior da região viável.

Para maiores esclarecimentos sobre o Método Subgradiente Afim Escala, consulte [24].

5.3. Trabalhos Recentes

Faremos uma breve exposição de alguns trabalhos mais recentes que utilizaram os Métodos Subgradientes e/ou suas extensões.

Zafar, em sua tese de mestrado [2], propõe algoritmos para resolver problemas de otimização não diferenciáveis, denominado Algoritmos de Memória Limitada com Dilatação e Redução do Espaço, que são variantes do r-algoritmo, motivadas pelas atualizações dos Métodos Quase Newton de Pequenas Memórias e Memórias limitadas. Vimos em 5.6 que o r-algoritmo emprega uma estratégia de dilatação do espaço na direção da diferença de dois subgradientes sucessivos. Porém, o método precisa armazenar a matriz de dilatação do espaço e atualiza-la em toda iteração, resultando em um trabalho computacional significativo para problemas maiores. Para contornar esta dificuldade, Zafar desenvolveu um esquema de atualização baseado nos Métodos Quase Newton de Pequenas Memórias, que em vez de utilizar a aproximação da Hessiana da iteração anterior para calcular a aproximação da Hessiana na iteração atual, usa a matriz identidade. Este novo esquema de atualização permite também uma combinação de dilatação do espaço com estratégias de redução.

Os artigos [13] e [24] de 2003 e 2004, respectivamente, definem os Métodos Subgradientes e Subgradientes Projetados, e apresentam algumas regras de tamanhos de passos. Em [13], Boyd e outros, através de gráficos, fazem uma análise da convergência dos Métodos Subgradientes utilizando tamanhos de passos distintos. Estes artigos também apresentam o Problema de Otimização na forma Lagrangeana Dual e propõem a aplicação dos Métodos Subgradientes para resolvê-lo.

Os Métodos Subgradientes são muito utilizados em problemas de otimização combinatória. Vários trabalhos utilizam Relaxações Lagrangeana para modelar esses problemas e aplicam os Métodos subgradientes puros ou com algumas modificações, para resolvê-lo.

A Relaxação Lagrangeana é uma técnica bem conhecida e usada freqüentemente na obtenção de limitantes para problema de Otimização Combinatória. Para instâncias ainda maiores, são aplicados Métodos Subgradientes melhorados por inúmeros recursos. Veremos agora, alguns trabalhos que os utilizam.

Bertsekas e Nédic [49], apresentam uma classe de métodos para minimização de funções convexas não diferenciáveis que consistem da soma de um número grande de funções componentes, ou seja, $f(x) = \sum_{i=1}^m f_i(x)$. Este tipo de minimização aparece em Relaxação Lagrangeana para problemas separáveis de grande escala com restrições acopladas. A idéia é processar iterações utilizando subgradientes das funções componentes com ajustes de variáveis após o cálculo de cada uma dessas funções. O método tem obtido sucesso na solução de problemas de mínimos quadrados de grande porte, tais como os que aparecem em treinamento de redes neurais. Neste trabalho, os autores apresentam propriedades de convergência para variantes determinísticas do método, bem como para variantes estocásticas. São apresentados também, resultados computacionais que mostram que a performance dos métodos são particularmente interessantes.

Em [6], Beltran e Heredia propõem um método heurístico, denominado Método Subgradiente Radar, cujo objetivo é refinar o cálculo do passo dos Métodos Subgradientes. Problemas de otimização de produção de eletricidade são utilizados para comparar o método proposto com o Método Subgradiente clássico.

Ellen Fukuda sugere em [27], algumas maneiras de resolver problemas de programação linear baseando-se na relaxação das restrições “difíceis”, com o cálculo de solução dual através dos Métodos Subgradientes Projetados e o Algoritmo do Volume. Ainda apresenta algumas técnicas de recuperação de soluções primais, tais como, os Métodos Subgradientes com convergência ergódica, que também pode ser encontrado com mais detalhes em [63], e as variantes do Algoritmo de Volume.

Os trabalhos que citaremos a seguir apresentam aplicações recentes dos métodos Subgradientes e/ou suas extensões em problemas práticos.

A evolução da comunicação sem fio abre caminho para contribuição da Pesquisa Operacional no desenvolvimento de modelos matemáticos que minimizem os custos da rede, atendendo à demanda dos usuários e garantindo a qualidade e confiabilidade dos serviços. Durante o projeto de uma rede de telefonia celular surgiu um problema, denominado Problema de Alocação Generalizada com Restrições de Diversidade e Capacidade (PAG-DC), que consiste na atribuição de estações de rádio base a *hubs*, a um custo mínimo, tal que as demandas de cada estação e a capacidade de cada *hubs* sejam atendidas. Adriana Rigolon propõe em [57], um novo modelo matemático para o PAG-DC e emprega a técnica de Relaxação Lagrangeana associada aos Métodos Subgradientes e Dilatação do Espaço para obter limitantes inferiores de boa qualidade. Ela mostra, através de alguns experimentos, que os algoritmos baseados nos Métodos Subgradientes são capazes de encontrar bons limitantes inferiores porém, os Métodos de Dilatação do Espaço são mais eficientes.

Sandra em [59], propõe um algoritmo para a solução de um problema linear inteiro misto de grande porte que otimiza a compra de equipamentos de uma companhia telefônica. Esse problema surgiu na Companhia Telefônica do Rio Janeiro (TELERJ) e trata da otimização da rede de transmissão digital em vigor nessa empresa. Nem o roteamento de sistemas e nem a compra de cabos e equipamentos eram feitas de forma otimizada, por isso, formulou-se um modelo matemático para minimizar os custos. Sandra utiliza a formulação Lagrangeana Dual do problema linear inteiro que modela esse sistema telefônico urbano, e para resolvê-lo, aplica o Método de Dilatação do Espaço na Direção de Dois Subgradientes Sucessivos.

Pimenta em [54], apresentou um trabalho cujo objetivo é otimizar uma rede de transporte de cargas e passageiros, contendo diferentes tipos de produtos e veículos, cujos custos pudessem ter seus valores alterados a cada instância do problema. Também permite o reaproveitamento de veículos, atendendo às restrições de capacidade, no menor custo possível, utilizando-se de técnicas já conhecidas. Primeiro utilizou o algoritmo de Dijkstra

para encontrar o menor caminho. Em seguida, para determinar a rota de custo mínimo, empregou técnicas de Relaxação Lagrangeana associada aos Métodos Subgradientes, além de duas heurísticas complementares.

Glaydston e Luiz Antonio, em [56], propuseram uma nova modelagem matemática a para o Problema da Rotulação Cartográfica de Pontos (PRCP), e apresentaram uma nova técnica para relaxar um problema de otimização combinatória de modo a obter limites de melhor qualidade. Essa nova técnica foi aplicada ao problema PRCP. Assim, como esse problema pode ser representado através de um grafo, essa relaxação consiste em dividir esse grafo em pequenos subproblemas (*clusters*) facilmente resolvidos. As arestas que conectam os clusters são então relaxadas utilizando a Relaxação Lagrangeana e, empregando um Algoritmo de Subgradiente, resolve-se os problemas separadamente reagrupando-os após cada iteração.

Alguns pesquisadores, [41], [47], [48], e [51], propuseram a combinação das Relaxações Lagrangeana e Surrogate com objetivo de conseguir melhores tempos computacionais em relação à Relaxação Lagrangeana.

A Relaxação Lagrangeana/*Surrogate*, denominada *Lagsur*, generaliza a Relaxação Lagrangeana tradicional, no qual, inicialmente, é derivada uma Relaxação Surrogate do conjunto de restrições, e assim, utiliza-se a informação local dessas restrições *surrogate* relaxadas na Relaxação Lagrangeana para acelerar os Métodos Subgradientes. Uma busca local é realizada a cada iteração do algoritmo de modo a corrigir os tamanhos de passos.

Marcelo e Luiz, em [48], utilizam a Relaxação Lagrangeana/*Surrogate* na aplicação de heurísticas baseadas em relaxações que utilizam os Métodos Subgradientes. Segundo os autores, a Relaxação *Lagsur* foi proposta como uma alternativa à relaxação *lagrangeana*, substituindo-a com vantagens em diversos problemas de otimização combinatória.

A Relaxação *Lagsur* é apropriada para problemas com um grande número de variáveis. Quanto maior for o número de variáveis do problema, melhor será o desempenho

da Relaxação *Lagsur* em relação à Relaxação Lagrangeana no que se refere ao tempo de execução.

Em [47], Narciso e Lorena examinam o comportamento da Relaxação *Lagsur* com um Método Subgradiente mais elaborado, aplicados ao problema do Caixeiro Viajante. Nesse Método Subgradiente, para obter uma nova direção de busca, combina-se os subgradientes de duas iterações consecutivas. Os resultados mostram que o ganho da Relaxação Lagrangeana/Surrogate é reduzido e esta passa a ter quase o mesmo comportamento, em relação ao tempo, que a Relaxação Lagrangeana usual. Neste caso, observou-se que não há diferença em utilizar as Relaxações Lagrangeana ou *lagsur*, pois os tempos e limites foram praticamente iguais. Portanto, deve-se substituir Lagrangeano usual pelo Lagrangeano/surrogate no caso de falta de informações, onde os passos dos Métodos Subgradientes são corrigidos melhor e com maior frequência. O artigo mostra também, que o Lagrangeano/surrogate pode ser usado com outros Métodos Subgradientes.

Lílian e Reinaldo, no artigo [51], modelam o Problema de Carregamento de Paletes (PCP) do produtor como um *Problema Inteiro 0 - 1*. O problema consiste em encontrar um arranjo de caixas em camadas horizontais (padrão de carregamento bidimensional) de tal forma que a utilização da superfície do palete seja a máxima possível. Para resolver (PCP) do produtor, os autores, apresentam métodos exatos baseados na aplicação de relaxações Lagrangeana e/ou *surrogate* e, para otimizar os limitantes superiores, utilizam os Métodos Subgradientes. Os métodos propostos são procedimentos de busca em árvore do tipo *branch and bound* que, em cada nó, utilizam limitantes derivados de Relaxações Lagrangeana e/ou *Surrogate*. Testes de redução do problema e heurísticas Lagrangeana e *surrogate* são aplicados aos Métodos Subgradientes para otimizar os limitantes inferiores (soluções factíveis). Um procedimento enumerativo de busca em árvore é usado para a obtenção da solução ótima do problema.

O artigo [41] apresenta os Métodos de Geração de Colunas e o de Decomposição de Dantzig-Wolfe que ficaram conhecidos como métodos eficientes para o tratamento de problemas de programação linear com grande número de variáveis. Um problema mestre

restrito é identificado e novas colunas são geradas através de um subproblema. Estes métodos sofrem de problemas de estabilização. Para amenizar estes problemas, as variáveis duais têm sido controladas de várias maneiras, em geral, restringindo-se sua norma para evitar grandes variações. A Relaxação Lagrangeana/*Surrogate* foi proposta nesse artigo para estabilização de Métodos Subgradientes. Considera-se a combinação do Método de Geração de Colunas e a relaxação Lagrangeana/*surrogate* como uma proposta de estabilização.

Capítulo 6

Alguns Exemplos Numéricos

O Algoritmo Subgradiente foi implementado em Matlab. O tamanho do passo utilizado é $t^k = \frac{1}{k}$. Aplicamos os Métodos Subgradientes em problemas extraídos de [44].

6.1. Problemas Teste

1. CB2

Dimensão 2

Função objetivo $f(x) = \max\{x_1^2 + x_2^2, (2 - x_1)^2 + (2 - x_2)^2, 2e^{-x_1 + x_2}\}$

Ponto ótimo $x^* = (1.139286, 0.899365)$

Valor ótimo $f(x^*) = 1.952225$

2. DEM

Dimensão 2

Função objetivo $f(x) = \max\{5x_1 + x_2, -5x_1 + x_2, x_1^2 + x_2^2 + 4x_2\}$

Ponto ótimo $x^* = (0, -3)$

Valor ótimo $f(x^*) = -3$

3. QL

Dimensão 2

Função objetivo $f(x) = \max\{f_1(x), f_2(x), f_3(x)\}$

No qual, $f_1(x) = x_1^2 + x_2^2$;

$$f_2(x) = x_1^2 + x_2^2 + 10(-4x_1 - x_2 + 4)$$

$$f_3(x) = x_1^2 + x_2^2 + 10(-x_1 - 2x_2 + 6)$$

Ponto ótimo $x^* = (1.2, 2.4)$

Valor ótimo $f(x^*) = 7.2$

4. MIFFLIN1

Dimensão 2

Função objetivo $f(x) = -x_1 + 20 \max\{x_1^2 + x_2^2 - 1, 0\}$

Ponto ótimo $x^* = (1, 0)$

Valor ótimo $f(x^*) = -1$

5. ROSEN

Dimensão 4

Função objetivo

$$f(x) = \max\{f_1(x), f_1(x) + 10f_2(x), f_1(x) + 10f_3(x), f_1(x) + 10f_4(x)\}$$

No qual,

$$f_1(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4;$$

$$f_2(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8$$

$$f_3(x) = x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10$$

$$f_4(x) = x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5$$

Ponto ótimo $x^* = (0, 1, 2, -1)$

Valor ótimo $f(x^*) = -44$

6.2. Resultados Obtidos

Problema 1: CB2

Ponto inicial : $x = (1.0000, -0.1000)$ e $f(x) = 0.8000$

iteração	x1	x2	f
1	1.420331	0.782696	1.01017
2	1.022527	0.514106	1.15021
3	1.198884	0.782194	1.24042
4	1.328847	0.979755	1.32153
5	1.216074	0.820125	1.36396
6	1.303048	0.951029	1.41944
7	1.219018	0.840090	1.45168
8	1.133056	0.756471	1.48694
9	1.193402	0.843031	1.52174
10	1.247305	0.920348	1.55516
11	1.192216	0.851486	1.57724
12	1.237379	0.915699	1.60519
13	1.190595	0.857638	1.62410
14	1.229484	0.912525	1.64812
15	1.188861	0.862312	1.66464
16	1.148007	0.818244	1.68182
17	1.180545	0.863376	1.70092
18	1.211145	0.905820	1.71959
19	1.178976	0.866337	1.73287
20	1.206594	0.904473	1.74969
21	1.177484	0.868770	1.76175
22	1.202655	0.903389	1.77705
23	1.176077	0.870802	1.78810
24	1.149412	0.840859	1.79944
25	1.171764	0.871319	1.81258
26	1.193192	0.900521	1.82552
27	1.170543	0.872796	1.83501
28	1.190476	0.899885	1.84702
29	1.169393	0.874075	1.85588
30	1.188028	0.899334	1.86710
31	1.168310	0.875190	1.87539
32	1.185805	0.898851	1.88591
33	1.167288	0.876170	1.89372
34	1.183776	0.898423	1.90362
35	1.166322	0.877038	1.91099
36	1.181913	0.898039	1.92034
37	1.165407	0.877810	1.92733
38	1.180195	0.897694	1.93619
39	1.164539	0.878501	1.94282
40	1.178602	0.897379	1.95123
41	1.163714	0.879122	1.95755
42	1.177121	0.897091	1.96557
43	1.162928	0.879682	1.97159

44	1.175738	0.896826	1.97925
45	1.162179	0.880190	1.98501
46	1.174442	0.896580	1.99233

=====
 A solução encontrada é: $x = (1.1744 , 0.8966)$ e $f(x) = 1.9923$

A solução ótima é: $x^* = (1.1393, 0.8994)$ e $f(x^*) = 1.9522$

Problema 2: DEM

Ponto inicial : $x = (1, -2)$ e $f(x) = -4$

iteração	x1	x2	f
67	0.000012	-2.921711	-3.07829
68	-0.014139	-2.924541	-3.07546
69	-0.000193	-2.927331	-3.07267
70	0.013553	-2.930080	-3.06992
71	0.000001	-2.932790	-3.06721
72	-0.013364	-2.935463	-3.06454
73	-0.000183	-2.938100	-3.06190
74	0.012821	-2.940700	-3.05930
75	-0.000009	-2.943266	-3.05673
76	0.012652	-2.945799	-3.05420
77	0.000155	-2.948298	-3.05170
78	-0.012181	-2.950765	-3.04923
79	-0.000001	-2.953201	-3.04680
80	0.012027	-2.955607	-3.04439
81	0.000148	-2.957983	-3.04202
82	-0.011587	-2.960330	-3.03967
83	0.000006	-2.962648	-3.03735
84	-0.011449	-2.964940	-3.03506
85	-0.000129	-2.967204	-3.03280
86	0.011060	-2.969441	-3.03056
87	0.000000	-2.971654	-3.02835
88	-0.010935	-2.973840	-3.02616
89	-0.000123	-2.976003	-3.02400
90	0.010569	-2.978141	-3.02186
91	-0.000005	-2.980256	-3.01974
92	0.010454	-2.982348	-3.01765
93	0.000107	-2.984417	-3.01558
94	-0.010130	-2.986465	-3.01354
95	-0.000001	-2.988490	-3.01151
96	0.010023	-2.990495	-3.00950
97	0.000103	-2.992479	-3.00752
98	-0.009716	-2.994443	-3.00556
99	0.000004	-2.996387	-3.00361
100	-0.009619	-2.998311	-3.00169
101	-0.000092	-3.000217	-2.99978
102	-0.000091	-2.991447	-2.99540
103	0.009251	-2.993316	-2.99354

=====

A solução encontrada é $x = (0.0093, -2.9933)$ e $f(x) = -2.9935$

A solução ótima é: $x^* = (0, -3)$ e $f(x^*) = -3$

Problema3: QL

Ponto inicial : $x = (1, 3)$ e $f(x) = 6$

iteração	x1	x2	f
8961	1.164100	2.417918	7.16709
8962	1.164151	2.418018	7.16710
8963	1.164103	2.417919	7.16712
8964	1.164153	2.418018	7.16712
8965	1.164106	2.417919	7.16714
8966	1.164156	2.418019	7.16715
8967	1.164108	2.417920	7.16717
8968	1.164159	2.418019	7.16718
8969	1.164111	2.417921	7.16720
8970	1.164161	2.418020	7.16720
8971	1.164114	2.417921	7.16722
8972	1.164164	2.418020	7.16723
8973	1.164117	2.417922	7.16725
8974	1.164167	2.418021	7.16726
8975	1.164119	2.417922	7.16728
8976	1.164169	2.418022	7.16728
8977	1.164122	2.417923	7.16731
8978	1.164172	2.418022	7.16731
8979	1.164125	2.417923	7.16733
8980	1.164175	2.418023	7.16734
8981	1.164127	2.417924	7.16736
8982	1.164178	2.418023	7.16737
8983	1.164130	2.417925	7.16739
8984	1.164180	2.418024	7.16739
8985	1.164133	2.417925	7.16741
8986	1.164183	2.418024	7.16742
8987	1.164135	2.417926	7.16744
8988	1.164186	2.418025	7.16745
8989	1.164138	2.417926	7.16747
8990	1.164188	2.418025	7.16747
8991	1.164141	2.417927	7.16749
8992	1.164191	2.418026	7.16750
8993	1.164144	2.417927	7.16752
8994	1.164194	2.418026	7.16753
8995	1.164146	2.417928	7.16755
8996	1.164099	2.417830	7.16757
8997	1.164149	2.417929	7.16757
8998	1.164102	2.417830	7.16759
8999	1.164152	2.417929	7.16760

A solução encontrada é: $x = (1.1642, 2.4179)$ e $f(x) = 7.1676$

A solução ótima é: $x^* = (1.2000, 2.4000)$ e $f(x^*) = 7.2000$

Problema4: MIFFLIN1

Ponto inicial $x = (2.1000, -0.7000)$ e $f(x) = -2.0000$

iteração	x1	x2	f
1	1.174629	-0.391543	-1.77967
2	0.734796	-0.244932	-1.59245
3	0.734796	-0.244932	-1.25912
4	0.734796	-0.244932	-1.00912

A solução encontrada é: $x = (0.7348, -0.2449)$ e $f(x) = -1.00912$

A solução ótima é: $x^* = (1, 0)$ e $f(x^*) = -1$

Problema 5: ROSEN

Ponto inicial : $x = (0, 0, 0, 0)$ e $f(x) = -45$

iteração	x1	x2	x3	x4	f
8961	0.007591	0.969251	2.003017	-1.002238	-44.54261
8962	0.007628	0.969273	2.003112	-1.002275	-44.54260
8963	0.007615	0.969258	2.003021	-1.002213	-44.54260
8964	0.007652	0.969280	2.003117	-1.002250	-44.54259
8965	0.007599	0.969258	2.003023	-1.002233	-44.54259
8966	0.007636	0.969281	2.003119	-1.002269	-44.54258
8967	0.007623	0.969265	2.003028	-1.002208	-44.54258
8968	0.007660	0.969288	2.003124	-1.002245	-44.54257
8969	0.007607	0.969266	2.003030	-1.002227	-44.54257
8970	0.007644	0.969288	2.003125	-1.002264	-44.54256
8971	0.007631	0.969273	2.003034	-1.002203	-44.54256
8972	0.007668	0.969295	2.003130	-1.002240	-44.54255
8973	0.007615	0.969273	2.003036	-1.002222	-44.54255
8974	0.007652	0.969296	2.003132	-1.002259	-44.54254
8975	0.007640	0.969281	2.003041	-1.002198	-44.54254
8976	0.007676	0.969303	2.003137	-1.002235	-44.54253
8977	0.007624	0.969281	2.003043	-1.002217	-44.54253
8978	0.007660	0.969304	2.003138	-1.002254	-44.54252
8979	0.007608	0.969282	2.003044	-1.002237	-44.54251
8980	0.007644	0.969304	2.003140	-1.002273	-44.54251
8981	0.007632	0.969289	2.003049	-1.002212	-44.54250
8982	0.007668	0.969311	2.003145	-1.002249	-44.54250
8983	0.007616	0.969289	2.003051	-1.002231	-44.54249
8984	0.007652	0.969312	2.003146	-1.002268	-44.54249

8985	0.007640	0.969296	2.003056	-1.002207	-44.54248
8986	0.007587	0.969274	2.002962	-1.002190	-44.54248
8987	0.007624	0.969297	2.003057	-1.002226	-44.54247
8988	0.007571	0.969275	2.002963	-1.002209	-44.54247
8989	0.007608	0.969298	2.003059	-1.002246	-44.54246
8990	0.007555	0.969276	2.002965	-1.002228	-44.54246
8991	0.007592	0.969298	2.003060	-1.002265	-44.54245
8992	0.007579	0.969283	2.002970	-1.002204	-44.54245
8993	0.007616	0.969305	2.003065	-1.002241	-44.54244
8994	0.007563	0.969283	2.002971	-1.002223	-44.54244
8995	0.007600	0.969306	2.003067	-1.002260	-44.54243
8996	0.007547	0.969284	2.002973	-1.002243	-44.54243
8997	0.007584	0.969306	2.003068	-1.002279	-44.54242
8998	0.007571	0.969291	2.002978	-1.002218	-44.54242
8999	0.007608	0.969313	2.003073	-1.002255	-44.54241

=====

A solução encontrada é:

$$x = (0.007608, 0.969313, 2.003073, -1.002255) \text{ e } f(x) = -44.54241$$

$$\text{A solução ótima é: } x^* = (0, 1, 2, -1) \text{ e } f(x^*) = -44$$

6.3 Um Exemplo dos Métodos de Dilatação do Espaço na Direção do Subgradiente

Considere o problema não diferenciável: Minimizar $f(x, y) = |x| + |y|$.

Os resultados foram obtidos através do algoritmo 5.2.1.1.1 no qual se utiliza a série harmônica como tamanho do passo, ou seja, $t_k = \frac{1}{k}$ e $\frac{1}{\alpha} = 0.9$. Inicia-se com o ponto

$$x^1 = (2, 0).$$

k	x1	x2	f
1	1.0000	0.0000	1.0000
2	0.6818	-0.3182	1.0000
3	0.4673	-0.1037	0.5710
4	0.3066	0.0570	0.3636
5	0.1781	-0.0716	0.2496
6	0.0709	0.0356	0.1065
7	-0.0209	-0.0563	0.0772
8	0.0594	0.0241	0.0835
9	-0.0120	-0.0474	0.0593
10	0.0523	0.0169	0.0692
11	-0.0061	-0.0415	0.0476

12	0.0474	0.0121	0.0595
13	-0.0020	-0.0374	0.0394
14	0.0439	0.0085	0.0524
15	0.0010	-0.0343	0.0354
16	-0.0391	0.0059	0.0450
17	-0.0013	-0.0320	0.0333
18	0.0344	0.0037	0.0381
19	0.0006	-0.0301	0.0306
20	-0.0316	0.0021	0.0336
21	-0.0010	-0.0286	0.0295
22	0.0282	0.0007	0.0289
23	0.0003	-0.0273	0.0276
24	-0.0265	-0.0005	0.0270
25	-0.0008	0.0252	0.0260
26	0.0239	0.0005	0.0244
27	0.0001	-0.0233	0.0235
28	-0.0228	-0.0004	0.0232
29	-0.0007	0.0218	0.0225
30	0.0208	0.0004	0.0212
31	0.0000	-0.0204	0.0204
32	-0.0200	-0.0003	0.0203
33	-0.0006	0.0192	0.0198
34	0.0183	0.0003	0.0186
35	-0.0000	-0.0181	0.0181
36	0.0178	-0.0002	0.0180
37	0.0005	0.0172	0.0176
38	-0.0165	0.0003	0.0167
39	0.0000	-0.0162	0.0163
40	-0.0160	-0.0002	0.0162
41	-0.0004	0.0155	0.0159
42	0.0149	0.0002	0.0151

=====

Solução ótima $x^*=(0 \ 0)$ e $f(x^*)=0$

Solução encontrada = $(0.0149, 0.0002)$ e $f(x^*)= 0.0151$

Conclusões

A Otimização Não Diferenciável produz uma série de dificuldades que, na maioria das vezes, não podem ser superadas pelos algoritmos tradicionais para problemas diferenciáveis. Assim, os problemas não diferenciáveis necessitam de técnicas computacionais próprias.

Os Métodos Subgradientes, desenvolvidos por Shor, foram os pioneiros em otimização não diferenciáveis. Possuem uma estrutura simples, porém, apresentam baixa velocidade de convergência, na maioria dos casos, pior do que linear.

No capítulo 5, apresentamos vários métodos que foram propostos no sentido de melhorar a performance dos Métodos Subgradientes. Dentre esses métodos, destacam-se os Métodos de Dilatação do espaço, que segundo [59] estão sendo reconhecidos como métodos eficazes para resolver problemas de otimização não diferenciável. Adriana Rigolon [57], através de alguns experimentos, mostrou que para o problema de alocação generalizada com restrições de diversidade e capacidade, os Métodos de Dilatação do Espaço foram mais eficientes do que os Métodos Subgradientes.

Alguns trabalhos [27], [54], [57] e [59], vistos na seção 5.3, mostraram que os Métodos Subgradientes puros ou com algumas modificações, tem-se revelados eficientes quando combinados com a Relaxação Lagrangeana e/ou Surrogate para resolver, principalmente, problemas de otimização combinatória.

Assim, de acordo com [6], apesar das desvantagens dos Métodos Subgradientes, eles são métodos populares por serem muito simples de implementar, e pelo seu baixo custo computacional.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] AKGÌUL, M. *Topics in Relaxation and Ellipsoidal Methods*. Boston: Pitman Advanced Pub. Program, 1984.
- [2] ANSARI, Z. A., *Limited Memory Space Dilation and Reduction Algorithms*, Thesis (Master), Faculty of the Virginia, Blacksburg, Virginia, 1998.
- [3] AZAM, R., *Studies on Quasi-Newton Methods for Nonsmooth Convex Optimization*, Thesis Doctor's, Graduate School of Information Science, Nara Institute of Science and Technology, 1998.
- [4] BAHN, O., DU MERLE, O., GOFFIN, J.L., VIAL, J.P., *A Cutting Plane Method from Analytic Centers for Stochastic Programming*, *Mathematical Programming*, vol. 69, pp. 45-73, 1995.
- [5] BAZARAA, M.S., SHERALI, H.D., SHETTY, C.M., *Nonlinear Programming, Theory and Algorithms*, 2 ed, John Wiley and Sons, New York, 1983.
- [6] BELTRAN, C., HEREDIA, F. J., *An Effective Line Search for the Subgradient Method*, University of Geneva, Switzerland, 2004.
- [7] BEN-TAL, A., ZOWE, J., *Necessary and Sufficient Optimality Conditions for a Class of Nonsmooth Minimization Problems*, *Mathematical Programming*, North-Holland, vol. 24, pp. 70-91, 1982.

- [8] BERTSEKAS, D.P., *Nondifferentiable Optimization Via Aproximation*, Mathematical Programming Study 3, North-Holland Publishing Company, 1-25, 1975.
- [9] BERTSEKAS, D.P., *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 1995.
- [10] BOCANEGRA, S., SANTOS, M.A., CAMPOS, F. F.F., *Uma Proposta de Solução para o Problema Não Linear de Fluxo Multiproduto Utilizando Pontos Interiores*, Universidade Federal de Minas Gerais, Belo Horizonte.
- [11] BONNANS, J.F., GILBERT, J.C., LEMARÉCHAL, C., SAGASTIZÁBAL, C., *Optimization Numérique, Aspects Théoriques et Pratiques, Mathématiques & Applications*, 27, Springer-Verlag, Berlin, 1997.
- [12] BORTOLOSSI, H. J., *Cálculo diferencial a Várias Variáveis, Uma Introdução à Teoria de Otimização*, Edições Loyola, Editora PUC-Rio, 2002.
- [13] BOYD, B., XIAO L., MUTAPCIC, A., *Subgradient Methods, Notes for EE392o*, Stanford University, Autumn, 2003.
- [14] BURKE, J. V., *Second Order Necessary and Sufficient Conditions for Convex Composite NDO*, *Mathematical Programming*, North-Holland, vol. 38, pp. 287-302, 1987.
- [15] CAMERINI, P., FRATTA, L., MAFFIOLI, F., *On Improving Relaxation Methods by Modified Gradient Techniques*, In *Math Programming Stud 3* (M.L. Balinski, P. Wolfe, eds.), pp 26-34, North Holland, Amsterdam, 1975.

- [16] CANTANE, D.R., *Métodos de Pontos Interiores Aplicados ao Problema de Regressão Pela Norma L_p* , Dissertação (Mestrado), ICMC/USP, São Carlos, 2004.
- [17] CHENEY, W., GOLDSTEIN, A.A., *Newton's Method for Convex Programming and Chebyshev Approximation*, Numerische Mathematik, 1-5, pp. 253-268, 1959.
- [18] CLARKE, F.H., *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.
- [19] DENNIS, J.E.,JR., SCHNABEL, R.B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Classics In Applied Mathematics, SIAM, 1996.
- [20] ELZINGA, J., MOORE, T.G., *A Central Cutting Plane Algorithm for the Convex Programming Problem*, Mathematical Programming, vol. 8, pp. 134-145, 1975.
- [21] FERREIRA, P.A.V., *Otimização Não-Linear*, Universidade Estadual de Campinas, Campinas, 1994.
- [22] FLETCHER, R., MAZA, E.S., *Nonlinear Programming and Nonsmooth Optimization by Successive Linear Programming*, Mathematical Programming, north-Holland, vol. 43, pp. 235-256, 1998.
- [23] FREIRE, W. P. *Um Algoritmo de Direções Viáveis para Otimização Convexa Não Diferenciável*, Tese (Doutorado), COPPE/UFRJ, Rio de Janeiro 2005

- [24] FRENK, J.B.G., STURM, J.F., ZHANG, S., *An Interior Point Subgradient Method for Linearly Constrained Nondifferentiable Convex Programming*, Erasmus University Rotterdam, Netherlands, 1996.
- [25] FREUND, R. M., *Subgradient Optimization, Generalized Programming, and Nonconvex Duality*, Institute of Technology, Massachusetts, 2004.
- [26] FRIEDLANDER, A., *Elementos de Programação Não-Linear*, Editora da Universidade Estadual de Campinas, Campinas, 1994.
- [27] FUKUDA, E.H., *Algoritmo do Volume e Otimização Não Diferenciável*, Dissertação (Mestrado), Universidade de São Paulo, São Paulo, 2007.
- [28] GOFFIN, J., VIAL, J., *Interior Point Methods for Nondifferentiable Optimization*, HEC Technical Report 97.24, 1998.
- [29] GOFFIN, J.L., VIAL, J.P., *Convex Nondifferentiable Optimization: A Survey Focused on the Analytic Center Cutting Plane Method*, *Optimization Methods and Software*, vol. 17, pp. 805-867, 2002.
- [30] HERSKOVITS, J., *A Two-Stage Feasible Directions Algorithm for Nonlinear Constrained Optimization*, *Mathematical Programming*, vol. 36, pp. 19-38, 1986.
- [31] HERSKOVITS, J., *A View on Nonlinear Optimization*, ed. *Herskovits, J., Advances in Structural Optimization*, Kluwer Academic Publishers, pp. 71-116, 1995.

- [32] HERSKOVITS, J., *A Feasible Directions Interior Point Technique for Nonlinear Optimization*, JOTA, Journal of Optimization Theory and Applications, vol. 99, pp. 121-146, 1998.
- [33] HIRIART-URRUTY, J., LEMARÉCHAL, C., *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, 1993.
- [34] JUSSIANI, L.F., *Desempenho do Método de Lagrangeano Aumentado com Penalidade Quadrática*, Dissertação (Mestrado), Universidade Federal do Paraná, Curitiba, 2004
- [35] KELLEY, J.E., *The Cutting Plane Method for Solving Convex Programs*, J. SIAM, 8, pp. 703-712, 1960.
- [36] KIWIEL, K.C., *Methods of Descent for Nondifferentiable Optimization*, Springer-Verlag, New York, 1985.
- [37] LASDON, L.S., *Optimization Theory for Large Systems*, Collier Macmillan Publishers, London, 1970.
- [38] LEMARÉCHAL, C., *Nondifferentiable Optimization*, in: Handbooks in OR&MS, eds. Nemhauser, G.L., et al., Elsevier Science Publishers, North-Holland, pp. 529-572, 1989.
- [39] LEMARECHAL, C., *Bundle Method in Nonsmooth Optimization*, *Nonsmooth Optimization: Proceedings of IIASA Workshop*, C. Lemarechal and R. Mifflin, eds., pp. 79-109, 1978.
- [40] LIMA, A.L.D.S, *Uma Estratégia de Decomposição por Relaxação Lagrangeana para a Otimização da Programação Diária da Operação de Sistemas Hidrotérmicos com*

Modelagem Detalhada da Rede Elétrica, Aplicação ao Sistema Brasileiro, Tese (Doutorado), COPPE/UFRJ, Rio de Janeiro, 2007.

[41] LORENA, A. N., PEREIRA, M. A., SALOMÃO, S. N. A., *A relaxação Lagrangeana/surrogate e o método de geração de colunas: novos limitantes e novas colunas*, Instituto Nacional de Pesquisas Espaciais - INPE São José dos Campos – SP.

[42] LUENBERGER, D.G., *Optimization by Vector Spaces Methods*, John Wiley, New York, 1969.

[43] LUENBERGER, D.G., *Linear and Nonlinear Programming*, 2 ed., Addison-Wesley Publishing Company, 1984.

[44] MAKELA, M.M., NEITTAANMAKI, P., *Nonsmooth Optimization, Analysis and Algorithms with Applications to Optimal Control*, World Scientific Publishing, 1992.

[45] MARTINEZ, J.M., SANTOS, S.A., *Métodos Computacionais de Otimização*, 20º Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro, 1995.

[46] MIFFLIN, R., *An Algorithm for Constrained Optimization with Semismooth Functions*, Mathematics of Operations Research, 2, pp. 191-207, 1977.

[47] NARCISO, M.G., LORENA, L.A.N., *Uma Aplicação da Relaxação Lagrangeana/Surrogate ao Problema Simétrico do Caixeiro Viajante Usando um Método de Subgradientes Melhorado*, 2006.

[48] NARCISO, M.G., LORENA, L.A.N., *Um método exato para multiplicadores lagrangeano/surrogate*, LAC/INPE- Instituto Nacional de Pesquisas Espaciais.

- [49] NEDIC, A., BERTSEKAS, D. P., *Incremental Subgradient Methods For Nondifferentiable Optimization*, SIAM J. on Optimization, Cambridge, Mass.,2001.
- [50] NOCEDAL, J., WRIGHT, S.J., *Numerical Optimization*, Springer-Verlag, New York, 1999.
- [51] OLIVEIRA, L.K., MORABITO, R., *Métodos Exatos Baseados em Relaxações Lagrangeana e Surrogate para o Problema de Carregamento de Paletes do Produtor*, Pesquisa Operacional, v.26, n.2, p.403-432, 2006.
- [52] PANIER, E.R., TITS, A.L. AND HERSKOVITS, J., *A QP-free Globally Convergent, Locally Superlinearly Convergent Algorithm for Inequality Constrained Optimization*, SIAM Journal of Control and Optimization, vol. 26, pp. 788-810, 1988.
- [53] PERESSINI, A.L., SULLIVAN, F.E., UHL, J.J.,JR., *The Mathematics of Nonlinear Programming*, Springer-Verlag, New York, 1988.
- [54] PIMENTA, D. J., *Algoritmo de Otimização para o Problema de Roteamento de Veículos no Transporte Conjunto de Cargas e de Passageiros*, Dissertação (Mestrado), Universidade Federal de Minas Gerais, Belo Horizonte, 2001.
- [55] POLYAK, B.t., *History of Mathematical Programming in the URRS: Analyzing the Phenomenon*, Springer-Verlag , pp. 401-416, 2001.
- [56] RIBEIRO, G.M., LORENA, L.A.N., *Relaxação Lagrangeana com Formação de Clusters Aplicada a Rotulação Cartográfica de Pontos*, Instituto Nacional de Pesquisas Espaciais, São José dos Campos – SP.

- [57] RIGOLON, A.A., *Métodos de Subgradientes e de Dilatação do Espaço Aplicados ao Projeto de Rede de Telefonia Móvel*, Dissertação (Mestrado), MIA/UNIFOR, 2006.
- [58] ROCKAFELLAR, R.T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
- [59] RODRIGUES, S.I.M., *Relaxação Lagrangeana e Subgradientes com Dilatação de Espaço Aplicados a um Problema de Grande Porte*, Tese, COPPE/UFRJ, Rio de Janeiro, 1994.
- [60] SANT'ANA, D.C., *Uma Nova Proposta Utilizando Métodos de Lagrangeano Aumentado com Penalidades Modernas na Resolução de Problemas de Contato*, Dissertação (Mestrado), Universidade Federal do Paraná, Curitiba, 2005.
- [61] SENNE, E. LE.F., LORENA, L. A. N, PEREIRA, M.A., *Um Algoritmo Branch-And-Price para Problemas de Localização de P-Medianas*, A pesquisa Operacional e os Recursos Renováveis, Natal RN, 2003.
- [62] SANTOS, M.A., *Algoritmos de Centros para Programação Convexa Não Suave*, Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, 1998.
- [63] SHOR, N.Z., *Minimization Methods for Nondifferentiable Functions*, Springer-Verlag, Berlim, 1985.
- [64] SHOR, N.Z., *Nondifferentiable Optimization and Polynomial Problems*, Kluwer Academic Publishers, Boston, 1998.

[65] SORIANO, R.M., VIVAS, C. F., *Una Caracterizacion de Disenos de Experimentos optimos por Subgradientes*, Trabajos de Estadística, Vol. 2, Num. 1, pp.3-13, 1987.

[66] WOLFE, P., *A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions*, Mathematical Programming Study 3, North-Holland Publishing Company, 145-173, 1975.