

Universidade Federal de Juiz de Fora

Pós-Graduação em Matemática

Mestrado em Matemática

Jesús Cernades Gómez

***MÉTODO DE DIREÇÕES INTERIORES AO
EPÍGRAFO PARA A SOLUÇÃO DE PROBLEMAS
DE OTIMIZAÇÃO NÃO-CONVEXOS E
NÃO-DIFERENCIÁVEIS VIA DUALIDADE
LAGRANGEANA***

Juiz de Fora

2013

Jesús Cernades Gómez

***MÉTODO DE DIREÇÕES INTERIORES AO
EPÍGRAFO PARA A SOLUÇÃO DE PROBLEMAS
DE OTIMIZAÇÃO NÃO-CONVEXOS E
NÃO-DIFERENCIÁVEIS VIA DUALIDADE
LAGRANGEANA***

Dissertação aprovada pela Comissão Examinadora abaixo elencada como requisito para a obtenção do título de Mestre em Matemática pelo Mestrado Acadêmico em Matemática do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora.

Orientador: Prof. Dr. Wilhelm Passarella Freire

Juiz de Fora

2013

Cernades Gómez, Jesús.

Método de Direções Interiores ao Epígrafo para a solução de problemas de Otimização não-Convexos e não-Diferenciáveis via Dualidade Lagrangeana / Jesús Cernades Gómez. - 2013.

99f. : il.

Dissertação (Mestrado em Matemática) – Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

1. Matemática 2. Programação não-linear. 3. Otimização não-diferenciável.
4. Dualidade Lagrangeana. I. Título.

Jesús Cernades Gómez

***MÉTODO DE DIREÇÕES INTERIORES AO
EPÍGRAFO PARA A SOLUÇÃO DE PROBLEMAS
DE OTIMIZAÇÃO NÃO-CONVEXOS E
NÃO-DIFERENCIÁVEIS VIA DUALIDADE
LAGRANGEANA***

Dissertação aprovada pela Comissão Examinadora abaixo elencada como requisito para a obtenção do título de Mestre em Matemática pelo Mestrado Acadêmico em Matemática do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora.

Prof. Dr. Wilhelm Passarella Freire
(Orientador)

Mestrado Acadêmico em Matemática
Instituto de Ciências Exatas - UFJF

Prof. Dr. Sandro Rodrigues Mazorche
Mestrado Acadêmico em Matemática
Instituto de Ciências Exatas - UFJF

Prof. Dr. José Herskovits Norman
COPPE-UFRJ

Juiz de Fora, 07 de junho de 2013.

À minha mãe Alejandra, meu pai Wilbert, minha esposa Flor e minha família que esteve sempre apoiando e incentivando diante das dificuldades, visando minha formação pessoal e profissional e dando-me suporte emocional. A vocês o meu eterno agradecimento.

AGRADECIMENTOS

A Deus, por que sem Ele nada disso seria possível, pois Dele são a sabedoria e o poder; é ele quem dá sabedoria aos sábios e entendimento aos inteligentes.

A minha mãe Alejandra, meu pai Wilbert, minha esposa Flor e minha família, por me apoiar o tempo todo, ajudando a alcançar os meus sonhos e objetivos.

Aos meus amigos pela compreensão, conhecimento compartilhado e incentivo, em especial a Juan, Gisely e Carolina.

Ao Prof. Dr. Wilhelm Passarella Freire, orientador deste trabalho pelo conhecimento, atenção e disponibilidade com qual fui acolhido e orientado. Minha sincera gratidão!

Aos professores Dr. Fabio Rodrigues Pereira e Dra. Flaviana Andréa Ribeiro pelo apoio e compreensão. Minha sincera gratidão!. Enfim, a todos que de alguma forma apoiaram a realização deste trabalho.

RESUMO

Este trabalho tem por finalidade apresentar um método para a solução de problemas de otimização não-convexos e não-diferenciáveis. O método, chamado IED (Interior Epigraph Directions), aplica-se a problemas de otimização cuja função objetivo é contínua e definida em um subconjunto compacto de \mathbb{R}^n , sujeita a restrições de igualdade e/ou desigualdade.

O método IED considera o problema dual induzido por uma função lagrangeana aumentada e obtém a solução primal gerando uma sequência de pontos no interior do epígrafo da função dual. Primeiramente, um subgradiente é usado para gerar uma aproximação linear do problema dual. Em seguida, usa-se esta aproximação linear para definir-se uma direção de busca interior ao epígrafo da função dual. Obtém-se então, a partir de um ponto no interior do epígrafo, um novo ponto interior e, consequentemente, uma sequência de pontos interiores é construída. Essa sequência produz uma sequência dual que por sua vez origina uma sequência primal, através da solução de um subproblema originado pela dualidade.

A análise de convergência do algoritmo é também apresentada bem como resultados numéricos da solução de problema extraídos da literatura.

Palavras-Chave: Programação não-linear. Otimização não-diferenciável. Dualidade Lagrangeana.

ABSTRACT

This work presents a method for solving constrained nonsmooth and nonconvex optimization problems. The method, called IED (Interior Epigraph Directions) can be applied to optimization problems with continuous objective functions defined over compact subsets of \mathbb{R}^n and subjected to equalities and/or inequalities constraints.

The IED method considers the dual problem induced by a generalized augmented Lagrangian function and obtains the primal solution by generating a sequence of iterates in the interior of the dual function. First, a subgradient is used to build a linear approximation to the dual problem. Then, this linear approximation is used to define a search direction in the interior of the dual function. From an interior point of the epigraph, a new point is obtained and an interior sequence to the epigraph is built. This sequence of interior points generates a dual sequence which in its turn generates a primal sequence by solving a problem originated by duality.

The convergence analysis is also presented as well as numerical results of several problems obtained from the literature.

Keywords: Nonlinear programming. Non-differentiable optimization. Lagrangian duality.

LISTA DE FIGURAS

1.1	Existencia do hiperplano suporte de um conjunto D	19
1.2	Função convexa	20
2.1	Direção de busca FDIPA	34
3.1	Passo do Método Subgradiente.	39
3.2	Exemplo de uma direção com sentido contrário a um subgradiente que não é uma direção de descida.	42
3.3	Método de Planos Cortantes	43
3.4	Falha do Método de Plano Cortantes	44
6.1	Passo sério do Método IED	74
6.2	Passo nulo e a direção d_{DSG}	74
6.3	Diagrama de fluxo do Método IED	76
7.1	Kissing Number Problem para dimensão 3	94

LISTA DE TABELAS

7.1	Parâmetros e pontos iniciais usados pelo método DSG.	89
7.2	n_{its} =número de iterações, n_L =número de vezes que a função lagrangeana é calculada	90
7.3	n_{its} =número de iterações, n_L =número de vezes que a função lagrangeana é calculada	91
7.4	n_{its} =número de iterações, n_L =número de vezes que a função lagrangeana é calculada	92
7.5	Alguns valores conhecidos e limitantes para k_n	93
7.6	IED frente a AUGPEN, DSG, Ipopt, ALBOX, LANCELOT	95

LISTA DE SÍMBOLOS

A, B, C	matriz
\mathbb{R}^n	espaço Euclidiano n-dimensional
$\nabla f(x)$	gradiente de f em x
$\nabla^2 f(x)$	matriz Hessiana de f em x
$\partial f(x)$	subdiferencial de uma função convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no ponto x
$\partial_\varepsilon f(x)$	ε -subdiferencial da função convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no ponto x
$\mathcal{N}_D(x)$	O cone normal de um conjunto D convexo não vazio
$\{x_k\}, (x_k) = (x_0, x_1, \dots, x^k, \dots)$	uma sequência de pontos do \mathbb{R}^n
$f'(x; d)$	derivada direcional generalizada de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ em x na direção d
$\mathcal{D}_f(x)$	o conjunto de direções de descida da função f a partir do ponto x
$B(\bar{x}, \delta) = \{x \in \mathbb{R}^n \mid \ x - \bar{x}\ \leq \delta\}$	bola fechada de raio δ centrada no ponto $\bar{x} \in \mathbb{R}^n$
$L_{f,X}(c)$	o conjunto de nível da função f a partir do ponto x
x_i	componente i do vetor x
$\text{epi}(f)$	epígrafo da função f
X°	interior do conjunto X
$\mathbb{R}^{n \times n}$	espaço de matrizes em \mathbb{R}
$x^T y$	produto interno usual em \mathbb{R}^n
$\ \cdot\ $	a norma Euclidiana n-dimensional
$\text{cl}X$	o fecho de conjunto X
$\text{fr}X$	a fronteira do conjunto X
\mathbb{R}_+	o conjunto dos números reais não negativos

SUMÁRIO

INTRODUÇÃO	14
1 NOÇÕES PRELIMINARES	16
1.1 PROBLEMA DE OTIMIZAÇÃO E DEFINIÇÕES BÁSICAS	16
1.1.1 Problema Primal	16
1.2 DIREÇÃO DE DESCIDA / DIREÇÃO VIÁVEL	18
1.3 CONJUNTOS CONVEXOS E PROPRIEDADES	19
1.4 FUNÇÕES CONVEXAS	20
1.4.1 Funções convexas diferenciáveis	21
1.4.2 Propriedades básicas	21
1.4.3 Funções convexas não diferenciáveis	22
1.5 ELEMENTOS DA TEORIA DE DUALIDADE	24
2 FDIPA - FEASIBLE DIRECTIONS INTERIOR POINT ALGORITHM - UM MÉ- TODO DE PONTOS INTERIORES PARA OTIMIZAÇÃO NÃO LINEAR DIFE- RENCIÁVEL	31
2.1 O MÉTODO FDIPA	31
2.2 HIPÓTESE PARA A CONVERGÊNCIA DO MÉTODO	32
2.3 ALGORITMO FDIPA	35
2.4 CONVERGÊNCIA DO MÉTODO FDIPA	37
3 MÉTODOS PARA OTIMIZAÇÃO CONVEXA NÃO DIFERENCIÁVEL	38
3.1 MÉTODO SUBGRADIENTES	38

3.1.1 Escolha do Tamanho do Passo	39
3.2 O MÉTODO DE PLANOS CORTANTES	42
3.3 MÉTODO DE FEIXE	44
3.3.1 Algoritmo do método de feixe com programação quadrática	45
4 NFDA - NONSMOOTH FEASIBLE DIRECTIONS ALGORITHM - PARA OTIMIZAÇÃO CONVEXA NÃO DIFERENCIÁVEL	48
4.1 INTRODUÇÃO	48
4.1.1 O método NFDA	49
4.2 ALGORITMO NFDA	50
4.3 CONVERGÊNCIA DO ALGORITMO NFDA	52
5 DSG - THE DEFLECT SUBGRADIENT METHOD - UM ALGORITMO TIPO SUBGRADIENTE PARA OTIMIZAÇÃO NÃO-CONVEXA E NÃO-DIFERENCIÁVEL	54
5.1 O MÉTODO DSG	54
5.2 O ALGORITMO DSG	55
5.2.1 Algoritmo DSG	55
5.3 CONVERGÊNCIA DO MÉTODO DSG	58
5.4 Uma escolha especial de s_k	61
6 MÉTODO DE DIREÇÕES INTERIORES AO EPÍGRAFO PARA A SOLUÇÃO DE PROBLEMAS DE OTIMIZAÇÃO NÃO-CONVEXOS E NÃO-DIFERENCIÁVEIS VIA DUALIDADE LAGRANGEANA	64
6.1 O MÉTODO IED	65
6.1.1 O passo do IED	73
6.2 O ALGORITMO DE DIREÇÕES INTERIORES AO EPÍGRAFO	75
6.2.1 Diagrama de fluxo do Método IED	76
6.3 CONVERGÊNCIA DO MÉTODO IED	77

7	RESULTADOS NUMÉRICOS	83
7.1	PROBLEMAS TESTE	83
7.1.1	Resultados Obtidos	88
7.2	APLICAÇÃO	93
7.2.1	Kissing Number Problem	93
8	CONCLUSÕES	96
	REFERÊNCIA	97

INTRODUÇÃO

Otimização é uma área da Matemática que vem sendo aplicada a um número cada vez maior de problemas reais. A Programação Matemática, em particular, tem sido utilizada para modelar e resolver diversos tipos de problemas de Engenharia, Economia, Administração, Física, Química, etc. Existem muitos problemas em otimização cujo interesse reside em encontrar pontos de máximo ou de mínimo de funções que não possuem derivadas em alguns pontos de seu domínio. É desse tipo de problema que trata a área da Programação Matemática denominada Otimização Não Diferenciável. Em particular, a Otimização Convexa se ocupa de funções convexas não necessariamente diferenciáveis.

Para resolver esta classe de problemas são necessárias técnicas que substituem o cálculo diferencial clássico. Tais técnicas são oriundas de uma área da Matemática chamada Análise Convexa. No caso de funções não convexas, estende-se os conceitos da Análise Convexa a funções localmente lipschitzianas.

Apresentamos neste trabalho um algoritmo proposto por Freire e outros [21] para minimização de funções não-convexas e não diferenciáveis.

O problema em estudo é definido por:

$$(P) \begin{cases} \text{minimizar } f_0(x) \\ f(x) = 0 \\ x \in X, \end{cases}$$

onde $X \subset \mathbb{R}^n$ é compacto, $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são contínuas.

Embora (P) seja estabelecido como um problema com restrições de igualdade apenas, restrições de desigualdade também podem ser incorporados através do uso do operador $a^+ = \max\{a, 0\}$, $a \in \mathbb{R}$.

Dualidade zero, propriedades de pontos de sela e ausência de convexidade valem para funções lagrangeanas tais como

$$L(x, (u, c)) = f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle,$$

onde $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $c \in \mathbb{R}_+ = [0, \infty)$, $A \in \mathbb{R}^{m \times m}$ é uma matriz real e simétrica e $\sigma : \mathbb{R}^m \rightarrow \mathbb{R}$ é uma função contínua que satisfaz $\sigma(z) > 0, \forall z \in \mathbb{R}^m - \{0\}$ e $\sigma(0) = 0$. O problema dual gerado por lagrangeanas do tipo acima mencionadas é (não-diferenciável) convexo o qual é usualmente resolvido por técnicas de otimização convexa não-diferenciável tais como os métodos subgradientes.

Estas lagrangeanas induzem variáveis duais $(u, c) \in \mathbb{R}^m \times \mathbb{R}_+$ onde u é o multiplicador associado do termo linear e $c \geq 0$ e o parâmetro de penalidade associado ao termo de aumento.

Dado um par de iterados (u_k, c_k) , um iterado primal x_k é calculado pela regra

$$x_k \in \text{Argmin}_{x \in X} \{f_0(x) + c_k \sigma(f(x)) - \langle Au_k, f(x) \rangle\}.$$

O iterado x_k é usado para a obtenção de um subgradiente, o qual dá origem a uma direção de busca que produz o próximo iterado dual (u_{k+1}, c_{k+1}) .

O método IED é a mistura de dois outros algoritmos que são o DSG (Deflected Subgradient Method), veja [6] e NFDA (Nonsmooth Feasible Directions Algorithm), veja [18].

De maneira geral, o método IED pode ser descrito como segue. Começando em um ponto no interior do epígrafo da função dual, um subgradiente é usado para definir-se uma aproximação linear do epígrafo da função dual. Usa-se essa aproximação para encontrar uma direção de busca que fornece um ponto auxiliar. Se este ponto auxiliar estiver no interior do epígrafo, temos um passo sério e o ponto auxiliar será o novo iterado, a partir do qual o processo se repete. Caso contrário, temos um passo nulo e aplica-se uma atualização como definida no algoritmo DSG a partir do ponto original. Devido as propriedades do método DSG, neste caso, o novo iterado também se encontra no interior do epígrafo. O novo ponto assume o lugar do ponto original, a partir do qual o processo se repete até que um passo sério ocorra. Se ocorrem infinitos passos nulos, a convergência será garantida pelo método DSG. O método IED gera uma sequência no interior do epígrafo da função dual bem como uma sequência primal. Isto motiva o nome dado ao método.

Este trabalho está organizado em oito capítulos. No capítulo 1 apresenta-se as preliminares. Nos capítulos 2, 4, 5, e 6 são apresentados os métodos FDIPA, NFDA, DSG e IED, respectivamente. No capítulo 3 apresenta-se alguns métodos de otimização convexa. No capítulo 7 apresenta-se os resultados numéricos obtidos pelo método IED. Para finalizar apresentamos a conclusão e as referências bibliográficas.

1 NOÇÕES PRELIMINARES

1.1 PROBLEMA DE OTIMIZAÇÃO E DEFINIÇÕES BÁSICAS

1.1.1 Problema Primal

O problema a considerar neste trabalho é:

$$(P) \quad \begin{cases} \text{minimizar } f(x) \\ x \in D, \end{cases}$$

onde o conjunto D será chamado conjunto viável do problema (P) , os pontos de D serão chamados pontos viáveis e f será chamada função objetivo.

Definição 1.1. Dizemos que um ponto $\bar{x} \in D$ é

1. *minimizador global de (P) , se*

$$f(\bar{x}) \leq f(x), \quad \forall x \in D.$$

2. *minimizador local de (P) , se existe uma vizinhança U de \bar{x} tal que*

$$f(\bar{x}) \leq f(x), \quad \forall x \in D \cap U.$$

Definição 1.2. Dizemos que $\bar{v} \in [-\infty, +\infty)$ definido por

$$\bar{v} = \inf_{x \in D} f(x),$$

é o valor ótimo do problema (P) .

Típicamente, o conjunto viável de um problema de otimização é definido pela interseção de um sistema de desigualdades e/ou igualdades e um conjunto onde se inclui cotas superiores e

inferiores das variáveis, isto é

$$D = \{x \in X \subset \mathbb{R}^n / g(x) \leq 0, h(x) = 0\},$$

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

Teorema 1.1 (Teorema de Weierstrass). *Sejam $D \subset \mathbb{R}^n$ um conjunto compacto não vazio e $f : D \rightarrow \mathbb{R}$ uma função contínua.*

Então, f atinge um mínimo e um máximo em D e o problema (P) têm soluções globais veja [8].

Definição 1.3. *O conjunto de nível da função $f : D \rightarrow \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por*

$$L_{f,D}(c) = \{x \in D / f(x) \leq c\}.$$

Corolário 1.1.1. *Sejam $D \subset \mathbb{R}^n$ um subconjunto não vazio e $f : D \rightarrow \mathbb{R}$ contínua no conjunto D . Suponhamos que existe $c \in \mathbb{R}$ tal que o conjunto de nível $L_{f,D}(c)$ seja não vazio e compacto.*

Então o problema (P) possui uma solução global.

Lembremos agora algumas definições básicas para funções.

Definição 1.4. *Dado $D \subset \mathbb{R}^n$, uma função $f : D \rightarrow \mathbb{R}$ diz-se Lipschitziana quando existe $k > 0$ tal que, para quaisquer $x, y \in D$, tem-se $\|f(x) - f(y)\| \leq k\|x - y\|$.*

É imediato ver que toda função Lipschitziana é contínua.

Definição 1.5. *Sejam $f : D \rightarrow \mathbb{R}$ definida no aberto $D \subset \mathbb{R}^n$, $a \in D$ e $v \in \mathbb{R}^n$. A derivada direcional de f no ponto a , segundo o vetor v , é, o limite*

$$\frac{\partial f}{\partial v}(a) = \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t}$$

quando existe.

A derivada de f em a ao longo do vetor d é denotada por $f'(a; d)$ e dada pelo seguinte limite se existir,

$$f'(a; d) = \lim_{\lambda \rightarrow 0^+} \frac{f(a + \lambda d) - f(a)}{\lambda}.$$

Quando $v = e_i$, $i = 1, \dots, n$

$$\frac{\partial f}{\partial e_i}(a) = \frac{\partial f}{\partial x_i}(a) = \lim_{t \rightarrow 0} \frac{f(a + te_i) - f(a)}{t},$$

será chamado derivada parcial.

Definição 1.6. Se as derivadas parciais de uma função f de n -variáveis existem no ponto a definimos o gradiente

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

Definição 1.7. Dada $f : D \rightarrow \mathbb{R}$, com $D \subset \mathbb{R}^n$, seja $a \in D$. Diremos que f é diferenciável no ponto a quando existirem as derivadas parciais $\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)$ e para todo vetor $v = (v_1, \dots, v_n) \in \mathbb{R}^n$, com $a + v \in D$, se tenha

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

Quando f é diferenciável em todo os ponto de D , dizemos simplesmente que f é diferenciável.

1.2 DIREÇÃO DE DESCIDA / DIREÇÃO VIÁVEL

Definição 1.8. Um conjunto $K \subset \mathbb{R}^n$ chama-se cone quando

$$d \in K \Rightarrow td \in K, \forall t \in \mathbb{R}_+.$$

Definição 1.9. Dizemos que $d \in \mathbb{R}^n$ é uma direção viável em relação ao conjunto D no ponto $a \in D$, quando existe $\varepsilon > 0$ tal que

$$a + td \in D, \forall t \in [0, \varepsilon].$$

Denotemos por $\mathcal{V}_D(a)$ o conjunto de todas as direções viáveis em relação ao conjunto D no ponto $a \in D$.

Definição 1.10. Dizemos que $d \in \mathbb{R}^n$ é uma direção de descida de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no ponto $a \in \mathbb{R}^n$, se existe $\varepsilon > 0$ tal que

$$f(a + td) < f(a), \forall t \in (0, \varepsilon].$$

Denotemos por $\mathcal{D}_f(a)$ o conjunto de todas as direções de descida da função f no ponto a .

Lema 1.1 (Direções de descida). Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no ponto $a \in \mathbb{R}^n$. Então

1. Para todo $d \in \mathcal{D}_f(a)$, tem-se $(f'(a))^T d \leq 0$.
2. Se $d \in \mathbb{R}^n$ satisfaz $(f'(a))^T d < 0$, tem-se que $d \in \mathcal{D}_f(a)$.

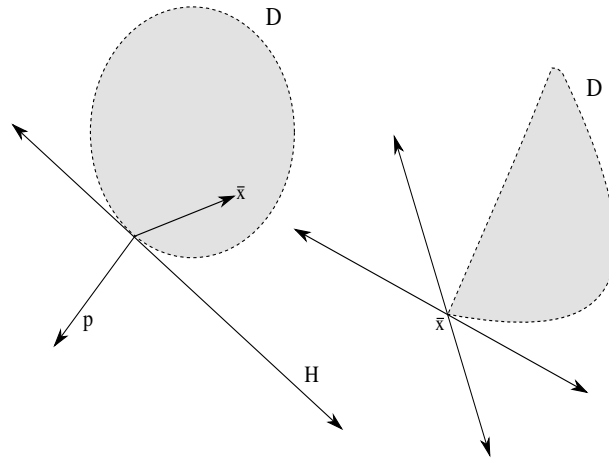


Figura 1.1: Existência do hiperplano suporte de um conjunto D

1.3 CONJUNTOS CONVEXOS E PROPRIEDADES

Definição 1.11. Um conjunto $D \subset \mathbb{R}^n$ é dito convexo se para quaisquer $x, y \in D$ e para todo $\lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda)y \in D.$$

Definição 1.12. Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $a \in D$. O cone normal no ponto a em relação ao conjunto D é dado por

$$\mathcal{N}_D = \{d \in \mathbb{R}^n / d^T(x - a) \leq 0 \ \forall x \in D\}.$$

Definição 1.13. Um hiperplano H em \mathbb{R}^n é um coleção de pontos da forma

$$H(p, \alpha) = \{x \in \mathbb{R}^n / p^T x = \alpha\}$$

onde p é um vetor não nulo em \mathbb{R}^n e α um escalar.

Um hiperplano H define dois semi-espacos fechados $H^+(p, \alpha) = \{x \in \mathbb{R}^n / p^T x \geq \alpha\}$ e $H^-(p, \alpha) = \{x \in \mathbb{R}^n / p^T x \leq \alpha\}$ e dois semi-espacos abertos $\{x \in \mathbb{R}^n / p^T x > \alpha\}$ e $\{x \in \mathbb{R}^n / p^T x < \alpha\}$.

Definição 1.14. Seja D um subconjunto não vazio de \mathbb{R}^n , e seja $a \in \text{fr}D$. Um hiperplano $H(p, \alpha) = \{x \in \mathbb{R}^n / p^T x = \alpha\}$, onde $\alpha = p^T a$, é chamado um hiperplano suporte de D em a se para qualquer $x \in D \subset H^+$, tem-se $p^T x \geq \alpha$ ou para qualquer $x \in D \subset H^-$, tem-se $p^T x \leq \alpha$.

Teorema 1.2. Seja D um conjunto convexo não vazio em \mathbb{R}^n , e seja $a \in \text{fr}D$. Então existe um hiperplano que suporta D em a , isto é, existe um vetor não nulo p tal que $p^T x \leq \alpha$, onde $\alpha = p^T a$, para cada $x \in \text{cl}D$ (Ver Teorema 2.4.7 [3]).

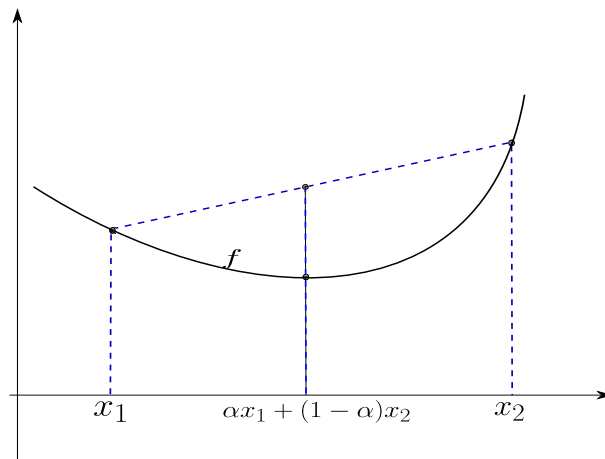


Figura 1.2: Função convexa

1.4 FUNÇÕES CONVEXAS

Definição 1.15. Seja $f : D \rightarrow \mathbb{R}$, onde D é conjunto convexo não vazio em \mathbb{R}^n . A função f é convexa em D se:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2),$$

para todo $x_1, x_2 \in D$ e para todo $\lambda \in [0, 1]$.

Definição 1.16. Seja $f : D \rightarrow \mathbb{R}$, onde D é conjunto convexo não vazio em \mathbb{R}^n . A função f é estritamente convexa em D se

$$f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2),$$

para quaisquer $x_1, x_2 \in D$ com $x_1 \neq x_2$ e todo $\lambda \in (0, 1)$.

A função f diz-se fortemente convexa com módulo $\gamma > 0$, quando para quaisquer $x_1, x_2 \in D$ e $\lambda \in (0, 1)$, tem-se

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) - \gamma \lambda(1 - \lambda) \|x_1 - x_2\|^2.$$

Definição 1.17. Seja $f : D \rightarrow \mathbb{R}$, onde D é um conjunto convexo não vazio em \mathbb{R}^n . A função f é concava em D se $-f$ é convexa em D .

Lema 1.2. Seja D um subconjunto convexo não vazio de \mathbb{R}^n , e $f : D \rightarrow \mathbb{R}$ uma função convexa. Então o conjunto de nível $L_{f,D}(\alpha) = \{x \in D / f(x) \leq \alpha\}$, onde α é um número real, é um conjunto convexo.

Definição 1.18. *Seja D um conjunto não vazio do \mathbb{R}^n , e seja $f : D \rightarrow \mathbb{R}$ uma função. O epígrafo de f , denotado por $\text{epi}(f)$, é um subconjunto de \mathbb{R}^{n+1} definido por*

$$\text{epi}(f) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} / y \in \mathbb{R}, y \geq f(x)\}.$$

Teorema 1.3. *Seja D um conjunto convexo não vazio de \mathbb{R}^n e seja $f : D \rightarrow \mathbb{R}$ uma função. Então f é convexa em D se, e somente se o $\text{epi}(f)$ é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

1.4.1 Funções convexas diferenciáveis

Nesta seção apresentaremos algumas propriedades de funções convexas diferenciáveis cujas demonstrações poder ser encontradas em [19].

Teorema 1.4. *Sejam $D \subset \mathbb{R}^n$ aberto e convexo, $f : D \rightarrow \mathbb{R}$, $f \in C^1(D)$. Então f é convexa se, e somente se $f(y) \geq f(x) + \nabla f(x)^T(y - x)$, para todo $x, y \in D$.*

Teorema 1.5. *Seja $D \subset \mathbb{R}^n$ um subconjunto aberto e convexo, e $f : D \rightarrow \mathbb{R}$ uma função tal que $f \in C^1(D)$. Então, f é convexa se, e somente se para todo $x, y \in D$, $\nabla f(x)^T(y - x) \leq \nabla f(y)^T(y - x)$.*

As funções convexas de classe C^2 são caracterizadas pelo seguinte resultado:

Teorema 1.6. *Seja $D \subset \mathbb{R}^n$ um conjunto aberto e convexo, e $f : D \rightarrow \mathbb{R}$ uma função com $f \in C^2(D)$. Então f é convexa se, e somente se $\nabla^2 f(x) \leq 0$ para todo $x \in D$.*

Definição 1.19. *Chamamos problema de programação convexa ao problema*

$$(P) \begin{cases} \text{minimizar } f(x) \\ x \in D, \end{cases}$$

onde D é um conjunto convexo e f é uma função convexa.

Teorema 1.7. *Em um problema de programação convexa, todo minimizador local é global. O conjunto dos minimizadores é convexo. Se f é estritamente convexa, não pode haver mais de um minimizador.*

1.4.2 Propriedades básicas

Proposição 1.1 (Convexidade do supremo de funções convexas). *Sejam $D \in \mathbb{R}^n$ um conjunto convexo e $f_i : D \rightarrow \mathbb{R}$, $i \in I$, funções convexas em D , onde I é um conjunto qualquer. Suponhamos que exista $\beta \in \mathbb{R}$ tal que $f_i(x) \leq \beta$ para todo $x \in D$ e $i \in I$.*

Então a função

$$f : D \rightarrow \mathbb{R}, f(x) = \sup_{i \in I} f_i(x)$$

é convexa em D .

Definição 1.20. Seja D um subconjunto convexo de \mathbb{R}^n . Dizemos que $g : D \rightarrow \mathbb{R}^m$ ($g = (g_1, g_2, \dots, g_m)$) é convexa em D se todas as funções $g_i : D \rightarrow \mathbb{R}$, $i = 1, \dots, m$, são convexas em D .

Corolário 1.4.1. Seja Ω um subconjunto convexo de \mathbb{R}^n . Sejam $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função convexa e $h : \mathbb{R}^n \rightarrow \mathbb{R}^l$ uma função afim. Então o conjunto

$$D = \{x \in \Omega / h(x) = 0, g(x) \leq 0\}$$

é convexo.

Teorema 1.8 (Continuidade de funções convexas). Seja D um conjunto convexo e aberto não vazio de \mathbb{R}^n e $f : D \rightarrow \mathbb{R}$ uma função convexa.

Então f é localmente Lipschitz-contínua em D . Em particular, f é contínua em D .

Teorema 1.9. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Suponhamos que exista $c \in \mathbb{R}$ tal que o conjunto de nível

$$L_{f, \mathbb{R}^n}(c) = \{x \in \mathbb{R}^n / f(x) \leq c\}$$

é não vazio e limitado.

Então $L_{f, \mathbb{R}^n}(t)$ é limitado para todo $t \in \mathbb{R}$.

Teorema 1.10 (Compacidade de conjuntos de nível de uma função fortemente convexa). Suponhamos que a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja fortemente convexa no \mathbb{R}^n . Então o conjunto de nível

$$L_{f, \mathbb{R}^n}(c) = \{x \in \mathbb{R}^n / f(x) \leq c\}$$

é compacto para todo $c \in \mathbb{R}$.

Corolário 1.4.2. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função fortemente convexa e $\emptyset \neq D \subset \mathbb{R}^n$ um conjunto fechado qualquer.

Então f tem um único minimizador em D .

1.4.3 Funções convexas não diferenciáveis

Nesta seção vamos ver algumas propriedades importantes de funções convexas. Veremos também a definição de subgradiente e algumas de suas propriedades importantes na teoria de

otimização.

Lema 1.3. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Então para todo $a \in \mathbb{R}^n$, existem as derivadas direcionais em cada direção $d \in \mathbb{R}^n$. Além disso,*

$$f(a + \alpha d) \geq f(a) + \alpha f'(a; d), \quad \forall \alpha \in \mathbb{R}_+.$$

Proposição 1.2. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Então para qualquer das sequências $\{x^k\} \rightarrow x$, $\{d^k\} \rightarrow d$ ($k \rightarrow \infty$), tem-se que*

$$\limsup_{k \rightarrow \infty} f'(x^k; d^k) \leq f'(x; d).$$

Mais ainda, se f é diferenciável em \mathbb{R}^n , então a derivada de f é contínua em \mathbb{R}^n .

Definição 1.21. *Seja D um subconjunto convexo não vazio de \mathbb{R}^n e seja $f : D \rightarrow \mathbb{R}$ uma função convexa. Então s é chamado um subgradiente de f em $a \in D$ se*

$$f(x) \geq f(a) + s^T(x - a), \quad \text{para todo } x \in D.$$

O conjunto de todos os subgradientes de f em a é chamado subdiferencial e é denotado por $\partial f(a)$. Além disso, pelo lema 1.3 temos

$$\partial f(a) = \{x \in \mathbb{R}^n / f'(a; d) \geq \langle x, d \rangle \quad \forall d \in \mathbb{R}^n\}.$$

Vejamos agora alguns importantes resultados sobre subgradientes que também pode ser encontrados em [1] e [3].

O teorema seguinte garante a existência de subgradiente de uma função convexa.

Teorema 1.11. *Seja D um conjunto convexo não vazio do \mathbb{R}^n e seja $f : D \rightarrow \mathbb{R}$ uma função convexa. Então para cada $a \in \text{int } D$, existe um vetor s tal que o hiperplano suporte ao $\text{epi}(f)$ no ponto $(a, f(a))$ é definido por*

$$H = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} / y = f(x) + s^T(x - a)\}.$$

Em particular

$$f(x) \geq f(a) + s^T(x - a), \quad \text{para todo } x \in D$$

isto é, s é um subgradiente de f em a .

Teorema 1.12 (O subdiferencial da função convexa). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Então para todo $a \in \mathbb{R}^n$, o conjunto $\partial f(a)$ é convexo, compacto e não vazio. Além disso, para*

todo $d \in \mathbb{R}^n$, tem-se

$$f'(a; d) = \max_{y \in \partial f(a)} \langle y, d \rangle.$$

Proposição 1.3. *Uma função convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é diferenciável no ponto $a \in \mathbb{R}^n$ se, e somente se o conjunto $\partial f(a) = \{\nabla f(a)\}$.*

Teorema 1.13. *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e $D \subset \mathbb{R}^n$ um conjunto convexo. Então $a \in \mathbb{R}^n$ é um minimizador de f em D se, e somente se*

$$\exists y \in \partial f(a) \text{ tal que } 0 \in \partial f(a) + \mathcal{N}_D(a).$$

Se $D = \mathbb{R}^n$ tem-se $0 \in \partial f(a)$.

Proposição 1.4 (Continuidade de subdiferencial). *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa, $\{x^k\} \rightarrow a$ ($k \rightarrow \infty$) e $y^k \in \partial f(x^k)$ para todo $k \in \mathbb{N}$. Então a sequência $\{y^k\}_{k \in \mathbb{N}}$ é limitada e todos os seus pontos de acumulação pertencem a $\partial f(a)$.*

Definição 1.22. *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e $\varepsilon \in \mathbb{R}_+$. Dizemos que $s \in \mathbb{R}^n$ é um ε -subgradiente de f no ponto $a \in \mathbb{R}^n$ quando $f(x) \geq f(a) + s^T(x - a) - \varepsilon$, $\forall x \in \mathbb{R}^n$.*

O conjunto de todos os ε -subgradiente de f em a , denotado por $\partial_\varepsilon f(a)$, se chama o ε -subdiferencial de f em a . Se $\varepsilon = 0$ chamaremos simplesmente subdiferencial.

Teorema 1.14 (O ε -subdiferencial). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Então para todo $x \in \mathbb{R}^n$ e todo $\varepsilon \in \mathbb{R}_+$, o conjunto $\partial_\varepsilon f(x)$ é convexo, compacto e não vazio.*

1.5 ELEMENTOS DA TEORIA DE DUALIDADE

Consideremos o problema de otimização restrita

$$(P) \quad \begin{cases} \text{minimizar } f_0(x) \\ f(x) = 0 \\ x \in X, \end{cases}$$

onde X é um conjunto compacto de \mathbb{R}^n e as funções $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são contínuas.

Seja $X_0 = \{x \in X / f(x) = 0\}$ o conjunto das restrições do problema (P). Seja

$$M_P = \inf_{x \in X_0} f_0(x)$$

o qual será o valor ótimo do problema (P). Seja Λ um subconjunto não vazio do $\mathbb{R}^m \times \mathbb{R}$ e definamos a função $L : \mathbb{R}^n \times \Lambda \rightarrow \mathbb{R}$ o qual chamaremos a função lagrangeana com variáveis duais em Λ do problema primal. Seja $u \in \mathbb{R}^m$ e $c \in \mathbb{R}$ então definimos cada elemento de Λ como um par ordenado (u, c) .

Agora definimos $H : \Lambda \rightarrow \mathbb{R} \cup \{-\infty\}$ como

$$H(u, c) = \inf_{x \in X} L(x, (u, c))$$

a função dual, então o problema dual para o problema primal (P) será definido como

$$(D) \quad \begin{cases} \text{maximizar } H(u, c) \\ (u, c) \in \Lambda, \end{cases}$$

e portanto o valor ótimo dual é dado como

$$M_D = \sup_{(u, c) \in \Lambda} H(u, c).$$

Definição 1.23. Dizemos que a dualidade é débil com respeito a L se vale

$$\inf_{x \in X} L(x, (u, c)) \leq M_P,$$

para cada $(u, c) \in \Lambda$

Definição 1.24. Dizemos que a dualidade é forte (dualidade zero) com respeito a L se vale

$$M_P = M_D, \text{ que é } \sup_{(u, c) \in \Lambda} \inf_{x \in X} L(x, (u, c)) = M_P.$$

Para mais detalhes veja [4], as condições para dualidade brecha zero são dados através de um conjunto de requerimentos sobre a lagrangeana e o dado do problema primal. Listaremos isto no requirements a seguir

(A_1) Para todo $\alpha > M_P$, o conjunto de nível $\{z \in X / f_0(z) \leq \alpha\}$ é compacto.

(H_1) $f_0(x) = L(x, (u, c))$ para todo $x \in X_0$ e todo $(u, c) \in \Lambda$.

Assumamos que existe un subconjunto $\Lambda_0 \subset \Lambda$ tal que

($H_2(\Lambda_0)$) $\forall \alpha < M_P$ e $\forall \sigma > 0$

$$\sup_{(u, c) \in \Lambda_0} \left[\inf_{x \in X} L(x, (u, c)) \right] > \alpha.$$

$$d(x, X_0) \geq \sigma$$

$(H_3(\Lambda_0))$

$$f_0(x) \leq L(x, (u, c))$$

para todo $x \in X$, $(u, c) \in \Lambda_0$

Teorema 1.15 (Dualidade forte). *Assuma que existe $\Lambda_0 \subset \mathbb{R}^n \times \mathbb{R}_+$ tal que (H_1) , $(H_2(\Lambda_0))$ e $(H_3(\Lambda_0))$ são satisfeitas e suponha também que a condição (A_1) vale para f_0 , então $M_P = M_D$.*

Demonstração. Seja $M_{D,0} = \sup_{\lambda \in \Lambda_0} H(\lambda)$. Mostremos que $M_P = M_{D,0}$. Então claramente $M_P = M_D$. Note que a dualidade fraca segue de (H_1) , assim $M_P \neq M_{D,0}$. Suponha que exista α tal que $M_P > \alpha > M_{D,0}$. Neste caso existe $\delta > 0$ tal que

$$X(\delta) = \{x \in X / d(x, X_0) \leq \delta\} \subset \{x \in X / f_0(x) > \alpha\}. \quad (1.1)$$

Com efeito, se um número δ não existe então para cada inteiro positivo podemos encontrar um elemento x_k tal que $d(x_k, X_0) \leq \frac{1}{k}$ e $f_0(x_k) \leq \alpha$. De (A_1) podemos assumir sem perda de generalidade que exista $\lim_{k \rightarrow \infty} x_k = x_0$. Então $x_0 \in X_0$ e $f_0(x_0) \leq \alpha < M_P = \inf_{x \in X_0} f_0(x)$ o qual é impossível.

Seja $\delta > 0$ um número tal que (5.2) mantem-se. Então para todo $\lambda \in \Lambda_0$ podemos escrever

$$\inf_{x \in X} L(x, \lambda) = H(\lambda) \leq M_{D,0} < \alpha \leq \inf_{x \in X(\delta)} f_0(x) \leq \inf_{x \in X(\delta)} L(x, \alpha), \quad (1.2)$$

onde usamos $(H_3(\Lambda_0))$ na última desigualdade. A expressão acima implica que

$$H(\lambda) = \inf_{\substack{x \in X \\ d(x, X_0) \geq \delta}} L(x, \lambda). \quad (1.3)$$

Para todo $\lambda \in \Lambda_0$. Como $\alpha < M_P$, podemos usar $(H_2(\Lambda_0))$ e encontrar $\bar{\lambda} \in \Lambda_0$ tal que

$$\inf_{\substack{x \in X \\ d(x, X_0) \geq \delta}} L(x, \bar{\lambda}) \geq \alpha = M_{D,0} + \varepsilon, \quad (1.4)$$

para $\varepsilon = \alpha - M_{D,0} > 0$. Usando (5.4) e (5.5), obtemos

$$M_{D,0} \geq H(\bar{\alpha}) \quad \inf_{\substack{x \in X \\ d(x, X_0) \geq \delta}} L(x, \bar{\lambda}) \geq \alpha = M_{D,0} + \varepsilon,$$

é uma contradição. □

Definição 1.25. Considere o conjunto de variáveis duais dado por $\Lambda = \mathbb{R}^n \times \mathbb{R}_+$ e a função contínua $\sigma : \mathbb{R}^m \rightarrow \mathbb{R}$ tal que $\sigma(z) > 0$ para todo $z \in \mathbb{R}^m \setminus \{0\}$ e $\sigma(0) = 0$. Seja $A \in \mathbb{R}^{m \times m}$ uma matriz simétrica. A função lagrangeana $L : \mathbb{R}^n \times \Lambda \rightarrow \mathbb{R}$ é definida como

$$L(x, u, c) = f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle.$$

A lagrangeana na definição 1.25 verifica a hipótese do Teorema 1.15 para a escolha $\Lambda_0 = \{0\} \times \mathbb{R}_+$. De fato, note que a hipótese básica em X y f_0 trivialmente satisfazem (A_1) . Também pode se verificar em uma maneira similar que as condições (H_1) , $(H_2(\Lambda_0))$ e $(H_3(\Lambda_0))$ também valem. Com efeito, $L(x, (u, c)) = f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle = f_0(x)$ para todo $x \in X_0$ e todo $(u, c) \in \Lambda$, desde que $\sigma(0) = 0$. Para todo $(u, c) \in \Lambda_0$ têm-se que $L(x, (u, c)) = f_0(x) + c\sigma(f(x)) \geq f_0(x)$, para todo $x \in X$ como consequência $(H_3(\Lambda_0))$ é válida. Mostremos agora que (A_1) implica a condição $(H_2(\Lambda_0))$. Assuma que a condição última não vale. Isto significa que existe $\alpha_0 < M_P$ e $\delta_0 > 0$ tal que

$$\inf_{\substack{x \in X \\ d(x, X_0) \geq \delta_0}} L(x, \lambda) \leq \alpha_0$$

para todo $\alpha \in \Lambda$, onde $\lambda = (u, c)$. Fixe $k_0 \in \mathbb{N}$ tal que $\alpha_0 + \frac{1}{k} < M_P$ para todo $k \geq k_0$ então para todo $\alpha \in \Lambda_0$ e $k \geq k_0$ temos

$$\inf_{\substack{x \in X \\ d(x, X_0) \geq \delta_0}} L(x, \alpha) \leq \alpha_0 + \frac{1}{k},$$

tomando a sequência $(k, 0) \in \Lambda_0$, conclui-se que existe uma sequência $\{x_k\} \subset \{x \in X / d(x, X_0) \geq \delta_0\}$ com

$$f_0(x_k) \leq f_0(x_k) + k\sigma(f(x_k)) \leq \alpha_0 + \frac{1}{k},$$

onde usamos também que $\sigma(f(x_k)) > 0$, por (A_1) para $\alpha = \alpha_0 + \frac{1}{k_0}$, a sequência é limitada e

portanto temos uma subsequência $\{x_{k_j}\}$ que converge para algum \bar{x} . Então temos

$$f_0(\bar{x}) = \lim_j f_0(x_{k_j}) \leq \lim_j f_0(x_{k_j}) + k_j \sigma(f(x_{k_j})) \leq \lim_j \alpha_0 + \frac{1}{k_j} = \alpha_0,$$

$\lim_j \sigma(f(x_{k_j})) = 0$, e pela continuidade $f(\bar{x}) = \lim_j f(x_{k_j}) = 0$. Consequentemente, $\bar{x} \in X_0$ e então $f_0(\bar{x}) \geq M_P > \alpha_0$, em contradição com a expressão acima. Portanto $(H_2(\Lambda_0))$ vale. Como uma consequência, o esquema de dualidade possui a propriedade de dualidade brecha zero. Definindo agora a função dual $H : \mathbb{R}^m \times \mathbb{R}_+ \rightarrow \mathbb{R}$ como

$$H(u, c) = \min_{x \in X} [f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle].$$

O seguinte teorema mostra que a função dual é côncava.

Teorema 1.16. *Para qualquer problema primal do tipo (P), a função dual H é côncava.*

Demonstração. Pela linearidade da lagrangeana em relação aos multiplicadores, para todo $x \in X$, $u_1, u_2 \in \mathbb{R}^n$, $c_1, c_2 \in \mathbb{R}_+$ e $\alpha \in [0, 1]$, têm-se que

$$\begin{aligned} L(x, \alpha u_1 + (1 - \alpha)u_2, \alpha c_1 + (1 - \alpha)c_2) \\ = \alpha L(x, u_1, c_1) + (1 - \alpha)L(x, u_2, c_2). \end{aligned}$$

Por isso,

$$\begin{aligned} H(\alpha u_1 + (1 - \alpha)u_2, \alpha c_1 + (1 - \alpha)c_2) \\ = \inf_{x \in X} L(x, \alpha u_1 + (1 - \alpha)u_2, \alpha c_1 + (1 - \alpha)c_2) \\ = \inf_{x \in X} \{ \alpha L(x, u_1, c_1) + (1 - \alpha)L(x, u_2, c_2) \} \\ \geq \alpha \inf_{x \in X} L(x, u_1, c_1) + (1 - \alpha) \inf_{x \in X} L(x, u_2, c_2) \\ = \alpha H(u_1, c_1) + (1 - \alpha)H(u_2, c_2) \end{aligned}$$

o qual mostra a concavidade de H . □

Desde que H é côncava concluímos portanto que o problema dual de (P) é um problema convexo dado por

$$(D) \quad \begin{cases} \text{maximizar } H(u, c) \\ (u, c) \in \mathbb{R}^m \times \mathbb{R}_+ \end{cases} \quad (1.5)$$

e desde que H é finito em $\mathbb{R}^m \times \mathbb{R}_+$, (D) será um problema de otimização sem restrições.

Vejamos alguns casos particulares como por exemplo quando $A = I$ e $\sigma(\cdot) = \|\cdot\|$, onde $\|\cdot\|$ é a (Euclidiana) l_2 -norma, o lagrangeano dado na definição 1.25 obtém-se o denominado

lagrangeano forte, introduzido em [22], exemplo 11.58 e para o caso $A = 0$, a lagrangeana obtem-se a função de penalidade com termo penalidade dado por $c\sigma(f(x))$.

Denotaremos o conjunto solução do problema (P) como $S(P)$ tal que um elemento de $S(P)$ seja representado por \bar{x} . Denotaremos o conjunto solução do problema (D) como $S(D)$ tal que um elemento de $S(D)$ seja representado por $\bar{z} = (\bar{u}, \bar{c})$. Como representação auxiliar introduziremos para (u, c) o conjunto

$$X(u, c) = \text{Argmin}_{x \in X} [f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle],$$

que é o conjunto de minimizadores da função $G(x) = f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle$ no conjunto compacto X .

Lema 1.4. *Seja H a função dual do problema (P) . Para cada $(u, c) \in \mathbb{R}^m \times \mathbb{R}_+$ temos $\tilde{x} \in X(u, c)$. Então*

1. $(-Af(\tilde{x}), \sigma(f(\tilde{x}))(1 + \gamma)) \in \partial_{c\gamma\sigma(f(\tilde{x}))}H(u, c)$ para cada $\gamma \geq 0$.
2. Se $(u, c) \in S(D)$, então $(u, d) \in S(D)$ para cada $d > c$. Nesta situação, para cada $\hat{x} \in X(u, d)$ devemos ter que $f(\hat{x}) = 0$.

Demonstração. (a) Devemos provar que para todo $(u', c') \in \mathbb{R}^m \times \mathbb{R}_+$ vale

$$H(u', c') \leq H(u, c) + \langle u' - u, -Af(\tilde{x}) \rangle + (c' - c)(1 + \gamma)\sigma(f(\tilde{x})).$$

De fato, pela definição de H temos que

$$\begin{aligned} H(u', c') &= \min_{x \in X} L(x, (u', c')) \\ &\leq f_0(\tilde{x}) + c\sigma(f(\tilde{x})) - \langle Au, f(\tilde{x}) \rangle + (c' - c)\sigma(f(\tilde{x})) + \langle u' - u, -Af(\tilde{x}) \rangle \\ &\leq H(u, c) + \langle u' - u, -Af(\tilde{x}) \rangle + (c' - c)(1 + \gamma)\sigma(f(\tilde{x})) - (c' - c)\gamma\sigma(f(\tilde{x})) \\ &\leq H(u, c) + \langle u' - u, -Af(\tilde{x}) \rangle + (c' - c)(1 + \gamma)\sigma(f(\tilde{x})) + c\gamma\sigma(f(\tilde{x})) \end{aligned}$$

onde usamos o fato que $c' \geq 0$ na última desigualdade. A prova (a) é completada.

(b) Desde que $(u, c) \in S(D)$, devemos ter $H(u, c) \geq H(u, d)$. Por outro lado, tomando $\hat{x} \in X(u, d)$, onde $d > c$. Por item (a), temos

$$\begin{aligned} H(u, d) \leq H(u, c) &\leq H(u, d) + \langle u - u, -Af(\hat{x}) \rangle + (c - d)\sigma(f(\hat{x})) \\ &= H(u, d) + (c - d)\sigma(f(\hat{x})) \\ &\leq H(u, d). \end{aligned}$$

Onde também usamos o fato que $\sigma \geq 0$. Portanto, $H(u, c) = H(u, d)$ e (u, d) é também uma solução dual. Desde que $d - c > 0$ devemos ter $f(\hat{x}) = 0$.

□

Em particular temos o seguinte resultado.

Corolário 1.5.1. *Dado a função dual H do problema (P) e um par coordenado $(\bar{u}, \bar{c}) \in \mathbb{R}^m \times \mathbb{R}_+$, tome $\hat{x} \in X(\bar{u}, \bar{c})$, então $(-Af(\hat{x}), \sigma(f(\hat{x}))) \in \partial H(\bar{u}, \bar{c})$.*

Teorema 1.17. *Seja $M_P \geq M_D$ e suponha que $\bar{x} \in X(\bar{u}, \bar{c})$. Então \bar{x} é a solução do (P) e (\bar{u}, \bar{c}) é uma solução de (D) se, e somente se $f(\bar{x}) = 0$.*

Demonstração. É suficiente provar que \bar{x} é uma solução, uma vez que o inverso é trivial. Suponha que $f(\bar{x}) = 0$ e $\bar{x} \in X(\bar{u}, \bar{c})$, então $L(\bar{x}, \bar{u}, \bar{c}) = f_0(\bar{x}) \geq M_P$. Por outro lado, $L(\bar{x}, \bar{u}, \bar{c}) = \min_{x \in X} L(x, \bar{u}, \bar{c}) = H(\bar{u}, \bar{c}) \leq M_D \leq M_P$. Onde usamos dualidade fraca na última desigualdade. Assim temos $f_0(\bar{x}) = M_P$ e portanto \bar{x} é uma solução de (P) .

□

2 FDIPA - FEASIBLE DIRECTIONS INTERIOR POINT ALGORITHM - UM MÉTODO DE PONTOS INTERIORES PARA OTIMIZAÇÃO NÃO LINEAR DIFERENCIÁVEL

O algoritmo que apresentaremos neste capítulo, denominado FDIPA (Feasible Directions Interior Point Algorithm), foi proposto por José Herkovits em [10]. Trata-se de um método para solução de problemas não-lineares com restrições de igualdade e/ou desigualdade. É um algoritmo de fácil implementação onde não é necessária a solução de subproblemas quadráticos. Também não se trata de um método de penalidades ou barreiras. O método requer somente a solução de dois sistemas de equações lineares com a mesma matriz principal em cada iteração, seguida de uma busca linear inexata. O método converge globalmente para pontos de Karush-Kuhn-Tucker.

Apesar de ser um método para problemas de otimização não linear, este método pode ser utilizado em problemas de otimização lineares devido ao fato de que seus sistemas fornecem uma direção de busca viável.

2.1 O MÉTODO FDIPA

Consideremos o seguinte problema de otimização não linear diferenciável com restrições de desigualdade

$$(Q) \begin{cases} \text{minimizar } f(x) \\ g(x) \leq 0 \end{cases}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são funções diferenciáveis.

Denotaremos por $\nabla g(x) \in \mathbb{R}^{m \times n}$ a matriz das derivadas de g , chamaremos $\lambda \in \mathbb{R}^m$ o vetor de variáveis duais, $\Omega = \{x \in \mathbb{R}^n / g(x) \leq 0\}$ é o conjunto viável e Ω^0 seu interior. Define-se a função Lagrangiana associada ao problema (Q) por

$$L(x, \lambda) = f(x) + \lambda^T g(x),$$

cuja matriz Hessiana é dada por

$$\Gamma(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x).$$

Denotando por $G(x)$ a matriz diagonal tal que $G_{ii}(x) = g_i(x)$, as condições de otimalidade de primeira ordem de Karush-Kuhn-Tucker (K-K-T) podem ser escrita como

$$\nabla f(x) + \nabla g(x)\lambda = 0 \tag{2.1}$$

$$G(x)\lambda = 0 \tag{2.2}$$

$$g(x) \leq 0 \tag{2.3}$$

$$\lambda \geq 0. \tag{2.4}$$

Um ponto (x^*, λ^*) satisfazendo as condições (2.1) e (2.2) será chamado par estacionário do problema e um ponto de K-K-T se satisfizer as condições (2.1)-(2.4). Um vetor x^* é um ponto estacionário se existe λ^* tal que (x^*, λ^*) é um par estacionário.

2.2 HIPÓTESE PARA A CONVERGÊNCIA DO MÉTODO

Para a convergência global do método, as seguintes hipóteses são consideradas:

Hipótese 2.1. Existe um número real a tal que o conjunto $\Omega_a = \{x \in \Omega / f(x) \leq a\}$ é compacto e tem interior Ω_a^0 não vazio.

Hipótese 2.2 $\forall x \in \Omega_a^0$ tem-se $g(x) < 0$.

Hipótese 2.3 As funções f e g são continuamente diferenciáveis em Ω_a e suas derivadas satisfazem à condição de Lipschitz.

Hipótese 2.4 (Condição de regularidade) Para cada $x \in \Omega_a$ os vetores $\nabla g_i(x)$ para i tal que $g_i(x) = 0$, são linearmente independentes.

Dado um ponto interior inicial, o algoritmo FDIPA gera uma sequência $\{x^k\}_{k \in \mathbb{N}}$ de pontos

interiores ao conjunto viável que satisfaz

$$f(x^{k+1}) < f(x^k) \text{ e } g(x^k) < 0, \text{ para } i = 1, 2, \dots, m$$

e converge para um ponto de K-K-T. Em cada iteração uma direção de busca que é uma direção de descida viável para o problema é calculada através da solução de 2 sistemas lineares com a mesma matriz. Uma busca linear é então realizada para garantir que o novo ponto fique no interior de Ω e também para garantir a convergência global.

Consideremos os sistema de equações (2.1)-(2.2). Pondo

$$Y = (x, \lambda) \text{ e } \phi(Y) = \begin{bmatrix} \nabla f(x) + \nabla g(x)\lambda \\ G(x)\lambda \end{bmatrix},$$

obtemos

$$\nabla \phi(Y) = \begin{bmatrix} \Gamma(x, \lambda) & \nabla g(x) \\ \Lambda \nabla g(x)^T & G(x) \end{bmatrix}$$

onde $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$ é uma matriz $m \times m$. Uma iteração de Newton para resolver o sistema de equações lineares $\phi(x) = 0$, com o ponto $Y^k = (x^k, \lambda^k)$ na iteração k, define um novo ponto $Y^{k+1} = (x^{k+1}, \lambda^{k+1})$ que é solução do sistema linear

$$\nabla \phi(Y^k)(Y - Y^k)^T = -\phi(Y^k)^T$$

o qual pode ser reescrito como

$$\begin{bmatrix} B^k & \nabla g(x^k) \\ \Lambda^k \nabla g^T & G(x^k) \end{bmatrix} \begin{bmatrix} x - x^k \\ \lambda - \lambda^k \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) + \nabla g(x^k)\lambda^k \\ G(x^k)\lambda^k \end{bmatrix} \quad (2.5)$$

onde substitui-se $\Gamma(x^k, \lambda^k)$ por B^k .

A matriz B^k pode ser tomada igual a uma aproximação quasi-Newton de $\Gamma(x, \lambda^k)$ ou pela matriz identidade. Contudo, para a convergência global, B^k deve ser definida positiva.

Pondo $d_\alpha = x - x^k$ e $\lambda_\alpha = \lambda$, (2.5) pode ser reescrito como

$$B^k d_\alpha + \nabla g(x^k)\lambda_\alpha = -\nabla f(x^k) \quad (2.6)$$

e

$$\Lambda^k \nabla g^T(x^k)d_\alpha + G(x^k)\lambda_\alpha = 0, \quad (2.7)$$

A solução $(d_\alpha, \lambda_\alpha)$ deste sistema fornece uma direção d_α e uma nova estimativa λ_α para λ . Em

[10], Herskovits prova que d_α é uma direção de descida para f , e nula num ponto estacionário. Entretanto, d_α pode não ser uma direção viável, pois quando alguma restrição aproxima-se a zero, d_α tende a uma direção tangente do conjunto viável.

De fato, reescrevendo a equação (2.7) tem-se

$$\lambda_i^k \nabla g_i^T(x^k) d_\alpha + g_i(x^k) \lambda_{\alpha_i} = 0, \quad i = 0, 1, \dots, m \quad (2.8)$$

o que implica que $\nabla g_i(x^k) = 0$ para todo i tal que $g_i(x^k) = 0$. Para evitar esse problema, define-se um novo sistema em d e $\bar{\lambda}$, onde acrescenta-se no lado direito de (2.7) o vetor $-\rho^k \Lambda^k$ (ρ^k é chamado parâmetro de deflexão)

$$B^k d + \nabla g(x^k) \bar{\lambda} = -\nabla f(x^k)$$

$$\Lambda^k \nabla g^T(x^k) d + G(x^k) \bar{\lambda} = -\rho^k \Lambda^k, \quad (2.9)$$

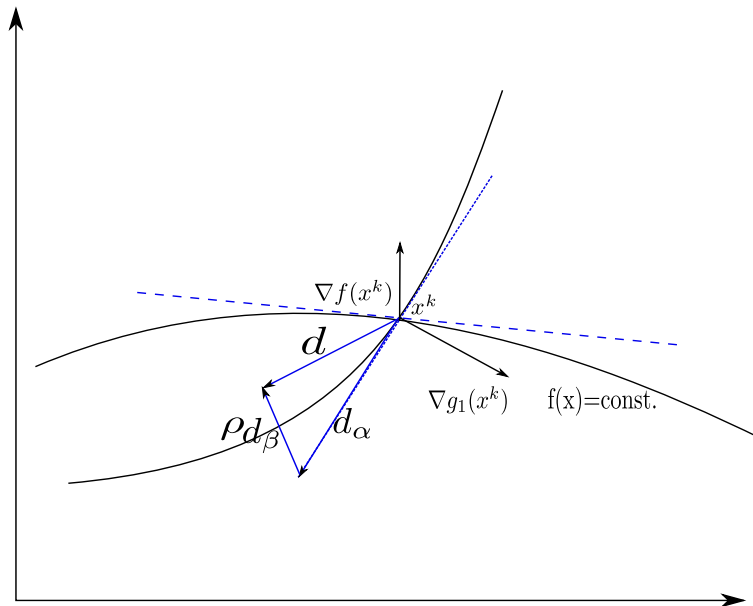


Figura 2.1: Direção de busca FDIPA

onde $\rho^k > 0$, d é a nova direção e $\bar{\lambda}$ é a nova estimativa de λ . Neste caso, (2.9) é equivalente a

$$\lambda_i^k \nabla g_i^T(x^k) d + g_i(x^k) \bar{\lambda}_i = -\rho^k \lambda_i^k, \quad i = 1, 2, \dots, m$$

e, consequentemente

$$\nabla g_i^T(x^k) d = -\rho^k < 0$$

para as restrições ativas. Logo d é uma direção viável.

A inclusão de um número negativo do lado direito da equação (2.7) produz uma deflexão

em d proporcional à ρ , na direção do interior da região viável. Para garantir que d seja uma direção de descida, ρ deve ser escolhido de maneira conveniente, veja [10].

Finalmente note que d pode ser obtido resolvendo-se os sistemas lineares,

$$\begin{aligned} B^k d_\alpha + \nabla g(x^k) \lambda_\alpha &= -\nabla f(x^k) \\ \Lambda^k \nabla g^T(x^k) d_\alpha + G(x^k) \lambda_\alpha &= 0 \end{aligned}$$

e

$$\begin{aligned} B^k d_\beta + \nabla g(x^k) \lambda_\beta &= 0 \\ \Lambda^k \nabla g^T(x^k) d_\beta + G(x^k) \lambda_\beta &= -\Lambda^k, \end{aligned}$$

de modo que

$$\begin{aligned} d &= d_\alpha + \rho d_\beta \\ \bar{\lambda} &= \lambda_\alpha + \rho \lambda_\beta \end{aligned}$$

2.3 ALGORITMO FDIPA

Apresentamos agora o algoritmo FDIPA.

Algoritmo 2.3.1. *Parâmetros $\xi \in \langle 0, 1 \rangle$, $\eta \in \langle 0, 1 \rangle$, $\varphi > 0$ e $\nu \in \langle 0, 1 \rangle$.*

Dados iniciais. $x \in \Omega_\alpha^0$, $0 < \lambda \in \mathbb{R}^m$, $B^k \in \mathbb{R}^{n \times n}$ matriz simétrica definida positiva.

Passo1. Cálculo da direção de busca.

(i) *Calcule $(d_\alpha, \lambda_\alpha)$ resolvendo o sistema linear*

$$B d_\alpha + \nabla g(x) \lambda_\alpha = -\nabla f(x), \quad (2.10)$$

$$\Lambda (\nabla g(x))^T d_\alpha + G(x) \lambda_\alpha = 0. \quad (2.11)$$

Se $d_\alpha = 0$, Parar.

(ii) *Calcule (d_β, λ_β) resolvendo o sistema linear*

$$B d_\beta + \nabla g(x) \lambda_\beta = 0, \quad (2.12)$$

$$\Lambda (\nabla g(x))^T d_\beta + G(x) \lambda_\beta = -\Lambda. \quad (2.13)$$

(iii) Se $d_\beta^T \nabla f(x) > 0$, tomar

$$\rho = \min[\varphi \|d_\alpha\|_2^2; (\xi - 1) \frac{d_\alpha^T \nabla f(x)}{d_\beta^T \nabla f(x)}]. \quad (2.14)$$

Caso contrário, tomar

$$\rho = \varphi \|d_\alpha\|_2^2. \quad (2.15)$$

(iv) Calcule a direção de busca

$$d = d_\alpha + \rho d_\beta \quad (2.16)$$

e também

$$\bar{\lambda} = \lambda_\alpha + \rho \lambda_\alpha \quad (2.17)$$

Passo 2. Busca linear

calcule t , o primeiro número da sequência $\{1, v, v^2, v^3, \dots\}$ satisfazendo

$$f(x + td) \leq f(x) + t\eta(\nabla f(x))^T d \quad (2.18)$$

e

$$g_i(x + td) < 0, \text{ se } \bar{\lambda}_i \geq 0 \quad (2.19)$$

ou

$$g_i(x + td) \leq g_i(x), \quad (2.20)$$

caso contrário.

Passo 3. Atualizar.

(i) Tomar

$$x = x + td$$

e defina novo

$$\lambda > 0$$

e

B

Este algoritmo é uma versão básica do FDIPA. Obtém-se uma família de algoritmos dependendo das regras na atualização da matriz B e do vetor λ . Algumas dessas regras encontram-se em [9] e [10].

2.4 CONVERGÊNCIA DO MÉTODO FDIPA

Nesta seção apresenta-se os lemas e teoremas da convergência do método FDIPA.

Lema 2.1. *Dado qualquer $x \in \Omega_a$, qualquer matriz definida positiva B e qualquer λ tal que $\lambda_i > 0$ onde $g_i(x) = 0$, a matriz*

$$M(x, \lambda, B) = \begin{bmatrix} B & \nabla g(x) \\ \Lambda \nabla g^T(x) & G(x) \end{bmatrix}$$

é não singular.

Lema 2.2. *O vetor d_α , definido no passo 1 do algoritmo, satisfaz*

$$d_\alpha^T \nabla f(x) \leq -d_\alpha^T B d_\alpha.$$

Lema 2.3. *A direção de busca d cumpre*

$$d^T \nabla f(x) \leq \xi d_\alpha^T \nabla f(x).$$

Proposição 2.1. *Seja $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ continuamente diferenciável e tal que $\nabla \phi$ satisfaz a condição Lipschitz em $\Gamma \in \mathbb{R}^n$, isto é, existe uma constante positiva k tal que para todo $x, y \in \Gamma$,*

$$\|\nabla \phi(y) - \nabla \phi(x)\|_2 \leq k \|y - x\|_2$$

Lema 2.4. *Existe $\tau > 0$ tal que para qualquer $x \in \Omega_a$ e d calculado no passo 1 do algoritmo, condições (2.18)-(2.20) são verificadas para qualquer $t \in [0, \tau]$.*

Lema 2.5. *Qualquer ponto de acumulação x^* da sequência $\{x^k\}$, gerada pelo algoritmo, é um ponto estacionário do problema. Além disso, $(x^*, \lambda_\alpha(x^*))$ constitui um par estacionário.*

Teorema 2.1. *Qualquer ponto de acumulação x^* de qualquer sequência gerada pelo algoritmo é um ponto Karush-Kuhn-Tucker do problema.*

A prova de convergência global, a inclusão de restrições de igualdade bem como teste numéricos podem ser encontrados em [10].

3 *MÉTODOS PARA OTIMIZAÇÃO CONVEXA NÃO DIFERENCIÁVEL*

No capítulo anterior estudamos um algoritmo para programação não linear diferenciável e pudemos ter uma breve ideia de como resolve-se um problema de otimização diferenciável. Dedicamos esse capítulo à otimização convexa não-diferenciável. Como veremos a seguir, problemas não-diferenciáveis precisam de técnicas computacionais próprias. Por exemplo, no caso diferenciável utiliza-se, para critério de parada, $\|\nabla f(x)\| < \varepsilon$ onde $\varepsilon > 0$ é previamente escolhido, o que não faz sentido no caso não-diferenciável. Outra dificuldade aparece na determinação da direção de busca, pois a direção de busca nem sempre é uma direção de descida e consequentemente a busca linear nem sempre é possível.

Neste capítulo faremos uma breve apresentação de alguns métodos para problemas não-diferenciáveis. Abordaremos os métodos Subgradientes, métodos de Planos Cortantes, os métodos de Feixe e o NFDA (Non-smooth Fesible Directions Algorithm), sendo esse último fundamentalmente importante para o algoritmo IED, tema central deste trabalho.

3.1 MÉTODO SUBGRADIENTES

Consideramos o seguinte problema irrestrito de otimização

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

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa não diferenciável.

Estes métodos, também chamados Métodos Gradientes Generalizados, veja [17], usam como direção de busca a direção dada por

$$d^k = \frac{-s^k}{\|s^k\|}, \text{ onde } s^k \in \partial f(x^k),$$

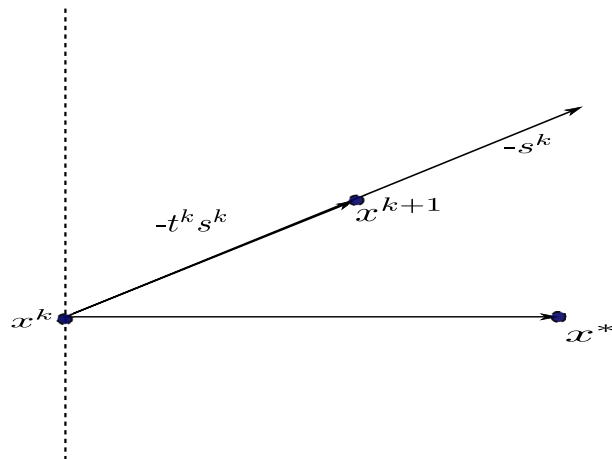


Figura 3.1: Passo do Método Subgradiente.

onde o próximo iterado fica então definido por $x^{k+1} = x^k + t_k d^k$ com $t_k > 0$ convenientemente escolhido de modo a garantir a convergência do método. Os métodos subgradientes possuem uma estrutura extremamente simples embora não apresentam bons resultados numéricos. Para maiores detalhes e aplicações destes métodos, vejam [7], [14] e [23].

Algoritmo 3.1.1 (Método Subgradiente).

Passo1. Escolha qualquer ponto $x^0 \in \mathbb{R}^n$ e tome $k = 0$

Passo2. Calcule $x^{k+1} = x^k - t^k \frac{s^k}{\|s^k\|}$, onde $s^k \in \partial f(x^k)$.

Passo3. Faça $k=k+1$ e volte ao passo 2.

3.1.1 Escolha do Tamanho do Passo

Apresentaremos agora uma condição sobre o passo que garante a convergência global dos Métodos Subgradientes, já que a escolha de um passo muito grande pode fazer o método oscilar desnecessariamente em torno da solução. Por outro lado se os passos forem muitos pequenos, o algoritmo se aproximará muito lentamente do ponto ótimo.

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

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

daí

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

segue-se

$$-(s^k)^T (x - x^k) \geq f(x^k) - f(x^*).$$

Como x^* é solução de (P) segue que $f(x^k) - f(x^*) \geq 0$. Portanto $-(s^k)^T(x - x^k) \geq 0$ consequentemente, o ângulo entre $-s^k$ e $x^* - x^k$ é agudo (como mostra-se na fig 3.1).

Então para t^k pequeno é fácil ver que

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

Assim, para $t^k > 0$ suficientemente pequeno, tem-se que x^{k+1} está mais próximo da solução x^* do que x^k , o que motiva a escolha a sequência $\{t^k\}$ tal que $\lim_{k \rightarrow \infty} t^k = 0$. O seguinte teorema apresenta uma cota para t^k (a prova pode ser encontrada em [16]).

Teorema 3.1. *Sejam x^* uma solução de (P) e $\{x^k\}$ uma sequência gerada pelo algoritmo subgradiente. Então*

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

sempre que

$$0 < t^k < 2 \frac{f(x^k) - f(x^*)}{\|s^k\|}.$$

No entanto, quando não conhecemos o valor ótimo do problema, não há como escolher um comprimento de passo que garanta esta propriedade de t^k e por isso devemos escolher t^k de tal maneira que a convergência fique garantida. Com efeito,

$$\|x^k - x^0\| \leq \|x^k - x^{k-1}\| + \|x^{k-1} - x^{k-2}\| + \dots + \|x^2 - x^1\| + \|x^1 - x^0\|$$

$$\text{como } x^{k+1} = x^k - t^k \frac{s^k}{\|s^k\|}$$

$$\text{tem-se } \|x^{k+1} - x^k\| = t^k$$

$$\text{então } \|x^k - x^0\| \leq t^{k-1} + t^{k-2} + \dots + t^0 = \sum_{j=0}^{k-1} t^k = r$$

$$\text{consequentemente } \|x^k - x^0\| \leq r$$

Assim, o método encontrará a solução x^* se, e somente se, $\forall k \in \mathbb{N}$, tem-se $x^* \in B(x^0, r)$, onde x^0 é o ponto inicial. Para contornar este problema, impõe-se que $\sum_{k=0}^{\infty} t^k = \infty$.

O próximo teorema cuja demonstração também encontra-se em [16] estabelece a convergência dos métodos subgradientes.

Teorema 3.2. *Suponha que o conjunto X^* das soluções de (P) seja não vazio e limitado. Então, para qualquer ponto inicial x^0 , a sequência $\{x^k\}_{k \in \mathbb{N}}$ definida por $x^{k+1} = x^k - t^k \frac{s^k}{\|s^k\|}$ com t^k satisfazendo $\lim_{k \rightarrow \infty} t^k = 0$ e $\sum_{k=0}^{\infty} t^k = +\infty$, é limitada e todos os pontos de acumulação pertencem a X^* .*

Uma condição mais forte sobre a convergência dos métodos subgradientes é (veja a de-

monstração em [2]).

Teorema 3.3. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Suponhamos que o problema (P) tenha solução, e que a sequência $\{t^k\}$ satisfaça as condições*

$$\sum_{k=0}^{\infty} t_k = +\infty, \quad \sum_{k=0}^{\infty} t_k^2 < +\infty$$

Então toda sequência $\{x^k\}$ gerada pelo algoritmo(3.1.1) converge a uma solução do problema (P).

Os métodos Subgradientes não possuem critério de parada, embora pelo Teorema 1.13, se x é um mínimo de f é necessário e suficiente que $0 \in \partial f(x)$, porém isto é difícil verificar na prática.

Uma condição necessária e suficiente para a escolha de direções de descida d para função f no ponto x é

$$d \in D_f(x) \iff \langle y, d \rangle < 0, \quad \forall y \in \partial f(x)$$

o que implica conhecer todos os subgradientes num ponto x o que é quase impossível verificar na prática. Outra complicação de usar subgradientes é que nem sempre a direção do anti-subgradiente é uma direção de descida.

Vejamos isto com um exemplo: seja $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x) = |x_1| + 2|x_2|$ e considere um ponto $x = (x_1, 0)$, onde $x_1 > 0$ é arbitrário. É fácil de verificar que

$$y = (1, 2) \in \partial f(x).$$

No entanto, para todo $t > 0$ suficientemente pequeno, $f(x - ty) = x_1 + 3t > f(x)$ o que mostra que

$$-y \notin D_f(x).$$

Isto mostra que um anti-subgradiente pode não ser uma direção de descida como mostra-se na figura 3.2. Para maiores detalhes destes métodos veja [14], [7],[23].

O seguinte método que apresentaremos, chamados Métodos de Planos Cortantes, são mais estáveis do que os Métodos Subgradientes.

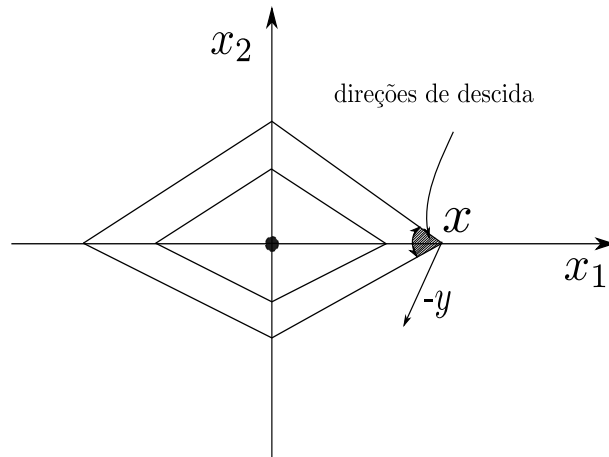


Figura 3.2: Exemplo de uma direção com sentido contrário a um subgradiente que não é uma direção de descida.

3.2 O MÉTODO DE PLANOS CORTANTES

Considere o problema

$$(C) \quad \begin{cases} \text{minimizar } f(x) \\ x \in D, \end{cases}$$

onde D é um conjunto convexo e compacto do \mathbb{R}^n e $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa.

A ideia principal do método de Planos Cortantes consiste na utilização das informações acumuladas ao longo das iterações anteriores para construir uma aproximação linear por partes (inferior) da função f , cada vez mais precisa. Consideremos que no início da iteração com índice $n+1$, tenhamos a seguinte informação: $x^k \in \mathbb{R}^n$, $s^k \in \partial f(x^k)$, $k = 1, 2, \dots, n$. A iteração do método de Planos Cortantes consiste na resolução do problema

$$(C_1) \quad \begin{cases} \text{minimizar } \psi_n(x) \\ x \in D, \end{cases}$$

onde $\psi_n: \mathbb{R}^n \rightarrow \mathbb{R}$, $\psi_n(x) = \max_{k=0,1,\dots,n} \{f(x^k) + (s^k)^T(x - x^k)\}$. Note que este problema é equivalente a

$$\begin{cases} \text{minimizar } t \\ (t, x) \in U_n, \end{cases}$$

onde $U_n = \{(t, x) \in \mathbb{R} \times D / f(x^k) + (s^k)^T(x - x^k) \leq t, k = 0, \dots, n\}$. Pela Proposição 1.1 temos que todas ψ_n são convexas e linear por partes. Além disso, pela definição de subdiferencial temos

$$f(x) \geq \psi_{n+1}(x) \geq \psi_n(x), \quad \forall x \in \mathbb{R}^n.$$

Na medida que n cresce, a aproximação ψ_n da função f fica cada vez mais precisa. Podemos

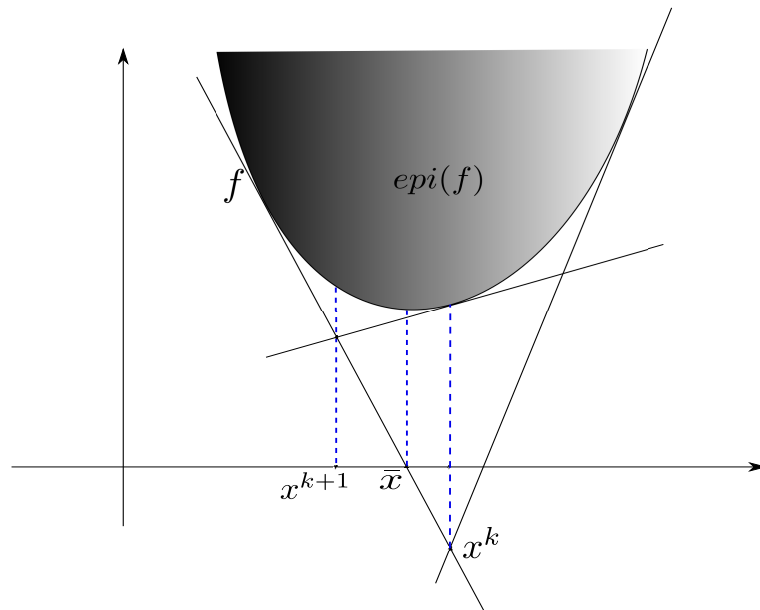


Figura 3.3: Método de Planos Cortantes

esperar, então, que as soluções dos subproblemas (C_1) se aproximam cada vez mais das soluções do problema (C) . Apresentemos o algoritmo de Planos Cortantes.

Algoritmo 3.2.1 (Algoritmo de Planos Cortantes). .

Passo 1. Escolha qualquer ponto $x^0 \in \mathbb{R}^n$, tomar $k = 0$

Passo 2. Calcular $f(x^k)$ e $s^k \in \partial f(x^k)$

Passo 3. Calcular x^{k+1} . uma solução do problema (C_1)

Passo 4. Fazer $k=k+1$, ir ao passo 2.

A vantagem mais importante deste método em comparação com os métodos subgradientes é a possibilidade de construir uma regra de parada, implementável através do cálculo do decréscimo nominal, definido por

$$\delta^k = f(x^k) - \psi_{k-1}(x^k).$$

Observa-se que o algoritmo termina quando δ^k é pequeno.

A principal deficiência do Métodos de Planos Cortantes é o acúmulo de subproblemas a serem resolvidos (métodos com muita memória). Assim como no método de subgradientes, no método de planos cortantes não se garante o decréscimo da função objetivo em cada iteração. Tal fato pode ser observado na figura 3.4. Note que no ponto x^3 foi encontrado $\delta^3 > \text{tol}$, onde tol é uma tolerância de parada, mas o processo continua obtendo $f(x^4) > f(x^3)$ e por isso diz-se que algoritmo não está livre de instabilidade.

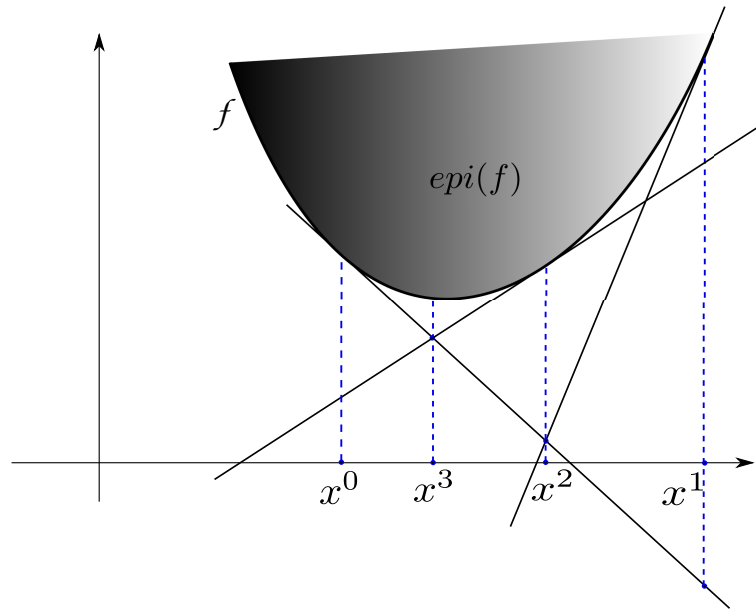


Figura 3.4: Falha do Método de Plano Cortantes

3.3 MÉTODO DE FEIXE

O método apresentado a seguir baseia-se na ideia de aproximações lineares por partes como no método de planos cortantes, porém com um mecanismo de estabilização, uma regra de parada confiável e uma técnica que permite total controle do tamanho dos subproblemas a serem resolvidos. Estes algoritmos se baseiam principalmente no material apresentado em [2]. Considere o problema de minimização irrestrita

$$(F) \quad \begin{cases} \text{minimizar } f(x) \\ x \in \mathbb{R}^n, \end{cases}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa em geral não diferenciável.

Suponha que $x^k \in \mathbb{R}^n$ seja uma aproximação atual da solução do problema (F) (tal que $f(x^k) < f(x^{k-1})$) e de modo que a iteração $k+1$ tenha acumulado os pontos $z^i \in \mathbb{R}^n$ e os subgradiantes $y^i \in \partial f(z^i)$, $i = 1, \dots, k$. O conjunto de dados é chamado o feixe. Então o iterado seguinte z^{k+1} é calculado achando uma solução do problema

$$\begin{cases} \text{minimizar } \psi_k(x) + \frac{\gamma_k}{2} \|x - x^k\|, \\ \text{sujeito a } x \in \mathbb{R}^n \end{cases} \quad (3.1)$$

onde

$$\psi_k : \mathbb{R}^n \rightarrow \mathbb{R}. \quad \psi_k(x) = \max_{i=0,1,\dots,k} \{f(x^i) + \langle y^i, x - z^i \rangle\} \quad (3.2)$$

é a aproximação de planos cortantes da função objetivo f , $\gamma_k > 0$ é um parâmetro estabilizador e o termo quadrático consiste em não permitir um deslocamento muito grande do ponto z^{k+1} em relação ao ponto x^k considerado como centro de estabilização.

A ideia dos métodos de feixe é mudar o centro de estabilização somente quando isso é realmente justificado, isto é, quando o passo de x^k a z^{k+1} resulta um decréscimo suficiente da função objetivo (isto é, que seja uma direção de descida). Se isso acontecer tomamos então $x^{k+1} = z^{k+1}$. Este tipo de passo se chama passo sério. Caso contrário, o centro de estabilização não muda, isto é, tomamos $x^{k+1} = x^k$. Este tipo de iteração se chama passo nulo.

3.3.1 Algoritmo do método de feixe com programação quadrática

Antes de implementar o algoritmo, acharemos uma forma de determinar a solução do problema (3.1), a qual é única, já que a função objetivo é fortemente convexa (veja o Teorema 1.7), e jogaremos fora parte da informação acumulada da equação (3.2), porém sem perder a convergência do proceso iterativo à solução do problema.

Seja

$$\psi_k(x) = f(x^k) + \max_{i=0,1,\dots,k} \{ \varepsilon_i^k + (y^i)^T (x - x^k) \} \quad (3.3)$$

onde

$$\varepsilon_i^k = f(x^k) - f(z^i) - (y^i)^T (x - z^i) \geq 0, \quad i = 0, 1, \dots, k. \quad (3.4)$$

Então, pela definição de ε -subdiferencial temos

$$y^i \in \partial_{\varepsilon_i^k} f(x^k), \quad i = 0, 1, \dots, k. \quad (3.5)$$

Agora, suponhamos que eliminamos alguns planos cortantes de nosso feixe. Então

$$\psi_k(x) = f(x^k) + \max_{i \in B_k} \{ \varepsilon_i^k + (y^i)^T (x - x^k) \} \quad (3.6)$$

onde B_k é o conjunto de índices que define o feixe atual, e tem-se que

$$y^i \in \partial_{\varepsilon_i^k} f(x^k), \quad i \in B_k \quad (3.7)$$

Para uma convergência adequada vamos a definir o que é uma função agregada. Para isso primeiro redefinimos (3.1) com restrição (3.6) ao problema equivalente

$$\text{minimizar } t + \frac{\gamma^k}{2} \|x - x^k\|^2 \quad (3.8)$$

$$(t, x) \in \{(t, x) \in \mathbb{R} \times \mathbb{R}^n / f(x^k) - \varepsilon_i^k + (y^i)^T (x - x^k) \leq t, i \in B_k\}. \quad (3.9)$$

Então em vez de resolver (3.8)-(3.9) podemos levar a seu dual, que pode ser escrito como

$$\text{minimizar } \frac{1}{2} \left\| \sum_{i \in B_k} v_i y^i \right\|^2 + \gamma_k \sum_{i \in B_k} v_i \varepsilon_i^k \quad (3.10)$$

$$v \in \{v \in \mathbb{R}_+^{|B_k|} / \sum_{i \in B_k} v_i = 1\}. \quad (3.11)$$

Temos o seguinte resultado.

Lema 3.1. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Para quaisquer $x^k \in \mathbb{R}^n$, $y^i \in \partial_{\varepsilon_i^k} f(x^k)$, $i \in B_k$, e $\gamma_k > 0$, seja*

$$d^k = \sum_{i \in B_k} v_i^k y_i^k \quad (3.12)$$

onde $v^k \in \mathbb{R}^{|B_k|}$ é uma solução de (3.10)-(3.11). Então, a única solução z^{k+1} do problema (3.1) onde a função objetivo é definida por (3.6), vem dado por

$$z^{k+1} = x^k - \frac{1}{\gamma_k} d^k \quad (3.13)$$

Além disso, tem-se que

$$d^k \in \partial_{\varepsilon_k} f(x^k) \quad (3.14)$$

onde

$$\varepsilon_k = \sum_{i \in B_k} v_i^k \varepsilon_i^k \quad (3.15)$$

Feito este lema agora podemos definir o que é a função agregada:

$$\psi_k^a : \mathbb{R}^n \rightarrow \mathbb{R}, \psi_k^a(x) = f(x^k) - \varepsilon_k (d^k)^T \cdot (x - x^k) \quad (3.16)$$

Então a técnica para remover planos cortantes sem alterar a convergência é agregando a função agregada ψ_k^a , veja os detalhes em [2].

Definindo agora as seguintes expressões para a implementação do método

$$e_{k+1}^{k+1} = f(x^k) - f(z^{k+1}) - (y^{k+1})^T \cdot (x^k - z^{k+1}) \geq 0 \quad (3.17)$$

e

$$\varepsilon_i^{k+1} = \varepsilon_i^k + f(x^{k+1}) - f(x^k) + (y^i)^T \cdot (x^k - x^{k+1}), \forall i \in B_{k+1}/\{k+1\}. \quad (3.18)$$

Feito os resultados e as definições prévias apresenta-se o algoritmo de feixe para programação quadrática.

Algoritmo 3.3.1 (O Método de Feixe). *Fixar um número inteiro $M \geq 2$ (o tamanho máximo permitido de feixe). Escolher $x^0 \in \mathbb{R}^n$ e tomar $k = 0$, $\mathcal{K} = \emptyset$. Escolher $\sigma \in (0,1)$. Calcular $f(x^0)$, e $y^0 \in \partial f(x^0)$, tomar $z^0 = x^0$, $\varepsilon_0^0 = 0$. Tomar $B_0 = \{0\}$, e definir o feixe inicial, que contém o único par (y^0, ε_0^0)*

Passo 1. *Escolher γ_k e calcular z^{k+1} , a solução do problema (3.1), com função objetivo dado por (3.6).*

Passo 2. *calcular $f(x^{k+1})$, $y^{k+1} \in \partial f(z^{k+1})$*

$$\Delta_k = f(x^k) - \psi_k(z^{k+1}) - \frac{\gamma_k}{2} \|z^{k+1} - x^k\|^2 \quad (3.19)$$

Passo 3. *Se $f(x^k) - f(z^{k+1}) \geq \sigma \Delta_k$*

*Tomar $x^{k+1} = z^{k+1}$ e incluir o índice k no conjunto \mathcal{K} (**passo sério**)*

*Caso contrário, tomar $x^{k+1} = x^k$ (**passo nulo**)*

Passo 4. *Se $|B_k| < M$, tomar $B_{k+1} = B_k \cup \{k+1\}$.*

Se $|B_k| = M$, escolher índices $i_1, i_2 \in B_k$ e tomar $B_{k+1} = (B_k/\{i_2\}) \cup \{k+1\}$,

$y^{i_1} = d^k$, $\varepsilon_{i_1}^k = \varepsilon_k$ onde $d \in \mathbb{R}^n$, e $\varepsilon_k \geq 0$ são definidos no lema anterior.

Passo 5. *Se $k \in \mathcal{K}$, calcular ε_i^{k+1} , $i \in B_{k+1}/\{k+1\}$, pela fórmula (3.18).*

Caso contrário, tomar $\varepsilon_i^{k+1} = \varepsilon_i^k$, $i \in B_{k+1}/\{k+1\}$. Calcular ε_{k+1}^{k+1} pela fórmula (3.17).

Definir o novo feixe como o conjunto de pares $(y^i, \varepsilon_i^{k+1})$, $i \in B_{k+1}$.

Passo 6. *Tomar $k = k+1$ e retornar ao **Passo 1**.*

4 NFDA - NONSMOOTH FEASIBLE DIRECTIONS ALGORITHM - PARA OTIMIZAÇÃO CONVEXA NÃO DIFERENCIÁVEL

Neste capítulo é apresentado um algoritmo para problemas convexos não diferenciáveis, chamado NFDA (Nonsmooth Feasible Directions Algorithm), desenvolvido por Freire [18] em sua tese de doutorado. NFDA foi desenvolvido a partir do FDIPA [10] para lidar com problemas irrestritos do tipo minimizar $f(x)$, onde $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa não necessariamente diferenciável. O problema original é substituído pelo problema equivalente minimizar $z \in \mathbb{R}$ tal que $f(x) \leq z$ que agora se torna um problema onde se quer minimizar uma função linear sujeita a uma restrição não diferenciável de desigualdade. Para resolver este último, constrói-se uma sequência de problemas auxiliares onde a restrição do problema equivalente é aproximada por hiperplanos suporte ao epígrafo da função f . Em cada iteração uma direção de busca para o problema equivalente é encontrada resolvendo-se dois sistemas lineares, análogos àqueles encontrados no FDIPA, obtidos agora dos problemas auxiliares. O algoritmo então constrói uma sequência de pontos interiores ao epígrafo de f , cujos pontos de acumulação são as soluções do problema original.

4.1 INTRODUÇÃO

NFDA foi desenvolvido para lidar com o seguinte problema:

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

onde $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa não necessariamente diferenciável.

Assume-se que os conjuntos de nível da função f são limitados. Assim, existe $a \in \mathbb{R}$ tal

que $\Gamma_a = \{x \in \mathbb{R}^n / f(x) \leq a\}$ é compacto. Além disso, pressupõe-se que dado $x \in \mathbb{R}^n$ é possível determinar um subgradiente $s \in \partial f(x)$.

4.1.1 O método NFDA

O problema irrestrito (P) pode ser reformulado como um problema de otimização linear com restrição de desigualdade não diferenciável, isto é,

$$(EP) \quad \begin{cases} \text{minimizar } z \\ f(x) \leq z \\ (x, z) \in \mathbb{R}^{n+1}. \end{cases}$$

Aqui $z \in \mathbb{R}$ é uma variável auxiliar. Com a presente aproximação, emprega-se a ideia de planos cortantes, para construir aproximações lineares do problema restrito (EP). Tomando $g_i^k(x, z)$ o conjunto do planos cortantes na iteração k , onde

$$g_i^k(x, z) = f(y_i^k) + (s_i^k)^T (x - y_i^k) - z, \quad i = 0, 1, \dots, l,$$

$y_i^k \in \mathbb{R}^n$ são pontos auxiliares, $s_i^k \in \partial f(y_i^k)$ e l representa o número de planos cortantes. Chamando agora

$$\tilde{g}_l^k(x, z) = [g_0^k(x, z), \dots, g_l^k(x, z)]^T, \quad \tilde{g}_l^k : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^{l+1}$$

o problema na iteração e fica definido por

$$(AP_l) \quad \begin{cases} \text{minimizar } \psi(x, z) = z \\ \tilde{g}_l^k(x, z) \leq 0 \\ (x, z) \in \mathbb{R}^{n+1}. \end{cases}$$

O algoritmo NFDA não resolve o problema (AP_l). O método procura uma direção de busca d_l^k , como no método FDIPA, usando o problema (AP_l). Note que d_l^k pode ser achado ainda se o problema (AP_l) não tem solução mínima.

O maior comprimento de passo é dado por

$$t = \max\{t / \tilde{g}_l^k((x^k, z^k) + t d_l^k) \leq 0\}.$$

Caso t seja infinito, toma-se

$$t_l^k = \min\{t_{max}/\mu, t\},$$

onde $\mu \in (0, 1)$ e $t_{max} > 0$ é prefixado. O ponto

$$(x_{l+1}^k, z_{l+1}^k) = (x^k, z^k) + t_l^k d_l^k \quad (4.1)$$

é viável com respeito ao problema (AP_l) . Então um ponto auxiliar é calculado

$$(y_{l+1}^k, w_{l+1}^k) = (x^k, z^k) + \mu t_l^k d_l^k. \quad (4.2)$$

Se (y_{l+1}^k, w_{l+1}^k) é estritamente viável como respeito ao problema (EP) , isto é, se $w_{l+1}^k > f(y_{l+1}^k)$ considera-se que o conjunto atual de planos cortantes é uma boa aproximação de $f(x)$ numa vizinhança de x^k . Diz-se então que o passo é **sério** e uma nova iteração é definida $(x^{k+1}, z^{k+1}) = (y_{l+1}^k, w_{l+1}^k)$. Caso contrário, um novo plano cortante $g_{l+1}^k(x, z)$ é agregado ao problema aproximado e repete-se o procedimento até que um passo sério seja obtido. Feita esta descrição do método, apresentaremos o algoritmo NFDA.

4.2 ALGORITMO NFDA

Parâmetros. $\xi, \mu \in (0, 1)$, $\varphi > 0$ e $t_{max} > 0$.

Dado. $x^0, a > z^0 > f(x^0)$, $\lambda_0^0 \in \mathbb{R}_+$, $B \in \mathbb{R}^{(n+1) \times (n+1)}$ simétrica e definida positiva.

Tomar $y_0^0 = x^0$, $k = 0$ e $l = 0$.

Passo 1. Calcule $s_l^k \in \partial f(y_l^k)$. Um novo plano cortante na iteração atual

(x^k, z^k) é definido por

$$g_l^k(x, z) = f(y_l^k) + (s_l^k)^T (x - y_l^k) - z$$

Considere agora

$$\nabla g_l^k(x, z) = \begin{bmatrix} s_l^k \\ -1 \end{bmatrix} \in \mathbb{R}^{n+1}$$

defina

$$\tilde{g}_l^k(x, z) = [g_0^k(x, z), \dots, g_l^k(x, z)]^t \text{ in } \mathbb{R}^{l+1}$$

e

$$\nabla \tilde{g}_l^k(x, z) = [\nabla g_0^k(x, z), \dots, \nabla g_l^k(x, z)]^t \in \mathbb{R}^{(n+1) \times (l+1)}$$

Passo 2. Direção viável e de descida d_l^k para o problema (AP_l)

(ia) Calcule $d_{\alpha l}^k$ e $\lambda_{\alpha l}^k$, resolvendo

$$B^k d_{\alpha l}^k + \nabla \tilde{g}_l^k(x^k, z^k) \tilde{\lambda}_{\alpha l}^k = -\nabla \psi(x, z) \quad (4.3)$$

$$\tilde{\Lambda}_l^k [\nabla \tilde{g}_l^k(x^k, z^k)]^t d_{\alpha l}^k + \tilde{G}_l^k(x^k, z^k) \tilde{\lambda}_{\alpha l}^k = 0 \quad (4.4)$$

(ib) Calcule $d_{\beta l}^k$ e $\lambda_{\beta l}^k$, resolvendo

$$B^k d_{\beta l}^k + \nabla \tilde{g}_l^k(x^k, z^k) \tilde{\lambda}_{\beta l}^k = 0 \quad (4.5)$$

$$\tilde{\Lambda}_l^k [\nabla \tilde{g}_l^k(x^k, z^k)]^t d_{\beta l}^k + \tilde{G}_l^k(x^k, z^k) \tilde{\lambda}_{\beta l}^k = -\tilde{\lambda}_l^k \quad (4.6)$$

onde

$$\tilde{\lambda}_{\alpha l}^k = (\lambda_{\alpha 0}^k, \dots, \lambda_{\alpha l}^k), \quad \lambda_{\beta}^k = (\lambda_{\beta 0}^k, \dots, \lambda_{\beta l}^k)$$

$$\tilde{\lambda}_l^k = (\tilde{\lambda}_0^k, \dots, \tilde{\lambda}_l^k), \quad \tilde{\Lambda}_l^k = \mathbf{diag}(\lambda_0^k, \dots, \lambda_l^k)$$

$$\text{e } \tilde{G}_l^k(x, z) = \mathbf{diag}(g_0^k(x, z), \dots, g_l^k(x, z)).$$

(ii) Se $(d_{\beta l}^k)^T \nabla \psi(x, z) > 0$,

$$\rho = \varphi \|d_{\alpha l}^k\|^2$$

Caso contrário,

$$\rho = \min \left\{ \varphi \|d_{\alpha l}^k\|^2, (\xi - 1) \frac{(d_{\alpha l}^k)^t \nabla \psi(x, z)}{(d_{\beta l}^k)^t \nabla \psi(x, z)} \right\}$$

(iii) Calcule a direção de descida viável

$$d_l^k = d_{\alpha l}^k + \rho d_{\beta l}^k.$$

Passo 3. Calcule o comprimento do passo

$$t_l^k = \min \left\{ t_{max}/\mu, \min \{ t / \tilde{g}_l^k((x^k, z^k) + t d_l^k) \leq 0 \} \right\} \quad (4.7)$$

Passo 4. Calcule um novo ponto

i) Seja $(y_{l+1}^k, w_{l+1}^k) = (x^k, z^k) + \mu t_l^k d_l^k$,

ii) Se $w_{l+1}^k \leq f(y_{l+1}^k)$ temos um passo nulo. Então, definimos $\lambda_{l+1}^k > 0$ e tomamos $l = l + 1$.

Caso contrário, temos um passo sério. Então, chamamos $d^k = d_l^k, d_\alpha^k, = d_{\alpha l}^k$,

$$d_\beta^k = d_{\beta l}^k, \lambda_\alpha^k = \lambda_{\alpha l}^k, \lambda_\beta^k = \lambda_{\beta l}^k \text{ e } l^k = l, \text{ tomamos}$$

$$(x^{k+1}, z^{k+1}) = (y_{l+1}^k, w_{l+1}^k),$$

definimos $\lambda_0^{k+1} > 0, B^{k+1}$ simétrica e definida positiva e tomamos $k = k + 1, l = 0$,

$$y_0^k = x^k$$

iii) ir ao passo 1)

Os novos valores de λ e B deve satisfazer as seguintes hipóteses:

Hipótese 4.3 Existem números positivos σ_1 e σ_2 tais que

$$\sigma_1 \|v\|^2 \leq v^T B v \leq \sigma_2 \|v\|^2$$

para qualquer $v \in \mathbb{R}^{n+1}$.

Hipótese 4.4 Existem números positivos λ^l, λ^s , tais que $\lambda^l \leq \lambda_i \leq \lambda^s$ para $i = 0, 1, \dots, l$.

4.3 CONVERGÊNCIA DO ALGORITMO NFDA

Nesta seção, apresenta-se os resultados de convergência global do algoritmo. As demonstrações podem ser encontradas em [18]. A prova de convergência segue o seguinte roteiro: primeiro mostra-se que a direção de busca d_l^k é uma direção de descida para ψ . Mostra-se que o número de passos nulos em cada iteração é finito. Isto é, como $(x^k, z^k) \in \text{int}(\text{epi}(f))$, depois de um número finito de sub-iterações, obtem-se $(x^{k+1}, z^{k+1}) \in \text{int}(\text{epi}(f))$. Em consequência, a sequência $\{(x^k, z^k)\}_{k \in \mathbb{N}}$ é limitada e pertence ao interior do epígrafo de f . No que segue, mostra-se que d^k converge a zero quando $k \rightarrow \infty$. Este fato é usado para estabelecer um critério de parada para o algoritmo. Finalmente, mostra-se que a condição de otimalidade $0 \in \partial f(x^*)$ é satisfeita para os pontos de acumulação (x^*, z^*) da sequência $\{(x^k, z^k)\}_{k \in \mathbb{N}}$. Assim, qualquer ponto de acumulação da sequência $\{(x^k, z^k)\}_{k \in \mathbb{N}}$ é uma solução do problema (P).

Seguem abaixo os resultados.

Lema 4.1. *Suponha que os vetores no conjunto $\{\nabla g_i(x^k, z^k) / g_i(x^k, z^k) = 0\}$ são linearmente independente. Então para qualquer vetor $(x, z) \in \text{int}(\text{epi} f)$ e qualquer matriz definida positiva $B \in \mathbb{R}^{(n+1) \times (n+1)}$, a matriz*

$$\begin{bmatrix} B & \nabla \tilde{g}(x, z) \\ \tilde{\Lambda}[\nabla \tilde{g}(x, z)]^T & \tilde{G}(x, z) \end{bmatrix}$$

é não singular

Segue-se que d_α , d_β , λ_α e λ_β são limitados em Γ_a . Como ρ é limitado superiormente, temos também que $\bar{\lambda} = \lambda_\alpha + \rho\lambda_\beta$ é limitado.

Lema 4.2. *O vetor d_α satisfaz*

$$d_\alpha^T \nabla \psi(x, z) \leq -d_\alpha^T B d_\alpha$$

Como consequência, tem-se que a direção de busca d_α é descida para a função objetivo do problema AP_l .

Proposição 4.1. *A direção $d = d_\alpha + \rho d_\beta$ é uma direção de descida para a função objetivo do problema (AP_l) em um ponto (x, z) .*

Como consequência da proposição prévia, sendo $z^{k+1} = z^k + \mu t^k d_z^k$, tem-se que $z^{k+1} < z^k$ para todo k . Assim a sequência $\{(x^k, z^k)\}_{k \in \mathbb{N}}$ gerada pelo presente algoritmo pertence ao conjunto limitado $\text{int}(\text{epi} f) \cap \{(x, z) \in \mathbb{R}^{n+1} / z < z^0\}$.

Lema 4.3. *Seja $X \subseteq \mathbb{R}^n$ um conjunto convexo. Considere $x_0 \in \text{int}(X)$ e $\tilde{x} \in X$. Seja $\{\tilde{x}^k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n - X$ uma sequência tal que $\tilde{x}^k \rightarrow \tilde{x}$. Seja $\{x^k\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n$ uma sequência definida por $x^k = x_0 + \mu(\tilde{x}^k - x_0)$ com $\mu \in (0, 1)$. Então $\exists k_0 \in \mathbb{N}$ tal que $x^k \in \text{int}(X)$, $\forall k > k_0$.*

Proposição 4.2. *Seja $(\tilde{x}^k, \tilde{z}^k)$ um ponto de acumulação da sequência $\{(x_l^k, z_l^k)\}_{l \in \mathbb{N}}$ definida na eq.(4.1) para k fixo, então $\tilde{z}^k = f(\tilde{x}^k)$.*

Proposição 4.3. *Seja $(x^k, z^k) \in \text{int}(\text{epi} f)$. O iterado seguinte $(x^{k+1}, z^{k+1}) \in \text{int}(\text{epi} f)$ é obtido depois de um número finito de sub-iterações.*

Lema 4.4. *Para qualquer $(x, z) \in \text{int}(\text{epi}(f))$ e qualquer direção d dada pelo algoritmo, existe $\tau > 0$ tal que $\tilde{g}_l((x, z) + td) \leq 0$, $\forall t \in [0, \tau]$.*

Proposição 4.4. *Seja d_α^* um ponto de acumulação da sequência $\{d_\alpha^k\}_{k \in \mathbb{N}}$. Então $d_\alpha^* = 0$.*

Proposição 4.5. *Para qualquer ponto de acumulação (x^*, z^*) da sequência $\{(x^k, z^k)\}_{k \in \mathbb{N}}$, temos $0 \in \partial f(x^*)$.*

Para resultados numéricos e ter uma melhor ideia do método, veja [18].

5 *DSG - THE DEFLECT SUBGRADIENT METHOD - UM ALGORITMO TIPO SUBGRADIENTE PARA OTIMIZAÇÃO NÃO-CONVEXA E NÃO-DIFERENCIÁVEL*

Neste capítulo apresenta-se um método para solução de problemas de otimização não-convexos e não-diferenciáveis com restrições. Este método, chamado DSG (Deflected Subgradient Method) foi desenvolvido por Burachik e outros [6]. O método usa dualidade lagrangeana cuja função lagrangeana aumentada induz um método primal-dual com a propriedade da dualidade forte ou seja com brecha de dualidade zero. O método converge para a solução do problema dual se e somente se tal solução existe. É mostrado que todos pontos de acumulação da sequência primal gerada pelo algoritmo são soluções do problema primal. Veremos que, dependendo da escolha da função lagrangeana, o método se reduz ao método clássico de penalidades.

5.1 O MÉTODO DSG

Considere o problema de programação não linear:

$$(P) \quad \begin{cases} \text{minimizar } f_0(x) \\ f(x) = 0 \\ x \in X. \end{cases}$$

Define-se a função lagrangeana associada ao problema (P) como segue:

$$L(x, (u, c)) = f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle, \quad (5.1)$$

onde σ , A e L são como na definição (1.25).

Então, define-se a função Dual $H : \mathbb{R}^m \times \mathbb{R}_+ \rightarrow \mathbb{R}$, por

$$H(u, c) = \min_{x \in X} \{f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle\} \quad (5.2)$$

e então o problema dual fica

$$(D) \quad \begin{cases} \max H(u, c) \\ (u, c) \in \mathbb{R}^m \times \mathbb{R}_+. \end{cases}$$

Denota-se o conjunto das soluções viáveis do problema (P) como $S(P)$ e um elemento de $S(P)$ será denotado por \bar{x} . O conjunto das soluções viáveis do problema (D) é denotado por $S(D)$ e um elemento de $S(D)$ será denotado por $\bar{z} = (\bar{u}, \bar{c})$. Para facilitar a análise define-se o conjunto

$$X(\bar{u}, \bar{c}) = \text{Argmin}_{x \in X} \{f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle\} \quad (5.3)$$

5.2 O ALGORITMO DSG

Descreveremos agora o algoritmo correspondente ao método DSG:

5.2.1 Algoritmo DSG

Passo 0 Escolha (u_0, c_0) com $c_0 \geq 0$. Seja $k = 1$.

Passo k Dado (u_k, c_k) :

Passo k.1 encontrar o vetor $x_k \in X(u_k, c_k)$.

Se $f(x_k) = 0$, parar. Caso contrário,

Passo k.2 Seja

$$\begin{aligned} u_{k+1} &= u_k - s_k A(f(x_k)) \\ c_{k+1} &= c_k + (s_k + \varepsilon_k) \sigma(f(x_k)) \end{aligned}$$

onde $s_k, \varepsilon_k > 0$. Seja $k = k + 1$ e ir ao passo k.

Obs: note que o critério de parada do algoritmo se baseia no Teorema 1.17 e portanto vale destacar sua eficiência.

A seguinte hipótese sobre σ e A é importante.

Hipótese (L1) $\sigma(z) \geq \|A(z)\|$ para qualquer $z \in \mathbb{R}^m$.

Com esta hipótese, mostra-se na seguinte proposição que o algoritmo gera uma sequência $\{H(c_k, u_k)\}$ estritamente crescente. Denota-se $H_k = H(c_k, u_k)$ e seja \bar{H} valor ótimo do problema dual.

Proposição 5.1. *Considere a notação e definição do algoritmo DSG. Então*

$$a) H_{k+1} - H_k \leq s_k(\|Af(x_k)\|^2 + \sigma(f(x_k))^2) + \varepsilon_k \sigma(f(x_k))^2.$$

b) *Valem as afirmações:*

(i) *Se a função aumentada σ verifica (L1), então o passo DSG produz uma sequência $\{H_k\}$, que é estritamente crescente. Mais precisamente, se $(u_k, c_k) \notin S(D)$ então $H_{k+1} > H_k$.*

(ii) *Assuma que $\sigma(-z) = \sigma(z)$ para qualquer $z \in \mathbb{R}^m$. Se, para qualquer problema do tipo (P), o passo DSG produz uma sequência $\{H_k\}$ estritamente crescente, então σ deve verificar (L1).*

Demonstração. Aplicando o Lema 1.4(a), obtemos

$$H_{k+1} - H_k \leq \langle u_{k+1} - u_k, -Af(x_k) \rangle + (c_{k+1} - c_k) \sigma(f(x_k)).$$

Usando também a definição do algoritmo, o lado direito da expressão acima pode ser reescrito como

$$H_{k+1} - H_k \leq s_k \|Af(x_k)\|^2 + (s_k + \varepsilon_k) \sigma(f(x_k))^2,$$

o que implica facilmente a conclusão. Isto prova o item (a). Vamos agora provar (i) no item (b). Note que pelo Teorema 1.17 temos que $(u_k, c_k) \notin S(D)$ implica que $f(x_k) \neq 0$. Portanto $\sigma(f(x_k)) > 0$. Usando a definição do algoritmo escrevemos:

$$H_{k+1} = H(u_k, c_k + \varepsilon_k \sigma(f(x_k))).$$

Seja $\hat{x}_k \in X(u_k, c_k + \varepsilon_k \sigma(f(x_k)))$. Assuma primeiro que $f(\hat{x}_k) = 0$. Neste caso, o Teorema 1.17 nos dá $(u_k, c_k + \varepsilon_k \sigma(f(x_k))) \in S(D)$. Por outro lado, como $(u_k, c_k) \notin S(D)$ devemos ter

$H(u_k, c_k) < H(u_k, c_k + \varepsilon_k \sigma(f(x_k))) \leq H_{k+1}$. Portanto a conclusão mantêm-se nestes casos. Assuma agora que $f(\hat{x}_k) \neq 0$. Então $\sigma(f(\hat{x}_k)) > 0$ e

$$H_{k+1} \geq H_k + \varepsilon_k \sigma(f(x_k)) \sigma(f(\hat{x}_k)).$$

Portanto,

$$H_{k+1} \geq H_k + \varepsilon_k \sigma(f(x_k)) \sigma(f(\hat{x}_k)) > H_k. \quad (5.4)$$

Mostra-se agora que, se (L1) não vale, então pode-se encontrar um problema (P) onde a sequência $\{H_k\}$ gerada por DSG não seja estritamente crescente. Se (L1) não vale, então existe $\bar{z} \in \mathbb{R}^m$ tal que $\sigma(\bar{z}) - \|A\bar{z}\| < 0$. Portanto $\bar{z} \neq 0$. Seja $r = \|\bar{z}\| > 0$. Considere o problema (P) com $f_0(x) = \sigma(x) - \|Ax\|$, $X = B(0, r)$ a bola fechada de centro zero e raio r e a função restrição $f(x) = x$. É claro que $M_p = 0$. Seja $\bar{x} \in \text{Argmin}_{x \in B(0, r)} \{\sigma(x) - \|Ax\|\}$. Pela Hipótese (L1) temos que $f_0(\bar{x}) \leq f_0(\bar{z}) < 0$. Escolhendo agora $(u_0, c_0) = (0, 0)$, toma-se:

$$x_0 = \bar{x} \in \text{Argmin}_{x \in B(0, r)} \sigma(x) - \|Ax\| = X(u_0, c_0),$$

e $H_0 = H(u_0, c_0) = f_0(x_0) < 0 = M_p$. Como $f(x_0) = x_0 \neq 0$ realiza-se um passo DSG. Tomando s_0, ε_0 tal que

$$\frac{s_0}{\varepsilon_0} > \frac{\sigma(x_0)^2}{\|Ax_0\|^2 - \sigma(x_0)^2} > 0,$$

do passo DSG, temos que $(u_1, c_1) = (-s_0 Ax_0, (s_0 + \varepsilon_0) \sigma(x_0))$. Assim,

$$\begin{aligned} H_1 &= \min_{x \in B(0, r)} \{ \sigma(x) - \|Ax\| + (s_0 + \varepsilon_0) \sigma(x_0) \sigma(x) + s_0 \langle Ax_0, Ax \rangle \} \\ &\leq \sigma(-x_0) - \|A(-x_0)\| + s_0 (\sigma(-x_0)^2 - \|Ax_0\|^2) + \varepsilon_0 \sigma(-x_0) \sigma(x_0) \\ &= \sigma(x_0) - \|Ax_0\| + s_0 (\sigma(x_0)^2 - \|Ax_0\|^2) + \varepsilon_0 \sigma(x_0)^2 \\ &< \sigma(x_0) - \|Ax_0\| = H_0, \end{aligned}$$

O que é uma contradição. Logo, devemos ter que (L1) é verdadeira. \square

Corolário 5.2.1. *Considere a notação e a definição do algoritmo DSG e assuma que (L1) valha. Então, para qualquer k temos*

$$\varepsilon_k \sigma(f(x_k)) \sigma(f(\hat{x})) \leq H_{k+1} - H_k \leq s_k \|Af(x_k)\|^2 + (s_k + \varepsilon_k) \sigma(f(x_k))^2$$

onde $\hat{x} \in X(u_k, c_k + \varepsilon_k \sigma(f(x_k)))$.

Apresentaremos agora uma condição necessária e suficiente para o comprimento de passo s_k e ε_k que garantem a limitação da sequência dual $z_k = (u_k, c_k)$.

Lema 5.1. *Considere a notação e definição do algoritmo DSG. Os seguintes resultados são equivalentes:*

1. $\sum_{k=0}^{\infty} (s_k + \varepsilon_k) \sigma(f(x_k)) < \infty$.
2. A sequência $\{z_k\}$ é limitada.

Demonstração. Do algoritmo DSG

$$c_{m+1} - c_0 = \sum_{k=0}^m (s_k + \varepsilon_k) \sigma(f(x_k)) \quad \text{e} \quad \|u_{m+1} - u_0\| \leq \sum_{k=0}^m s_k \|Af(x_k)\|. \quad (5.5)$$

Por (L1) temos $\|Af(x_k)\| \leq \sigma(f(x_k))$. Se (a) vale, então claramente a sequência $\{c_k\}$ é limitada. Por outro lado,

$$\|u_{m+1} - u_0\| \leq \sum_{k=0}^m s_k \|Af(x_k)\| \leq \sum_{k=0}^m s_k \sigma(f(x_k)) < +\infty,$$

o qual limita à sequência $\{u_k\}$. Assuma agora que (b) valha. Por (5.5) devemos ter a condição (a).

□

5.3 CONVERGÊNCIA DO MÉTODO DSG

Lema 5.2. *Seja \bar{H} o valor ótimo de (P) (i.e $\bar{H} = M_D = M_P$). Assuma que a sequência $\{z_k\}$ gerada pelo algoritmo DSG é limitada e que o comprimento de passo s_k satisfaz*

$$s_k \geq \eta \frac{(\bar{H} - H_k)}{\sigma(f(x_k))^2}, \quad (5.6)$$

para algum $\eta > 0$ fixo. Então $\{H_k\}$ converge a \bar{H} e cada ponto de acumulação de $\{z_k\}$ é uma solução dual. Em particular, $S(D) \neq \emptyset$.

Demonstração. Pela eq.(5.5), a limitação da sequência $\{z_k\}$ implica que

$$\sum_{k=0}^{\infty} s_k \sigma(f(x_k)) < +\infty. \quad (5.7)$$

Seja (\bar{u}, \bar{c}) um ponto de acumulação de $\{(u_k, c_k)\}$, e denote por \mathcal{K} o conjunto infinito de índice tal que

$$\begin{aligned} \lim_{\substack{k \in \mathcal{K}, \\ k \rightarrow \infty}} (u_k, c_k) &= (\bar{u}, \bar{c}). \end{aligned}$$

Mostraremos que $(\bar{u}, \bar{c}) \in S(D)$. Pela limitação de $\{x_k\}$, pode-se também assumir que a sequência completa $\{x_k\}_{k \in \mathcal{K}}$ converge a algum ponto \bar{x} . Se $f(\bar{x}) = 0$, afirmamos que $\bar{x} \in X(\bar{u}, \bar{c})$. Neste caso, o Teorema 1.17 implica que $(\bar{u}, \bar{c}) \in S(D)$. De fato, pela definição de x_k , temos que

$$f_0(x_k) + c_k \sigma(f(x_k)) - \langle Au_k, f(x_k) \rangle \leq f_0(x) + c_k \sigma(f(x)) - \langle Au_k, f(x) \rangle,$$

para todo $x \in X$ e para todo k . Tomando limites para $k \in \mathcal{K}$, $k \rightarrow \infty$ na expressão acima, obtemos

$$f_0(\bar{x}) + \bar{c} \sigma(f(x_k)) - \langle A\bar{u}, f(\bar{x}) \rangle \leq f_0(x) + \bar{c} \sigma(f(x)) - \langle A\bar{u}, f(x) \rangle,$$

para todo $x \in X$. Portanto $\bar{x} \in X(\bar{u}, \bar{c})$ e assim $(\bar{u}, \bar{c}) \in S(D)$. Assuma agora que $f(\bar{x}) \neq 0$. Este fato, junto com a eq.(5.7), implica que a sequência $\{s_k\}_{k \in \mathcal{K}}$ converge a zero. Usando também a eq.(5.6) para $k \in \mathcal{K}$, concluí-se que a subsequência dos valores duais $\{H_k\}_{k \in \mathcal{K}}$ converge a \bar{H} . Por continuidade de H temos que

$$\begin{aligned} H(\bar{u}, \bar{c}) &= \limsup_{\substack{k \in \mathcal{K} \\ k \rightarrow \infty}} H(u_k, c_k) = \bar{H}. \end{aligned}$$

Isto mostra que $H(\bar{u}, \bar{c})$ tem valor ótimo \bar{H} e portanto $(\bar{u}, \bar{c}) \in S(D)$. Além disso, do Corolário 5.2.1 sabemos que $\{H_k\}$ é estritamente crescente. Com isto tem-se uma subsequência convergente e então a sequência completa deve convergir para \bar{H} . Portanto, isto completa a prova. □

Para o seguinte lema definamos $d_k = \|z - z_k\|$.

Lema 5.3. *Fixe $z = (u, c) \in \mathbb{R}^m \times \mathbb{R}_+$. Então*

$$d_{k+1} - d_k \leq (s_k^2 \|Af(x_k)\|^2 + (s_k + \varepsilon_k)^2 \sigma(f(x_k))^2) + 2[s_k(H_k - H(u, c)) - \varepsilon_k \sigma(f(x_k))(c - c_k)].$$

Teorema 5.1. *Se a sequência $\{z_k\}$ gerada pelo algoritmo DSG é limitada, então ela é convergente.*

Demonstração. Assuma que a sequência $\{z_k\}$ é limitada e seja \hat{z} um ponto de acumulação de $\{z_k\}$. Seja $\{z_{k_j}\}_j$ uma sequência convergente a \hat{z} . Usando o Lema 5.3 para a escolha de

$z = \hat{z} = (\hat{u}, \hat{c})$, concluí-se que a sequência $\{d(\hat{z}, z_k)\}$ verifica

$$d(\hat{z}, z_{k+1}) - d(\hat{z}, z_k) \leq s_k^2 \|Af(x_k)\|^2 + (s_k + \varepsilon_k)^2 \sigma(f(x_k))^2 - 2s_k(H(\hat{z}) - H_k) - 2\varepsilon_k \sigma(f(x_k))(\hat{c} - c_k).$$

Pelo Corolário 5.2.1, $\{H_k\}$ é estritamente crescente e portanto $\lim_k H_k = \sup_k H_k = \lim_j H_{k_j} = H(\hat{z})$. Usando agora semicontinuidade superior de H , obtemos

$$H(\hat{z}) \geq \limsup_j H_{k_j} = \lim_j H_{k_j} = \sup_j H_{k_j} \geq H_k \text{ para qualquer } k \in \mathbb{N}.$$

Assim, $(H(\hat{z}) - H_k) \geq 0$ para todo k . Usando também que $\{c_k\}$ é uma sequência estritamente crescente, temos que $(\hat{c} - c_k) \geq 0$ para todo k . Portanto,

$$d(\hat{z}, z_{k+1}) - d(\hat{z}, z_k) \leq s_k^2 \|Af(x_k)\|^2 + (s_k + \varepsilon_k)^2 \sigma(f(x_k))^2.$$

Desde que $\{z_k\}$ é limitada, podemos usar o Lema 5.1 para concluir que a série com termo geral $a_k = s_k^2 \|A(f(x_k))\|^2 + (s_k + \varepsilon_k)^2 \sigma(f(x_k))^2$ é convergente, e isto implica que a sequência $\{d(\hat{z}, z_k)\}_k$ é convergente. Mas a subsequência $\{d(\hat{z}, z_{k_j})\}_j$ desta sequência converge a zero, e assim toda a sequência converge a zero, dando a unicidade do ponto de acumulação.

□

O seguinte teorema mostra que toda sequência $\{(u_k, c_k)\}$ gerada pelo algoritmo DSG converge a uma solução dual e que todo ponto de acumulação da sequência $\{x_k\}$ converge a uma solução primal.

Teorema 5.2 (Convergência Primal-Dual). *Assuma que a sequência $\{(u_k, c_k)\}$ gerada pelo algoritmo é limitada. Assuma também que para algum $\eta > 0$ o comprimento de passo s_k satisfaz*

$$s_k \geq \eta \frac{(\bar{H} - H_k)}{\sigma(f(x_k))^2}. \quad (5.8)$$

Então a sequência $\{z_k\}$ converge a uma solução dual. Adicionalmente, todo ponto de acumulação de $\{\tilde{x}_k\}$ é solução de (P).

Para demonstração ver [6].

5.4 Uma escolha especial de s_k

Nesta seção estudaremos uma escolha especial para o parâmetro s_k , para o qual $S(D)$ não-vazio é equivalente à limitação de $\{z_k\}$. O comprimento de passo que consideramos é como segue:

$$\eta \frac{\bar{H} - H_k}{\sigma(f(x_k))^2} \leq s_k \leq 2 \frac{\bar{H} - H_k}{\sigma(f(x_k))^2}, \quad (5.9)$$

com $\eta \in (0, 2)$.

Para estabelecer o fato anunciado, precisamos de um resultado auxiliar.

Lema 5.4. *Assuma que $S(D) \neq \emptyset$ e seja $\{z_k\}$ a sequência gerada pelo algoritmo DSG com comprimento de passo $\{s_k\}$ satisfazendo*

$$\liminf_k \left[2 \frac{\bar{H} - H_k}{\sigma(f(x_k))^2} - s_k \right] > -\infty. \quad (5.10)$$

Então $\{z_k\}$ é limitada.

Demonstração. Fixemos a solução dual $(\bar{u}, \bar{c}) \in S(D)$. Para tal propósito, suponhamos por contradição, que $\{z_k\}$ não é limitada. Isto significa que quaisquer $\{u_k\}$ ou $\{c_k\}$ são não limitadas. Se $\{u_k\}$ é não limitada, então

$$\infty = \sum_k s_k \|Af(x_k)\| \leq \sum_k s_k \sigma(f(x_k)) \leq \sum_k (s_k + \varepsilon_k) \sigma(f(x_k)),$$

Assim $\{c_k\}$ deve ser não limitada. Portanto, em ambos os casos devemos ter $\{c_k\}$ não limitada. Como a sequência é estritamente crescente, ela tende a infinito. Por outro lado, pela definição do algoritmo DSG,

$$\begin{aligned} \|\bar{u} - u_{k+1}\|^2 &= \|\bar{u} - u_k + s_k Af(x_k)\|^2 \\ &= \|\bar{u} - u_k\|^2 + 2s_k \langle \bar{u} - u_k, Af(x_k) \rangle + s_k^2 \|Af(x_k)\|^2. \end{aligned} \quad (5.11)$$

Para estimar o termo do meio da expressão acima usamos a desigualdade do subgradiente (veja a demonstração do lema 1.4),

$$\bar{H} - H_k \leq \langle \bar{u} - u_k, -Af(x_k) \rangle + (\bar{c} - c_k) \sigma(f(x_k)). \quad (5.12)$$

Multiplicando ambos os lados por $2s_k$ e reorganizando a expressão resultante, obtemos

$$2s_k \langle \bar{u} - u_k, Af(x_k) \rangle \leq -2s_k(\bar{H} - H_k) + 2s_k(\bar{c} - c_k) \sigma(f(x_k))$$

Combinando este fato com (L1) e a eq.(5.11) obtemos

$$\begin{aligned} \|\bar{u} - u_{k+1}\|^2 &\leq \|\bar{u} - u_k\|^2 - 2s_k(\bar{H} - H_k) + 2s_k(\bar{c} - c_k)\sigma(f(x_k)) + s_k^2\|Af(x_k)\|^2 \\ &\leq \|\bar{u} - u_k\|^2 + s_k\sigma(f(x_k)) \left[s_k - \frac{2(\bar{H} - H_k)}{\sigma(f(x_k))^2} + \frac{2(\bar{c} - c_k)}{\sigma(f(x_k))} \right]. \end{aligned} \quad (5.13)$$

Pela condição da eq. (5.10) temos que existe uma constante $\rho \in \mathbb{R}$ e um índice k_0 tal que

$$s_k - \frac{2(\bar{H} - H_k)}{\sigma(f(x_k))^2} \leq \rho,$$

para todo $k \geq k_0$. Como dito acima, $\{c_k\}$ tende ao infinito e por isso existe um índice $k_1 \geq k_0$ tal que

$$\rho \leq \frac{2(c_k - \bar{c})}{\sigma(f(c_k))} \quad \text{para todo } k \geq k_1,$$

onde usamos o fato da sequência $\{\sigma(f(x_k))\}$ ser limitada. Portanto, concluí-se que para todo $k \geq k_1$,

$$s_k - \frac{2(\bar{H} - H_k)}{\sigma(f(x_k))^2} + \frac{2(\bar{c} - c_k)}{\sigma(f(x_k))} \leq 0.$$

Este fato, combinado com a eq.(5.13) nos dá $\|\bar{u} - u_{k+1}\| \leq \|\bar{u} - u_k\|$ para todo $k \geq k_1$ e isto implica que $\{u_k\}$ é limitada. Usando desigualdade de Cauchy-Schwarz em (5.12), obtemos

$$(c_k - \bar{c})\sigma(f(x_k)) \leq -(\bar{H} - H_k) + \|Af(x_k)\|\|\bar{u} - u_k\| \leq \|Af(x_k)\|\|\bar{u} - u_k\| \leq \sigma(f(x_k))\|\bar{u} - u_k\|,$$

onde usamos a Hipótese (L1) na desigualdade da direita. Note que se $f(x_{k_0}) = 0$ para algum k_0 , então o correspondente $(u_{k_0}, c_{k_0}) \in S(D)$ e o DSG para em k_0 . Neste caso a sequência é finita e portanto, limitada. Assim a hipótese de não limitação implica que $f(x_k) \neq 0$ para todo k . Usando este fato na expressão prévia, obtemos

$$(c_k - \bar{c}) \leq \|\bar{u} - u_k\|,$$

e portanto $\{c_k\}$ deve ser limitada, uma contradição. Isto implica que a sequência $\{z_k\}$ deve ser limitada. \square

No próximo teorema vamos ver que uma escolha do comprimento de passo como em (5.9), garante uma equivalência entre o conjunto solução $S(D)$ e a sequência gerada pelo algoritmo DSG.

Teorema 5.3. *Assuma que o comprimento de passo no algoritmo DSG é escolhido de acordo com (5.9). Então as seguintes condições são equivalentes;*

1. A sequência $\{z_k\}$ é limitada.
2. $S(D) \neq \emptyset$.

Demonstração. O fato de que (a) implica (b) é uma consequência do Teorema 5.1 e do lado esquerdo da desigualdade em (5.9). De fato, o Teorema 5.1 implica que cada ponto de acumulação de $\{z_k\}$ é uma solução dual. Em particular, $S(D)$ é não vazio. Para ver que (b) implica (a), basta observar que isto segue-se do Lema 5.4 e do fato de que o lado esquerdo da desigualdade em (5.9) implica (5.10). \square

O teorema seguinte é uma consequência dos Teoremas 5.2 e 5.3.

Teorema 5.4. *Assuma que o comprimento de passo no algoritmo DSG é escolhido de acordo com (5.9). Suponha também que $S(D) \neq \emptyset$. Então,*

- (i) *A sequência dual $\{z_k\}$ converge para a solução dual.*
- (ii) *A sequência de valores duais $\{H_k\}$ converge para um valor ótimo dual.*
- (iii) *Todo ponto de acumulação da sequência primal $\{\tilde{x}_k\}$ é solução do problema (P).*

O seguinte resultado simples é útil para uma implementação do algoritmo.

Proposição 5.2. *Assuma que uma das seguintes condições é satisfeita:*

1. *O comprimento de passo s_k satisfaz (5.6) e $\{z_k\}$ é limitada.*
2. *O comprimento de passo s_k satisfaz (5.9) e $S(D) \neq \emptyset$.*

Então existe uma solução dual (\bar{u}, \bar{c}) tal que $\bar{c} > z_k$ para todo k .

Demonstração. Para a demonstração veja [6] Proposição 5.2. \square

Para implementação numérica, escolha do comprimento do passo e uma melhor explicação do método DSG ver [6].

6 MÉTODO DE DIREÇÕES INTERIORES AO EPÍGRAFO PARA A SOLUÇÃO DE PROBLEMAS DE OTIMIZAÇÃO NÃO-CONVEXOS E NÃO-DIFERENCIÁVEIS VIA DUALIDADE LAGRANGEANA

Neste capítulo apresentaremos o método IED (Interior Epigraph Directions) que foi desenvolvido para resolver problemas de otimização não-convexos e não-diferenciáveis com restrições. O método IED, desenvolvido por Freire (veja [21]), combina os algoritmos DSG e NFDA estudados nos capítulos 4 e 5 respectivamente, de modo a obter um algoritmo que, usando o problema dual, constrói uma sequência de pontos duais no interior do epígrafo da função dual, da seguinte forma:

A partir de um ponto (u^k, c^k, z^k) no interior do epígrafo da função dual encontra-se uma direção d^k , resolvendo-se sistemas de equações semelhantes àsquelas empregadas pelo NFDA, que pertence a um certo cone definido por um subgradiente s^k da função dual. A direção d^k fornece então um novo ponto $(u^{k+1}, c^{k+1}, z^{k+1})$ que pode não estar no interior do epígrafo da função dual. Se o ponto $(u^{k+1}, c^{k+1}, z^{k+1})$ é interior então tem-se o que se chama passo sério e este ponto será o novo iterado. Caso $(u^{k+1}, c^{k+1}, z^{k+1})$ não esteja no interior do epígrafo da função dual, tem-se um passo nulo e utiliza-se a técnica empregada pelo DSG para gerar o novo iterado. Essa sequência dual gera uma sequência primal através da minimização da função Lagrangeana empregada pelo IED. Se o ponto solução deste problema satisfaz a restrição então este ponto é a solução do problema original e tem-se assim um critério de parada para o algoritmo.

Descreveremos na próxima seção o método IED para então apresentarmos o algoritmo e finalmente a análise de convergência do método.

6.1 O MÉTODO IED

Consideremos o problema de otimização restrita

$$(P) \quad \begin{cases} \text{minimizar } f_0(x) \\ f(x) = 0 \\ x \in X \end{cases}$$

onde X é um subconjunto compacto do \mathbb{R}^n e $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$, $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são funções contínuas.

Apesar de (P) conter apenas restrições de igualdade, restrições de desigualdade podem ser consideradas através do emprego do operador $a^+ = \max\{a, 0\}$, $a \in \mathbb{R}$.

Resolve-se o problema (P) , da seguinte maneira: Consideremos $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+ \rightarrow \mathbb{R}$ a função lagrangeana aumentada generalizada associada ao problema primal (P) , definida como

$$L(x, (u, c)) = f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle,$$

onde A , σ são definidas como em 1.25 do capítulo 1.

Então a função dual correspondente $H : \mathbb{R}^m \times \mathbb{R}_+ \rightarrow \mathbb{R}$ é definida como

$$H(u, c) = \min_{x \in X} \{f_0(x) + c\sigma(f(x)) - \langle Au, f(x) \rangle\}.$$

Portanto o problema dual associado ao problema primal (P) é dado por

$$(D) \quad \begin{cases} \text{maximizar } H(u, c) \\ (u, c) \in \mathbb{R}^m \times \mathbb{R}_+ \end{cases},$$

o qual é equivalente a

$$(D^*) \quad \begin{cases} \text{minimizar } -H(u, c) \\ (u, c) \in \mathbb{R}^m \times \mathbb{R}_+. \end{cases}$$

Feita esta equivalência, obtemos um problema de otimização convexa irrestrita, o qual é, por sua vez, equivalente a um problema de minimização de uma função linear sujeita a uma restrição de desigualdade não necessariamente diferenciável, a saber

$$(ED^*) \quad \begin{cases} \text{minimizar } \psi(u, c, z) = z \\ -H(u, c) \leq z \\ (u, c, z) \in \mathbb{R}^m \times \mathbb{R}_+ \times \mathbb{R}. \end{cases}$$

A partir do ponto $v^k = (c^k, u^k)$ traçamos um hiperplano suporte à restrição do problema (ED^*) definido pela equação $z = -H(v^k) + (s^k)^T(v^k - v)$ com $s^k \in \partial(-H(v^k))$.

Então uma aproximação do problema (ED^*) é dada por

$$(AC_k) \quad \begin{cases} \text{minimizar } \psi(v, z) = z \\ g^k(v, z) \leq 0, \end{cases}$$

onde $g^k(p) = -H(v^k) + (s^k)^T(v - v^k) - z$, com $p = (v, z) = ((u, c), z)$. (AC_k) é chamado problema auxiliar. Uma das diferenças entre NFDA e IED é que este último só utiliza um único hiperplano suporte em cada iteração. Aplicando-se a mesma técnica dos métodos FDIPA e NFDA ao problema auxiliar (AC_k) encontramos uma direção $d^k = d_\alpha^k + \rho^k d_\beta^k$ que é viável e de descida para o problema (AC_k) . As direções d_α^k e d_β^k são encontradas resolvendo-se os sistemas lineares:

$$\begin{aligned} B^k d_\alpha^k + \lambda_\alpha^k \nabla g^k(p^k) &= -\nabla \psi(p^k) \\ \lambda^k \nabla g^k(p^k)^T d_\alpha^k + \lambda_\alpha^k g^k(p^k) &= 0 \end{aligned} \quad (6.1)$$

$$\begin{aligned} B^k d_\beta^k + \lambda_\beta^k \nabla g^k(p^k) &= 0 \\ \lambda^k \nabla g^k(p^k)^T d_\beta^k + \lambda_\beta^k g^k(p^k) &= -\lambda^k \end{aligned} \quad (6.2)$$

e $\rho^k > 0$ é o parâmetro de deflexão. A direção d^k pode ser calculada mesmo se o problema (AC_k) não tiver mínimo finito. Para descrever o método IED precisamos de algumas definições e alguns resultados importantes.

Definição 6.1. *Se um vetor $(p^k, \lambda^k) = ((v^k, z^k), \lambda^k)$ verifica (2.3)-(2.4), então chamaremos à (p^k, λ^k) uma iteração NFDA viável. Se $g(p^k) < 0$ e $\lambda^k > 0$, então chamaremos (p^k, λ^k) uma iteração NFDA fortemente viável.*

O seguinte lema mostra que d_α^k é uma direção de descida para ψ .

Lema 6.1. *Suponha que $(p^k, \lambda^k) = ((v^k, z^k), \lambda^k)$ é uma iteração NFDA fortemente viável e B^k é uma matriz simétrica definida positiva. Então, a direção d_α^k satisfaz:*

$$(d_\alpha^k)^T \nabla \psi(p^k) = (d_\alpha^k)^T \nabla \psi(p^k) \leq -(d_\alpha^k)^T B^k d_\alpha^k. \quad (6.3)$$

Antes de mostrar o próximo lema requeremos uma hipótese para a matriz B^k .

Hipótese 6.1 Assuma que para qualquer $k \in \mathbb{N}$, a matriz $B^k \in \mathbb{R}^{(m+2) \times (m+2)}$ tenha seguinte estrutura:

$$B^k = \begin{bmatrix} C^k & 0 \\ 0 & \gamma^k \end{bmatrix}$$

onde $C^k \in \mathbb{R}^{(m+1) \times (m+1)}$ é definida positiva e $\gamma^k > \bar{\gamma} > 0$.

Lema 6.2. *Suponha que valha a hipótese 6.1. Sejam $s^k \neq 0$ um subgradiente e $(p^k, \lambda^k) = ((v^k, z^k), \lambda^k)$ uma iteração NFDA fortemente viável. Então temos os seguintes fatos:*

$$a) \nabla g^k(p^k)^T (B^k)^{-1} \nabla g^k(p^k) = (s^k)^T (C^k)^{-1} s^k + (1/\gamma^k) = \theta^k / \gamma^k > 0, e$$

$$\nabla g(p^k)^T (B^k)^{-1} \nabla \psi(p^k) = -(1/\gamma^k) < 0. \text{ Onde } \theta^k = (1 + \gamma^k (s^k)^T (C^k)^{-1} s^k).$$

$$b) \nabla g^k(p^k)^T d_\alpha^k = (1/\gamma^k)(1 - \lambda_\alpha^k \theta^k), e \nabla \psi(p^k)^T d_\alpha^k = (1/\gamma^k)(\lambda_\alpha^k - 1).$$

$$c) 0 < \lambda_\alpha^k < 1, \lambda_\beta^k = \gamma^k \lambda_\alpha^k. \text{ Portanto, } \nabla \psi(p^k)^T d_\alpha^k < 0.$$

$$d) 0 > \nabla g^k(p^k)^T d_\beta^k = -\lambda_\alpha^k \theta > -1, \nabla \psi(p^k)^T d_\beta^k = \lambda_\alpha^k.$$

e) Defina

$$\rho_1 = -\frac{\nabla g^k(p^k)^T d_\alpha^k}{\nabla g^k(p^k)^T d_\beta^k} e \rho_2 = -\frac{\nabla \psi(p^k)^T d_\alpha^k}{\nabla \psi(p^k)^T d_\beta^k}.$$

$$\text{Então } \rho_1^k > 0, \rho_2^k > 0 e \Delta = \rho_2^k - \rho_1^k > 0.$$

Demonstração. a) Segue diretamente da Hipótese 6.1.

b) Da primeira equação (6.1), temos que

$$d_\alpha^k = -(B^k)^{-1} (\nabla \psi(p^k) + \lambda_\alpha^k \nabla g^k(p^k)), \quad (6.4)$$

o que implica

$$\begin{aligned} \nabla g^k(p^k)^T d_\alpha^k &= -\nabla g^k(p^k)^T (B^k)^{-1} (\nabla \psi(p^k) + \lambda_\alpha^k \nabla g^k(p^k)) \\ &= 1/\gamma^k - \lambda_\alpha^k \theta / \gamma^k = 1/\gamma^k (1 - \lambda_\alpha^k \theta). \end{aligned}$$

De (6.4) também temos

$$\begin{aligned} \nabla \psi(p^k)^T d_\alpha^k &= -\nabla \psi(p^k)^T (B^k)^{-1} \nabla \psi(p^k) - \lambda_\alpha^k \nabla \psi(p^k)^T (B^k)^{-1} \nabla g^k(p^k) \\ &= -1/\gamma^k - \lambda_\alpha^k (-1/\gamma^k) = -(1/\gamma^k)(1 - \lambda_\alpha^k). \end{aligned}$$

Isto completa a prova de (b). (c) Usando (b) e a segunda equação de (6.1) obtemos $(1/\gamma^k)(1 - \lambda_\alpha^k \theta^k) = \nabla g^k(p^k)^T d_\alpha^k = -\frac{\lambda_\alpha^k}{\lambda^k} g^k(p^k)$. Combinando estas igualdades, reordenando a expressão resultante e usando a Hipótese 6.1 obtemos

$$\lambda_\alpha^k \left(-\frac{g^k(p^k)}{\lambda^k} + \theta^k/\gamma^k \right) = 1/\gamma^k > 0. \quad (6.5)$$

Como (p^k, λ^k) é uma iteração NFDA fortemente viável e $\theta^k/\gamma^k > 0$, as duas expressões entre parênteses são estritamente positivas, o que garante que $\lambda_\alpha^k > 0$. Agora usamos a primeira equação (6.2) para escrever

$$d_\beta^k = -\lambda_\beta^k (B^k)^{-1} \nabla g^k(p^k), \quad (6.6)$$

o que implica

$$\nabla g^k(p^k)^T d_\beta^k = -\lambda_\beta^k \nabla g^k(p^k)^T (B^k)^{-1} \nabla g^k(p^k) = -\lambda_\beta^k (\theta^k/\gamma^k). \quad (6.7)$$

Da segunda equação (6.2), temos:

$$\nabla g^k(p^k)^T d_\beta^k = -1 - \frac{\lambda_\beta^k g^k(p^k)}{\lambda^k}.$$

Combinando as duas últimas expressões temos

$$\lambda_\beta^k \left(\theta^k/\gamma^k - \frac{\nabla g^k(p^k)}{\lambda^k} \right) = 1. \quad (6.8)$$

Esta igualdade, junto com (6.5), implica $\lambda_\beta^k/\lambda_\alpha^k = \gamma^k$. Para finalizar a prova de (c), precisamos mostrar que $\lambda_\alpha^k < 1$. O último fato e (b) garante diretamente $\nabla \psi(p^k)^T d_\alpha^k < 0$. Por (6.8) e o fato de que $\lambda_\beta^k/\lambda_\alpha^k = \gamma$, podemos escrever

$$\lambda_\alpha^k = \lambda_\beta^k/\gamma^k = \frac{1}{\theta - \frac{\gamma^k g^k(p^k)}{\lambda^k}} < \frac{1}{\theta} \leq 1 \quad (6.9)$$

onde usamos $\theta^k = 1 + \gamma^k (s^k)^T (C^k)^{-1} s^k \geq 1$ e o fato de que $\theta^k - \frac{\gamma^k g^k(p^k)}{\lambda^k} > \theta^k \geq 1$,

o que implica

$$0 < \lambda_\alpha^k \theta^k < 1. \quad (6.10)$$

Portanto a prova está completa. Note que $\lambda_\alpha^k/\lambda_\beta^k = \gamma^k$ e (6.7) leva a $\nabla g^k(p^k)^T d_\beta^k = -\lambda_\alpha^k \theta^k$, o qual é a primeira igualdade em (d). Usando agora (6.10) obtemos $-1 < \nabla g^k(p^k)^T d_\beta^k < 0$. Para finalizar a prova de (d), use (6.6) e (a) para escrever

$$\nabla \psi(p^k)^T d_\beta^k = -\lambda_\beta^k \nabla \psi(p^k)^T (B^k)^{-1} \nabla g^k(p^k) = \frac{\lambda_\beta^k}{\gamma^k} = \lambda_\alpha^k$$

Para provar (e), usamos (b) e (d) para obter

$$\rho_1^k = -\frac{(1/\gamma^k)(1 - \lambda_\alpha^k \theta)}{-\lambda_\alpha^k \theta} = \frac{1}{\gamma^k} \left(\frac{1}{\lambda_\alpha^k \theta^k} - 1 \right)$$

e

$$\rho_2^k = \frac{1}{\gamma^k} \left(\frac{1}{\lambda_\alpha^k} - 1 \right)$$

Portanto $\rho_1^k > 0$ por (6.10). Também obtemos que

$$\Delta = \rho_2^k - \rho_1^k > 0 \text{ do fato de que } \theta = (1 + \gamma^k (s^k)^T (C^k)^{-1} s^k) > 1$$

De fato, o último mantém-se pela Hipótese 6.1 e do fato de que $s^k \neq 0$. Então a prova é completa. \square

O seguinte lema mostra que se o parâmetro de deflexão ρ^k é limitado, então d^k é uma direção de descida para ψ .

Lema 6.3. *Fixe $\xi \in (0, 1)$ e assuma as mesmas hipóteses do lema 6.2. Seja $d^k = d_\alpha^k + \rho^k d_\beta^k$, onde $\rho^k > 0$, e d_α^k e d_β^k são soluções dos sistemas (6.1) e (6.2). Se $0 < \rho^k \leq (\xi - 1) \frac{\nabla \psi(p^k)^T d_\alpha^k}{\nabla \psi(p^k)^T d_\beta^k}$, então $\nabla \psi(p^k)^T d^k < 0$, com $p^k = (v^k, z^k)$.*

Demonstração. Do Lema 6.2(d) temos $\nabla \psi(p^k)^T d_\beta^k = \lambda_\alpha^k > 0$. Agora segue-se facilmente de (6.3). De fato,

$$\begin{aligned} \nabla \psi(p^k)^T d^k &= \nabla \psi(p^k)^T d_\alpha^k + \rho \nabla \psi(p^k)^T d_\beta^k \\ &\leq \nabla \psi(p^k)^T d_\alpha^k + (\xi - 1) \nabla \psi(p^k)^T d_\beta^k \frac{\nabla \psi(p^k)^T d_\alpha^k}{\nabla \psi(p^k)^T d_\beta^k} \\ &= \nabla \psi(p^k)^T d_\alpha^k + (\xi - 1) \nabla \psi(p^k)^T d_\alpha^k \\ &= \xi \nabla \psi(p^k)^T d_\alpha^k < 0, \end{aligned}$$

onde usamos o Lema 6.2(c) na última desigualdade. \square

O NFDA minimiza funções convexas não diferenciáveis cujos conjuntos de nível são limitados. Nossa função convexa $-H$, em geral, não satisfaz esta propriedade. Portanto, introduziremos aqui uma estratégia para escolher direções adequadas para a função convexa $-H$ que temos. Precisamos impor uma condição adicional à direção d^k , isto é, d^k deve ser uma direção de descida para g^k , ou seja

$$\nabla g^k(v^k, z^k)^T d^k < 0. \quad (6.11)$$

Além disso, em cada iteração, queremos uma direção d^k que pertença ao interior do cone, definido na iteração k com iterato $p^k = (v^k, z^k) \in \mathbb{R}^{(m+1) \times 1}$ como segue:

$$K(p^k) = \{d \in \mathbb{R}^{m+1} \times \mathbb{R} / d^T \nabla g^k(p^k) < 0 \text{ e } d^T \nabla \psi(p^k) < 0\} \quad (6.12)$$

Apresenta-se algumas condições para o parâmetro de reflexão ρ^k de tal maneira que a direção de busca $d^k = d_\alpha^k + \rho^k d_\beta^k$ encontra-se sempre no cone $K(p^k)$. O seguinte lema mostra que a condição (6.11) mantêm-se sempre que $\rho^k > \rho_1^k$.

Lema 6.4. *Assuma que toda condição do Lema 6.2 mantêm-se e seja $d^k = d_\alpha^k + \rho^k d_\beta^k$, onde $\rho^k > 0$ e d_α^k e d_β^k resolvem os sistemas (6.1) e (6.2). Se $\rho^k > -\frac{(d_\alpha^k)^T \nabla g^k(p^k)}{(d_\beta^k)^T \nabla g^k(p^k)} = \rho_1^k$, com $p^k = (v^k, z^k)$ então (6.11) mantêm-se.*

Demonstração. Temos que

$$(d^k)^T \nabla g^k(p^k) = (d_\alpha^k)^T \nabla g^k(p^k) + \rho^k (d_\beta^k)^T \nabla g^k(p^k) < 0$$

se, e somente se $\rho^k (d_\beta^k)^T \nabla g^k(p^k) < -(d_\alpha^k)^T \nabla g^k(p^k)$. Pelo Lema 6.2(d) temos que $(d_\beta^k)^T \nabla g^k(p^k) < 0$ e portanto, a desigualdade acima é equivalente a $\rho^k > -\frac{(d_\alpha^k)^T \nabla g^k(p^k)}{(d_\beta^k)^T \nabla g^k(p^k)}$, como desejado. \square

Em seguida mostraremos a existência do parâmetro de reflexão ρ^k .

Corolário 6.1.1. *Dadas as condições do Lema 6.2, 6.3 e 6.4, existe $\rho^k > 0$ tal que*

$$\rho_1^k = -\frac{(d_\alpha^k)^T \nabla g^k}{(d_\beta^k)^T \nabla g^k} < \rho^k \leq (1 - \xi^k) \frac{(d_\alpha^k)^T \nabla \psi(p^k)}{(d_\beta^k)^T \nabla \psi(p^k)} < \rho_2^k.$$

Consequentemente, dado (p^k, λ^k) uma iteração NFDA fortemente viável, existe $\rho^k > 0$ tal que

$$d^k = d_\alpha^k + \rho^k d_\beta^k \in K(p^k).$$

Demonstração. O fato de ρ^k poder ser escolhido entre ρ_1^k e ρ_2^k segue do Lema 6.2(e). Se ρ^k também verifica a desigualdade restante, então $d^k \in K(p^k)$ como resultado do Lema 6.3 e 6.4. \square

Comentário 6.1 Sob as hipóteses do Lema 6.2, temos que $\rho_1^k < \rho_2^k$. Portanto, como demonstrado acima, existe $\xi^k \in (0, 1)$ tal que

$$\rho_1^k < (1 - \xi^k)\rho_2^k < \rho_2^k. \quad (6.13)$$

Na análise de convergência, escolhe-se o parâmetro ρ^k usado no Corolário 6.1.1 de forma que $\rho^k = (1 - \xi^k)\rho_2^k$ com

$$\xi^k \in \left(\frac{1}{3} \left(1 - \frac{\rho_1^k}{\rho_2^k} \right), \frac{2}{3} \left(1 - \frac{\rho_1^k}{\rho_2^k} \right) \right) \subset (0, 1). \quad (6.14)$$

A seguinte proposição mostra que todo ponto NFDA fortemente viável fornece um ponto NFDA fortemente viável.

Proposição 6.1. *Mantêm-se as hipóteses e notações dos Lemas 6.2, 6.3, 6.4 e o Corolário 6.1.1. Fixado qualquer $t > 0$ e assumindo que $\pi^k = (p^k, \lambda^k) = ((v^k, z^k), \lambda^k)$ é uma iteração NFDA fortemente viável, e $p^k \in \text{int}(\text{epi}(-H))$, então $\hat{\pi}^k = (p^k + td^k, \lambda^k + td_\lambda^k)$, onde $d_\lambda^k = \lambda_\alpha^k + \rho^k \lambda_\beta^k$, é uma iteração NFDA fortemente viável.*

Demonstração. A escolha de ρ^k fica clara considerando-se o Lema 6.2(c) e do fato de que $\lambda^k > 0$, tem-se $\lambda^k + td_\lambda^k > 0$. Portanto só precisamos mostrar que $g^k(p^k + td^k) < 0$. Denote $d^k = (d_v^k, d_z^k)$. Assim d_v^k e d_z^k são a v e z componentes e d^k lembrando que $d_z^k = \nabla \psi(p^k)^T d^k < 0$. Uma vez que $p^k \in \text{int}(\text{epi}(-H))$ temos

$$-H(v^k) < z. \quad (6.15)$$

Pelo Corolário 6.1.1 temos que $\nabla g^k(p^k)^T d^k < 0$ e assim obtemos

$$0 > \nabla g^k(p^k)^T d = (s^k)^T d_v^k + (-1)d_z^k = (s^k)^T d_v^k - \nabla \psi(p^k)^T d^k. \quad (6.16)$$

De (6.15) e (6.16) e da definição de g^k podemos escrever

$$\begin{aligned} g^k(p^k + td^k) &= g(v^k + td_v^k, z^k + td_z^k) = -H(v^k) + t(s^k)^T d_v^k - z^k - td_z^k \\ &= -H(v^k) - z^k + t((s^k)^T d_v^k - \nabla \psi(p^k)^T d^k) < 0, \end{aligned}$$

portanto $g^k(p^k + td^k) < 0$ e concluímos que $\hat{\pi}$ é NFDA fortemente viável. \square

Definiremos agora o que é um passo NFDA e o passo DSG.

Definição 6.2. *Seja $d \in \mathbb{R}^{m+1} \times \mathbb{R}$, ρ como no Corolário 6.1.1, e λ_α e λ_β soluções dos sistemas (6.1) e (6.2). Defina $d_\lambda = \lambda_\alpha + \rho\lambda_\beta$. Então o passo NFDA a partir de $\pi = (p, \lambda) = ((v, z), \lambda)$ é definido como*

$$\hat{\pi} = (p + td, \lambda + td_\lambda), \quad (6.17)$$

onde $t > 0$.

Definição 6.3. *Dado $v = (u, c) \in \mathbb{R}^m \times \mathbb{R}_+$. Seja A e σ como na definição 1.25 e x como no Corolário 1.5.1. Defina a direção $d_{DSG} = (Af(x), -(1 + \alpha)\sigma(f(x)))$, onde $\alpha > 0$. Seja $\pi = ((v, z), \mu)$. O passo DSG a partir de (v, z) com $r > 0$ é dado por*

$$(\tilde{v}, \tilde{z}) = (v + r(Af(x), -(1 + \alpha)\sigma(f(x))), \tilde{z}), \text{ com } \tilde{z} \in (-H(\tilde{v}), -H(v)).$$

Então o passo completo DSG é dado por

$$\tilde{\pi} = ((\tilde{v}, \tilde{z}), \mu)$$

Comentário 6.2 Note que λ é alterado no passo NFDA mas é mantido fixo no passo DSG.

Comentário 6.3 Como uma consequência da Proposição 5.1(i), temos que $-H_{k+1} = -H(u^{k+1}, c^{k+1}) < -H_k = -H(u^k, c^k)$. Portanto, pela escolha de $z \in (-H_{k+1}, -H_k)$ devemos ter $((u^{k+1}, c^{k+1}), z) \in \text{int}(\text{epi}(-H))$, isto é, $(v^k, z^k) + d_{DSG} \in \text{int}(\text{epi}(-H))$.

Para gerais iterações, iniciamos em um ponto no $\text{int}(\text{epi}(-H))$ e desejamos que o iterato seguinte também seja um ponto interior. Se começamos com um ponto NFDA fortemente viável (v, z, λ) tal que $(v, z) \in \text{int}(\text{epi}(-H))$ o passo NFDA gera um ponto $\hat{\pi} = (\hat{v}, \hat{z}, \hat{\lambda})$ NFDA fortemente viável, mas não garante que o novo ponto pertença ao $\text{int}(\text{epi}(-H))$. Então se $(\hat{v}, \hat{z}) \notin \text{int}(-H)$, voltamos ao passo original e fazemos um passo do subgradiente defletivo (DSG), isto é, obtemos um ponto $\tilde{\pi} = ((v, z) + d_{DSG}, \lambda)$, onde d_{DSG} é a direção DSG o qual garante que $(v, z) + d_{DSG} \in \text{int}(\text{epi}(H))$ e portanto realizamos novamente o passo NFDA a partir de $\tilde{\pi}$. Na seguinte seção apresenta-se o passo IED.

6.1.1 O passo do IED

Suponha que no passo do IED temos um ponto $p^k = (v^k, z^k) \in \text{int}(\text{epi}(-H))$, com $\lambda^k > 0$, o qual é um ponto NFDA fortemente viável e uma iteração $x^k \in X(v^k)$ tal que $f(x^k) \neq 0$. Então escolhemos uma direção $d^k = d_\alpha^k + \rho^k d_\beta^k$ onde d_α^k e d_β^k são soluções dos sistemas 6.1 e 6.2 e o parâmetro de deflexão ρ^k é tomado como no Corolário 6.1.1 tal que $d^k \in K(p^k)$. Logo, escolhemos o comprimento de passo da seguinte maneira

$$t^k = \max\left\{\eta_0, \frac{z^k - \hat{q}}{-\nabla \psi(p^k)^T d^k}, \frac{|(v^k)^T s^k|}{-\nabla \psi(p^k)^T d^k}\right\}, \quad (6.18)$$

onde $\eta_0 > 0$ e $-\hat{H}$ são parâmetros fixos, e $\hat{H} \geq \bar{H}$ com \bar{H} solução dual. Então produz-se um ponto auxiliar

$$\hat{\pi}^k = (\hat{p}^k, \hat{\lambda}^k) = (y^k, w^k, \hat{\lambda}^k) = (v^k, z^k, \lambda^k) + t^k(d^k, d_\lambda^k), \text{ onde } d_\lambda^k = \lambda_\alpha^k + \rho^k \lambda_\beta^k,$$

que é NFDA fortemente viável, e declaramos $i = 0$ índice da iteração nula. Daí, calculamos $\hat{x}^k \in X(y^k)$. Se $f(\hat{x}^k) = 0$, então pelo Teorema 1.17 os pontos y^k e \hat{x}^k são soluções dos problema dual e primal, respectivamente e então paramos. Isto é o critério de parada do método *IED*. Se $f(x^k) \neq 0$ e $w^k > -H(y^k)$ dizemos que o passo é atingido, declaramos este passo como **sério**, a nova iteração será

$$(p^{k+1}, \lambda^{k+1}) = (v^{k+1}, z^{k+1}, \lambda^{k+1}) = (y^k, w^k, \hat{\lambda}^k), \quad x^{k+1} = \hat{x}^k,$$

fazemos $k = k + 1$ (veja figura 6.1) e voltamos a passo k . Caso contrário se $w^k \leq -H(y^k)$, dizemos que o passo não é atingido e declaramos este passo **nulo**. Mantemos k fixo e fazemos $i = i + 1$. Realiza-se o passo DSG com uma direção d_{DSG} a partir do ponto p^k tal como se mostra na figura 6.2, para obtermos então o ponto

$$\tilde{\pi} = (\tilde{p}^{k,i}, \lambda^k) = (\hat{v}^{k,i}, \hat{z}^{k,i}, \lambda^k).$$

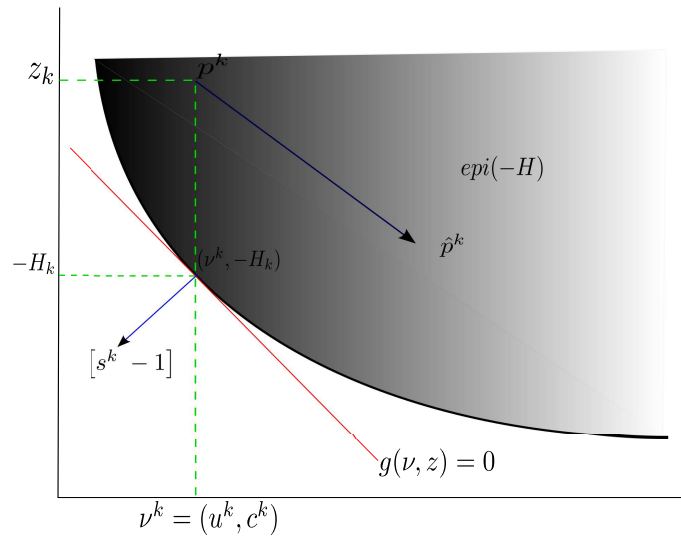


Figura 6.1: Passo s3rio do M3todo IED

Calculamos $\tilde{x}^{k,i} \in X(\tilde{v}^{k,i})$. Se $f(\tilde{x}^{k,i}) = 0$, novamente aplicamos o Teorema 1.17 e paramos. Casos contr3rio calcule

$$\tilde{z}^k \in \left(-H(\tilde{v}^{k,i}), -H(\nu^k) \right)$$

e tome a nova itera33o $\nu^k = \tilde{\nu}^{k,i}$, $z^k = \tilde{z}^k$, $x^k = \tilde{x}^{k,i}$ e volte ao passo k.

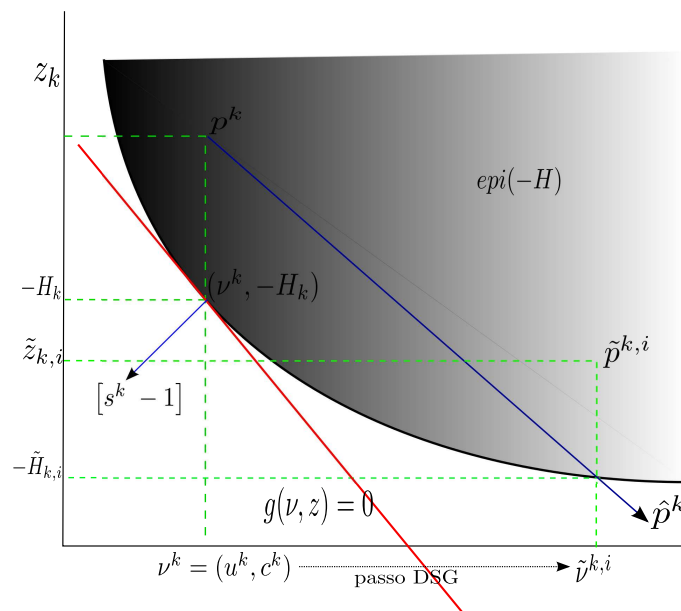


Figura 6.2: Passo nulo e a dire33o d_{DSG}

Explicado o passo principal, descrevemos agora o algoritmo IED.

6.2 O ALGORITMO DE DIREÇÕES INTERIORES AO EPÍ- GRAFO

Escolha $\eta_0 \in \langle 0, 1 \rangle$, $\hat{H} \geq \bar{H}$ com \bar{H} valor ótimo do problema dual, uma matriz simétrica A e uma função aumentada σ . Tome a sequência $\{\alpha_k\} \subset \mathbb{R}_+$ de números positivos e uma sequência $\{B^k\}$ de matrizes definidas positivas como nas hipóteses 6.1 e 6.2.

Passo 0 Fixe $u^0 \in \mathbb{R}^m$, $c^0 > 0$ e $\lambda_0 > 0$. Seja $v^0 = (u^0, c^0)$, $z^0 > -H(v^0)$. Tome

$$x^0 \in X(v^0) \text{ tal que } f(x^0) \neq 0. \text{ Seja } k = 0$$

Passo k Seja $p^k = (v^k, z^k) \in \text{int}(\text{epi}(-H))$, $\lambda^k > 0$ e $x^k \in X(v^k)$, tal que $f(x^k) \neq 0$. Calcule s^k como Corolário 1.5.1.

Passo k.1 Seja $i = 0$. Dado (p^k, λ^k) , encontre $d_\alpha^k, d_\beta^k, \lambda_\alpha^k$ e λ_β^k resolvendo os sistemas (6.1) e (6.2). Calcule ρ_1^k, ρ_2^k como no Corolário 6.1.1 e a direção de busca

$$d^k = d_\alpha^k + \rho^k d_\beta^k,$$

onde $\rho_1^k < \rho^k = (1 - \xi^k)\rho_2^k < \rho_2^k$, com ξ^k como no comentário 6.2. Calcule agora o comprimento de passo como

$$t^k = \max \left\{ \eta_0, \frac{z^k + \hat{H}}{-\nabla \psi(v^k, z^k)^T d^k}, \frac{|(v^k)^T s^k|}{-\nabla \psi(v^k, z^k)^T d^k} \right\}$$

Passo k.2 Calcule o ponto auxiliar $\hat{p}^k = (y^k, w^k) = p^k + t^k d^k$. Encontre $\hat{x}^k \in X(y^k)$. Se $f(\hat{x}^k) = 0$, pare. Sejam $x^k = \hat{x}^k$ e $v^k = y^k$.

Passo k.3.a (Passo sério) Se $\hat{p}^k \in \text{int}(\text{epi}(q))$ então $p^{k+1} = \hat{p}^k$, $x^{k+1} = \hat{x}^k$, $\lambda^{k+1} = \lambda^k + t^k(\lambda_\alpha^k + \rho^k \lambda_\beta^k) = \hat{\lambda}^k$, $k = k + 1$ e ir ao passo k.3.a

Passo k.3.b (Passo Nulo) Faça $i = i + 1$. Escolha $s_k > 0$ como no passo 2 do algoritmo DSG no capítulo 5. Calcule um novo ponto no $\text{int}(\text{epi}(-H))$ realizando o passo DSG a partir de p^k :

$$\tilde{u}^{k,i} = u^k + s^k A(f(x^k));$$

$$\tilde{c}^{k,i} = c^k - (1 + \alpha_k) s^k \sigma(f(x^k)) :$$

$$\tilde{v}^k = (\tilde{u}^{k,i}, \tilde{c}^{k,i}).$$

Encontre $\tilde{x}^{k,i} \in X(\tilde{v}^{k,i})$. Se $f(\tilde{x}^{k,i}) = 0$, pare. Caso contrário, calcule $-H(\tilde{v}^{k,i})$

e tome $\tilde{z}^{k,i}$ tal que $-H(\tilde{v}^{k,i}) < \tilde{z}^{k,i} < -H(v^k)$. Seja

$$p^k = (\tilde{v}^{k,i}, \tilde{z}^{k,i});$$

$$x^k = \tilde{x}^{k,i};$$

vá para o passo k.

Comentário 6.4 Mencionamos que quando o contador i é incrementado em cada passo nulo o contador k para o passo sério é mantido constante e λ^k não é atualizado.

6.2.1 Diagrama de fluxo do Método IED

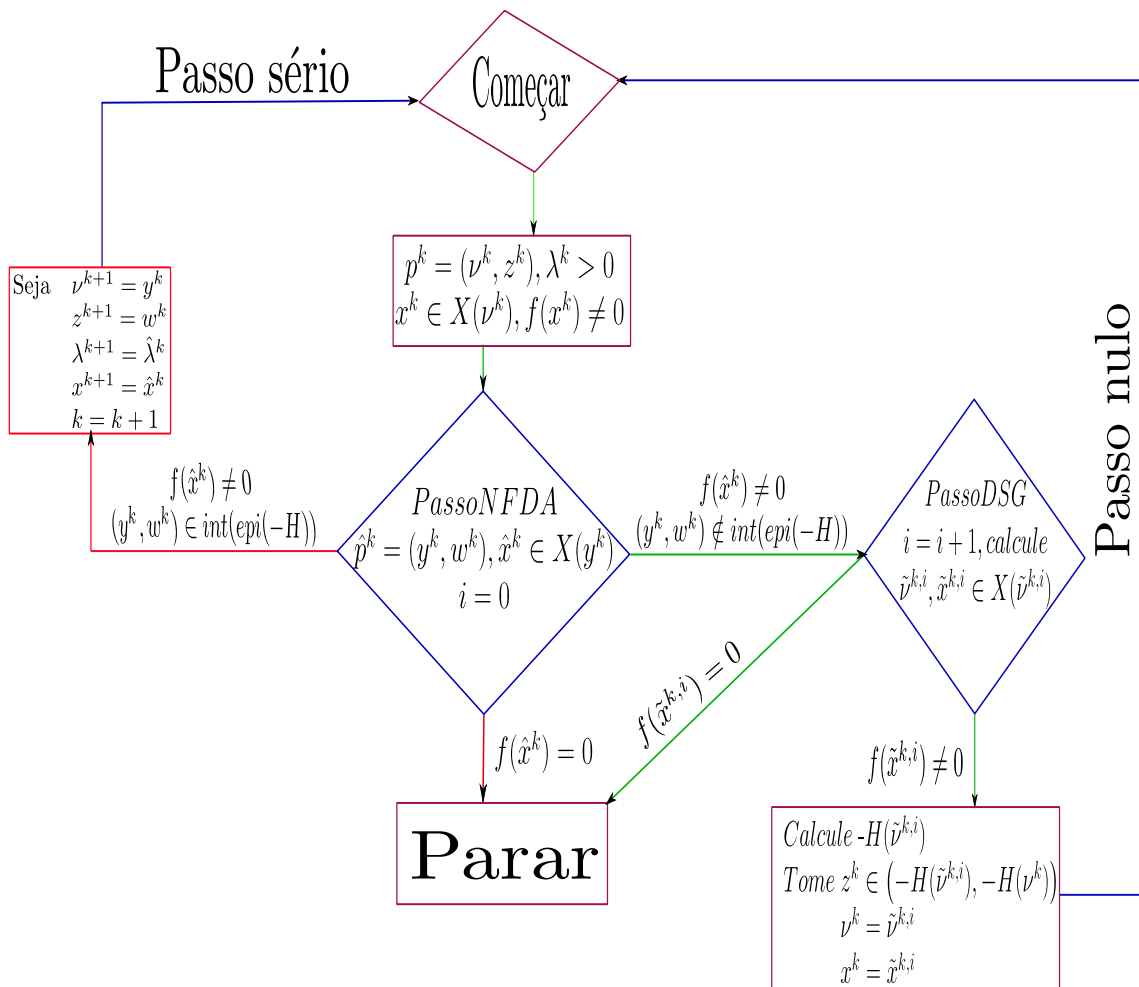


Figura 6.3: Diagrama de fluxo do Método IED

6.3 CONVERGÊNCIA DO MÉTODO IED

Nesta seção mostraremos que todo ponto de acumulação da sequência $\{x^k\}$ gerada pelo algoritmo IED é solução do problema (P) . Para isto, considera-se primeiramente o caso em que IED gera um número finito k_0 de passos sérios e um número finito i_0 de passos nulos. Neste caso, o ponto x^{k_0} será solução de (P) . Depois, considera-se o caso em que temos infinitos passos nulos e um número finito de passos sérios. A convergência neste caso é garantida pelo DSG. Para o caso em que ocorrem infinitos passos sérios, mostra-se primeiro que as sequências $\{-H^k\}$ e $\{z^k\}$ são decrescentes e ambas convergem para a solução do dual. Em seguida, mostra-se que se a sequência $\{\xi^k\}$ converge a zero então $s^k \rightarrow 0$. Finalmente, mostra-se que todo ponto de acumulação de $\{x^k\}$ é viável.

Para mostrar a convergência do método IED, precisamos da seguinte hipótese.

Hipótese 6.2 Para qualquer $k \in \mathbb{N}$, existem números positivos σ_1 e σ_2 tais que

$$\sigma_1 \|v\|^2 \leq v^T (C^k)^T v \leq \|v\|^2 \sigma_2, \quad \forall v \in \mathbb{R}^n.$$

Teorema 6.1. *Se o algoritmo para em um passo sério $k = k_0$ e um passo nulo $i = i_0$ então x^{k_0} é uma solução de (P) e v^{k_0} é uma solução dual.*

Demonstração. Caso IED pare em $k = k_0$ e $i = i_0$, isto acontecerá no passo $k_0.2$ ou $k_0.3.b$. Em ambos os casos temos que $f(x^{k_0}) = 0$ com $x^{k_0} \in X(v^{k_0})$. Pelo Teorema 1.17 temos que x^{k_0} é uma solução primal e v^{k_0} é uma solução dual. \square

A prova do seguinte resultado é uma consequência direta do Teorema 5.4.

Teorema 6.2. *Assuma que o problema dual tenha solução e que o comprimento do passo no algoritmo DSG satisfaz (5.9). Se para algum k_0 fixo o método fornece infinitos passos nulos, então temos que cada ponto de acumulação da sequência primal $\{\hat{x}^{k_0,i}\}$ é uma solução de (P) . Além disso, a sequência das variáveis duais $\{\hat{v}^{k_0,i}\}$ converge à solução dual.*

Para a análise em que temos infinitos passos sérios, começaremos com alguns lemas simples. Lembre-se que para uma direção dada $d = ((u, c), z) \in \mathbb{R}^{n+1} \times \mathbb{R}$, podemos denotar $d_z = \nabla \psi(v, z)^T d$. O seguinte lema mostra que a sequência $\{z_k\}$ gerada pelo algoritmo IED é uma sequência decrescente.

Lema 6.5. *Suponha que na iteração k tenhamos um passo sério e denote por (v^k, z^k) e (v^{k+1}, z^{k+1}) os correspondentes iterados consecutivos. Denote por $H_k = H(v^k)$ e $H_{k+1} = H(v^{k+1})$. Temos*

que

$$z^k > -H_k > z^{k+1} > -H_{k+1}.$$

Demonstração. Do passo k.2 e k.3 temos

$$z^{k+1} - z^k = \nabla \psi(v^k, z^k)^T (p^{k+1} - p^k) = t^k \nabla \psi(v^k, z^k)^T (d^k) = t^k d_z^k.$$

Escolhendo ρ^k como no Corolário 6.1.1, temos que $\nabla \psi(v^k, z^k)^T d^k = d_z^k < 0$. Como $t^k > 0$, conclui-se que $z^k > z^{k+1}$. Desde que $(v^k, z^k) \in \text{int}(\text{epi}(-H))$ para todo k , temos que $z^k > -H_k$. Para completar a prova, use (6.18) para escrever

$$t^k \geq \left[\frac{z^k + \hat{H}}{-\nabla \psi(v^k, z^k)^T d^k} \right] \geq \frac{z^k + H_k}{-d_z^k} \quad (*)$$

onde usa-se o fato que $\hat{H} \geq H_k$. Reorganizando a expressão (*) obtemos

$$-H_k > z^k + t^k d_z^k = z^{k+1}.$$

Assim

$$z^k > -H_k > z^{k+1} > -H_{k+1}$$

onde a última desigualdade segue-se do fato que $(v^{k+1}, z^{k+1}) \in \text{int}(\text{epi}(-H))$. \square

O seguinte resultado mostra que a sequência dual gerada pelo algoritmo IED converge à solução dual.

Teorema 6.3. *Se a sequência $\{p^k\}_{k \in \mathbb{N}} \in \text{int}(\text{epi}(-H))$ gerada pelo IED é infinito, então os seguintes resultados valem:*

1. As sequências $\{-H_k\}_{k \in \mathbb{N}}$ e $\{z^k\}_{k \in \mathbb{N}}$ são decrescentes e converge a \bar{H} .
2. $\hat{H} = \bar{H}$.
3. $\lim_{k \rightarrow \infty} (v^k)^T s^k = 0$.

Demonstração. Pelo lema 6.5 sabemos que $\{-H_k\}_{k \in \mathbb{N}}$ e $\{z^k\}_{k \in \mathbb{N}}$ são decrescentes. Como estas sequências são limitadas inferiormente por $-\bar{H}$, então elas convergem. Elas também têm o mesmo limite L uma vez que

$$0 \leq z^k + H_k < z^k - z^{k+1} = -t^k d_z^k,$$

e o lado direito tende a zero. Como ambas são limitados inferiormente por $-\bar{H}$, seus limites

comuns L verificam $L \geq -\bar{H}$. Usando agora (6.18) temos

$$0 = \lim_{k \rightarrow \infty} -d_z^k t^k \geq \lim_{k \rightarrow \infty} z^k + \hat{H} \geq \lim_{k \rightarrow \infty} z^k + H_k = 0,$$

o que implica $L = -\hat{H}$. De outro modo, obtemos $-\hat{H} = L \geq -\bar{H}$. Uma vez que $-\hat{H} \leq -\bar{H}$, conclui-se que $\hat{H} = \bar{H}$. Isto prova as duas primeiras afirmações. Use agora (6.18) para escrever

$$0 = \lim_{k \rightarrow \infty} -d_z^k t^k \geq \lim_{k \rightarrow \infty} |(\mathbf{v}^k)^T s^k| \geq 0,$$

o que conclui a prova. \square

Antes de mostrar que todo ponto de acumulação \bar{x} da sequência $\{x^k\}$ gerada pelo algoritmo IED é viável mostraremos o seguinte lema.

Lema 6.6. *Seja $\{\xi^k\}_{k \in \mathbb{N}} \subset (0, 1)$ a sequência gerada pelo Passo k.1 com $\rho^k = (1 - \xi^k)\rho_2^k$. Assuma também que a sequência $\{\xi^k\}_{k \in \mathbb{N}}$ é dada como no comentário 6.1. Se $\lim_{k \rightarrow \infty} \xi^k = 0$, então $\lim_{k \rightarrow \infty} s^k = 0$.*

Demonstração. Do Comentário 6.1, temos que

$$\xi^k > \frac{1}{3} \left(1 - \frac{\rho_1^k}{\rho_2^k} \right) > 0.$$

Portanto, se $\lim_{k \rightarrow \infty} \xi^k = 0$, então $\lim_{k \rightarrow \infty} \frac{\rho_1^k}{\rho_2^k} - 1 = 0$.

Sabemos que

$$\rho_1^k = \frac{1}{a_k} \left(\frac{1}{\lambda_\alpha^k \theta^k} - 1 \right) \text{ e } \rho_2^k = \frac{1}{a_k} \left(\frac{1}{\lambda_\alpha^k} - 1 \right).$$

Seja

$$\begin{aligned} \bar{\rho}^k &= \frac{\rho_1^k}{\rho_2^k} - 1 = \frac{\frac{1}{a_k} \left(\frac{1}{\lambda_\alpha^k \theta^k} - 1 \right)}{\frac{1}{a_k} \left(\frac{1}{\lambda_\alpha^k} - 1 \right)} - 1 \\ &= \frac{\frac{1}{\theta^k} - 1}{1 - \lambda_\alpha^k}. \end{aligned}$$

Obtemos $\theta^k = \frac{1}{1 + \bar{\rho}^k(1 - \lambda_\alpha^k)}$. Como $1 - \lambda_\alpha^k \in (0, 1)$ e $\bar{\rho}^k \rightarrow 0$ temos que

$$\lim_{k \rightarrow \infty} \theta^k = 1.$$

Então $\lim_{k \rightarrow \infty} \{1 + (s^k)^T (C^k)^{-1} s^k = 1\}$ o que implica

$$\lim_{k \rightarrow \infty} (s^k)^T (C^k)^{-1} s^k = 0.$$

Pela Hipótese 6.2 temos que $\lim_{k \rightarrow \infty} (s^k) = 0$. □

A proposição seguinte mostra que todo ponto de acumulação \bar{x} da sequência $\{x^k\}$ gerada pelo algoritmo IED é viável.

Proposição 6.2. *Assuma que IED realiza infinitos passos sérios. Seja \bar{x} um ponto de acumulação da sequência $\{x^k\}_{k \in \mathbb{N}}$ gerada pelo algoritmo. Então $f(\bar{x}) = 0$.*

Demonstração. Se IED realiza infinitos passos sérios então temos que $f(x^k) \neq 0$, para todo $k \in \mathbb{N}$. Com a notação do Lema 6.5, temos que

$$z^{k+1} - z^k = t^k d_z^k < 0 \tag{6.19}$$

Pelo Lema 6.5 a sequência $\{z^k\}_{k \in \mathbb{N}}$ converge. Tomando limites em ambos lados da (6.19), obtemos $\lim_{k \rightarrow \infty} t^k d_z^k = 0$. A definição de t^k no Passo k.1 implica que $t^k \geq \eta_0$ e, portanto, conclui-se que $\lim_{k \rightarrow \infty} d_z^k = 0$. Por definição de d^k , temos

$$d_z^k = d_{\alpha z}^k + \rho^k d_{\beta z}^k,$$

onde $\rho^k = (1 - \xi^k) \rho_2^k = (1 - \xi^k) \left[\frac{d_{\alpha z}^k}{-d_{\beta z}^k} \right]$. Combinando as duas últimas expressões obtemos

$$d_z^k = d_{\alpha z}^k + \rho^k d_{\beta z}^k = d_{\alpha z}^k + (1 - \xi^k) \left[\frac{d_{\alpha z}^k}{-d_{\beta z}^k} d_{\beta z}^k \right] = \xi^k d_{\alpha z}^k.$$

Isto implica que $\lim_{k \rightarrow \infty} d_z^k = \lim_{k \rightarrow \infty} \xi^k d_{\alpha z}^k = 0$. Como $\{\xi^k\}$ é limitada então conclui-se da última expressão que $\{d_{\alpha z}^k\}$ é limitada. Portanto existem subsequências $\{\xi^{k_j}\}$ e $\{d_{\alpha z}^{k_j}\}$ tais que $\lim_{k_j \rightarrow \infty} \xi^{k_j} = \xi$ e $\lim_{k_j \rightarrow \infty} d_{\alpha z}^{k_j} = d_{\alpha z}$. Então obtemos que $\xi = 0$ ou $d_{\alpha z} = 0$. Considere agora dois casos:

Caso 1: $\lim_{k_j \rightarrow \infty} \xi^{k_j} = 0$. Neste caso, pelo lema 6.6 obtemos que

$$\lim_{k_j \rightarrow \infty} s^{k_j} = \lim_{k_j \rightarrow \infty} (A f(x^{k_j}), -\sigma(f(x^{k_j}))) = 0.$$

Então tem-se $\lim_{k_j \rightarrow \infty} \sigma(f(x^{k_j})) = 0$ e pela continuidade de σ temos $\sigma(\lim_{k_j \rightarrow \infty} f(x^{k_j})) = 0$, o

que implica $\lim_{k_j \rightarrow \infty} f(x^{k_j}) = 0$ e portanto o ponto de acumulação \bar{x} verifica $f(\bar{x}) = 0$.

Caso 2: $\lim_{k_j \rightarrow \infty} d_{\alpha z}^{k_j} = 0$. Usando agora lema 6.1 obtemos

$$d_{\alpha z}^{k_j} \leq -(d_{\alpha}^{k_j})^T B^{k_j} d_{\alpha}^{k_j} \leq 0.$$

Como o lado esquerdo tende a zero, então $\lim_{k_j \rightarrow \infty} (d_{\alpha}^{k_j})^T B^{k_j} d_{\alpha}^{k_j} = 0$ e pela Hipótese 6.2 temos

$$\lim_{k_j \rightarrow \infty} d_{\alpha}^{k_j} = 0. \quad (6.20)$$

Usando também a primeira igualdade da equação (6.1) temos

$$\mathbb{R}^{n+1} \times \mathbb{R} \ni 0 = \lim_{k_j \rightarrow \infty} B^{k_j} d_{\alpha}^{k_j} = \lim_{k_j \rightarrow \infty} \left[-\lambda_{\alpha}^{k_j} s^{k_j}, \lambda_{\alpha}^{k_j} \right]^T + [0, -1]^T,$$

o que implica que $\lim_{k_j \rightarrow \infty} \lambda_{\alpha}^{k_j} s^{k_j} = 0$ e $\lim_{k_j \rightarrow \infty} \lambda_{\alpha}^{k_j} = 1$ e então obtemos que sequência $\{s^{k_j}\}$ cumpre $\lim_{k_j \rightarrow \infty} s^{k_j} = 0$. Então, de acordo com o caso 1, conclui-se que $f(\bar{x}) = 0$.

Antes de estabelecer o resultado final da convergência do método, precisamos de um lema que caracteriza a otimalidade primal.

□

Lema 6.7. *Assuma que IED realiza infinitos passos sérios. Seja \bar{x} um ponto de acumulação da sequência $\{x^k\}_{k \in \mathbb{N}}$ gerada pelo algoritmo IED. Então os seguintes itens são equivalentes:*

(a) $\lim_{k \rightarrow \infty} (s^k)^T v^k = 0.$

(b) \bar{x} é uma solução primal.

Demonstração. O fato de $x^k \in X(v^k)$ implica que

$$\begin{aligned} H(v^k) &= f_0(x^k) + c^k \sigma(f(x^k)) - \langle Af(x^k), u^k \rangle \\ &= f_0(x^k) + (s^k)^T v^k. \end{aligned} \quad (6.21)$$

Assuma que (a) valha. Tomando limites na igualdade acima e usando o fato de que no caso de passos sérios devemos ter $\bar{H} = \lim_{k \rightarrow \infty} H(v^k)$, obtemos

$$\bar{H} = f_0(\bar{x})$$

Uma vez que \bar{x} é viável pela Proposição 6.2, a igualdade acima implica que \bar{x} é uma solução

primal. Para a prova inversa, assumamos que \bar{x} é solução primal. Neste caso devemos ter

$$0 = f_0(\bar{x}) - \bar{H} = \lim_{k \rightarrow \infty} f_0(x^k) - H(v^k) = \lim_{k \rightarrow \infty} (s^k)^T v^k,$$

o qual implica em (a). \square

O seguinte teorema mostra que todo ponto de acumulação da sequência $\{x^k\}$ gerada pelo algoritmo IED é uma solução do problema (P).

Teorema 6.4. *Assuma que IED realiza infinitos passos sérios. Seja \bar{x} um ponto de acumulação da sequência $\{x^k\}_{k \in \mathbb{N}}$ gerada pelo algoritmo IED. Então \bar{x} resolve (P).*

Demonstração. A demonstração segue-se do Lema (6.7) e do Teorema 6.3. \square

7 *RESULTADOS NUMÉRICOS*

O algoritmo proposto foi implementado em Matlab 7.10.0, versão (R2010a) e usando um microcomputador Intel(R) core(TM) i7 de 3.4 GHz com 8.00 GB de RAM. Para verificar a confiabilidade do método resolvem-se alguns problemas extraídos de [11], os quais são amplamente utilizados para testar novos algoritmos. Para a solução do subproblema originado pela dualidade, emprega-se as rotinas `fminsearch` do Matlab que se baseia no método Nelder-Mead e também o algoritmo `SolvOpt` desenvolvido por Schor [12] para problemas não-diferenciáveis. No final deste capítulo apresentaremos o Kissing Number Problem e os resultados obtidos pelo IED e também por outros métodos já existentes.

7.1 PROBLEMAS TESTE

PQR-T1-7

Dimensão	3
Função objetivo	$f = -x_1x_2x_3$
Restrições	$g_1 = x_1^2 + 2x_2^2 + 4x_3^2 - 48$ $h_1(x) = \max(g_1, 0)$ $h_2(x) = \max(x_1 - 5, 0)$ $h_3(x) = \max(-x_1, 0)$ $h_4(x) = \max(x_2 - 5, 0)$ $h_5(x) = \max(-x_2, 0)$ $h_6(x) = \max(x_3 - 5, 0)$ $h_7(x) = \max(-x_3, 0)$
Ponto inicial	$x = -5[1 \ 1 \ 1]$
Ponto ótimo	$x^* = [4 \ 2.8284 \ 2]$

SQR-P1-1

Dimensão	2
Função objetivo	$a = [8 \ 8 \ 10 \ 10 \ 10 \ 10 \ 12 \ 12 \ 12 \ 12 \ 14 \ 14 \ 14 \ 16 \ 16 \ 16 \ 18 \ 18 \ 20 \ 20 \ 20 \ 22$ $22 \ 22 \ 24 \ 24 \ 24 \ 26 \ 26 \ 26 \ 28 \ 28 \ 30 \ 30 \ 30 \ 32 \ 32 \ 34 \ 36 \ 36 \ 38 \ 38 \ 40 \ 42]$ $b = [.49 \ .49 \ .48 \ .47 \ .48 \ .47 \ .46 \ .46 \ .45 \ .43 \ .45 \ .43 \ .43 \ .44 \ .43 \ .43 \ .46 \ .45$ $.42 \ .42 \ .43 \ .41 \ .41 \ .40 \ .42 \ .40 \ .40 \ .41 \ .40 \ .41 \ .41 \ .40 \ .40 \ 0.40 \ .38$ $.41 \ .40 \ .40 \ .41 \ .38 \ .40 \ .40 \ .39 \ .39]$ $f(x) = \sum_{i=1}^{44} (b_i - x_1 - (0.49 - x_1) \exp(-x_2(a_i - 8)))^2$
Restrições	$f_1(x) = \max(-0.49x_2 + x_1x_2 + 0.09, 0)$ $f_2(x) = \max(0.4 - x_1, 0)$ $f_3(x) = \max(-4 - x_2, 0)$
Ponto inicial	$x_0 = [0 \ 0]$
Ponto ótimo	$x^* = [0.41995 \ 1.28485]$

GQR-P1-1

Dimensão	2
Função objetivo	$f(x) = -75.196 + 3.8112x_1 + 0.0020567x_1^3 - 1.0345e - 5x_1^4 + 6.8306x_2$ $-0.030234x_1x_2 + 1.28134e - 3x_2x_1^2 + 2.266e - 7x_1^4x_2 - 0.25645x_2^2$ $+0.0034604x_2^3 - 1.3514e - 5x_2^4 + 28.106/(x_2 + 1) + 5.2375e - 6x_1^2x_2^2$ $+6.3e - 8x_1^3x_2^2 - 7e - 10x_1^3x_2^3 - 3.405e - 4x_1x_2^2 + 1.6638e - 6x_1x_2^3$ $+2.8673\exp(0.0005x_1x_2) - 3.5256e - 5x_1^3x_2 - 0.12694x_1^2$
Restrições	$f_1 = \max(-x_1x_2 + 700, 0)$ $f_2 = \max(-x_2 + x_1^2/125, 0)$ $f_3 = \max(-(x_2 - 50)^2 + 5(x_1 - 55), 0)$ $f_4 = \max(-x_1, 0)$ $f_5 = \max(x_1 - 75, 0)$ $f_6 = \max(-x_2, 0)$ $f - 7 = \max(x_2 - 65, 0)$
Ponto inicial	$x_0 = [90 \ 10]$
Ponto ótimo	$x^* = [13.5501 \ 51.66018]$

QQR-P1-1

Dimensão	3
Função objetivo	$f(x) = 4x_1^2 + 2x_2^2 + 2x_3^2 - 33x_1 + 16x_2 - 24x_3$
Restrições	$f_1(x) = 3x_1 - 2x_2^2 - 7$ $f_2(x) = 4x_1 - x_3^2 - 11$
Ponto inicial	$x_0 = [0 \ 0 \ 0]$
Ponto ótimo	$x^* = [5.32677 \ -2.11899 \ 3.21046]$

GLR-P1-1

Dimensão	3
Função objetivo	$f(x) = -32.174(255 \log((x_1 + x_2 + x_3 + 0.03)/(0.09x_1 + x_2 + x_3 + 0.03)))$ $+280 \log((x_2 + x_3 + 0.03)/(0.07x_2 + x_3 + 0.03))$ $+290 \log((x_3 + 0.03)/(0.13x_3 + 0.03))$
Restrições	$f_1(x) = \text{abs}(x_1 + x_2 + x_3 - 1)$ $f_2(x) = \max(\max([x_1 - 1 \ -x_1]), 0)$ $f_3(x) = \max(\max([x_2 - 1 \ -x_2]), 0)$ $f_4(x) = \max(\max([x_3 - 1 \ -x_3]), 0)$
Ponto inicial	$x_0 = [0.7 \ 0.2 \ 0.1]$
Ponto ótimo	$x^* = [0.617813 \ 0.328202 \ 0.053985]$

QQR-P1-2

Dimensão	3
Função objetivo	$f(x) = 1000 - x_1^2 - 2x_2^2 - x_3^2 - x_1x_2 - x_1x_3$
Restrições	$f_1(x) = 8x_1 + 14x_2 + 7x_3 - 56$ $f_2(x) = x_1^2 + x_2^2 + x_3^2 - 25$ $f_3(x) = \max(-x_1, 0)$ $f_4(x) = \max(-x_2, 0)$ $f_5(x) = \max(-x_3, 0)$
Ponto inicial	$x_0 = 2[1 \ 1 \ 1]$
Ponto ótimo	$x^* = [3.51212 \ 0.21699 \ 3.55217]$

PPR-P1-2

Dimensão	3
Função objetivo	$f(x) = 5x_1 + 50000/x_1 + 20x_2 + 72000/x_2 + 10x_3 + 144000/x_3$
Restrições	$f_1(x) = \max(-1 + 4/x_1 + 32/x_2 + 120/x_3, 0)$ $f_2(x) = \max(1e - 5 - x_1, 0)$ $f_3(x) = \max(1e - 5 - x_2, 0)$ $f_4(x) = \max(1e - 5 - x_3, 0);$
Ponto inicial	$x_0 = [1 \ 1 \ 1]$
Ponto ótimo	$x^* = [108.73472 \ 85.12614 \ 204.32471]$

LGR-P1-1

Dimensão	3
Função objetivo	$f(x) = 0.2x_3 - 0.8x_1$
Restrições	$f_1(x) = \max(\exp(x_1) - x_2, 0)$ $f_2(x) = \max(\exp(x_2) - x_3, 0)$ $f_3(x) = \max(x_1 - 100, 0)$ $f_4(x) = \max(-x_1, 0)$ $f_5(x) = \max(x_2 - 100, 0)$ $f_6(x) = \max(-x_2, 0)$ $f_7(x) = \max(x_3 - 10, 0)$ $f_8(x) = \max(-x_3, 0)$
Ponto inicial	$x_0 = [1 \ 1 \ 1]$
Ponto ótimo	$x^* = [0.18412 \ 1.20216 \ 3.32732]$

GQR-T1-5

Dimensão	3
Função objetivo	$f(x) = \log(x_3) - x_2$
Restrições	$f_1(x) = x_2^2 + x_3^2 - 4$ $f_2(x) = x_3 - 1 - x_1^2$
Ponto inicial	$x_0 = [1 \ 1 \ 1]$
Ponto ótimo	$x^* = [0 \ 1.7321 \ 1]$

PPR-P1-3

Dimensão	4
Função objetivo	$f(x) = x_1x_4(x_1 + x_2 + x_3) + x_3$
Restrições	$g(x) = 25 - x_1x_2x_3x_4$ $f_1(x) = \max(g, 0)$ $f_2(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 - 40$ $f_3(x) = \max(1 - x_1, 0)$ $f_4(x) = \max(x_1 - 5, 0)$ $f_5(x) = \max(1 - x_2, 0)$ $f_6(x) = \max(x_2 - 5, 0)$ $f_7(x) = \max(1 - x_3, 0)$ $f_8(x) = \max(x_3 - 5, 0)$ $f_9(x) = \max(1 - x_4(x), 0)$ $f_{10}(x) = \max(x_4 - 5, 0)$
Ponto inicial	$x_0 = 5[1 \ 1 \ 1 \ 1]$
Ponto ótimo	$x^* = [1 \ 4.7430 \ 3.8211 \ 1.3794]$

LPR-P1-1

Dimensão	4
Função objetivo	$f(x) = 1 + x_1 + x_2 + x_3 + x_4$
Restrições	$f_1(x) = \max(4/x_1 + 2.25/x_2 + 1/x_3 + 0.25/x_4 - 0.0401, 0)$ $f_2(x) = \max(0.16/x_1 + 0.36/x_2 + 0.64/x_3 + 0.64/x_4 - 0.010085, 0)$ $f_3(x) = \max(x_1 - 4e5, 0)$ $f_4(x) = \max(0.001 - x_1, 0)$ $f_5(x) = \max(x_2 - 3e5, 0)$ $f_6(x) = \max(0.001 - x_2, 0)$ $f_7(x) = \max(x_3 - 2e5, 0)$ $f_8(x) = \max(0.001 - x_3, 0)$ $f_9(x) = \max(x_4 - 1e5, 0)$ $f_{10} = \max(0.001 - x_4, 0)$
Ponto inicial	$x_0 = [1 \ 1 \ 1 \ 1]$
Ponto ótimo	$x^* = [193.4071 \ 179.5475 \ 185.0186 \ 168.7062]$

PGR-P1-3

Dimensão	5
Função objetivo	$f(x) = (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_3 - 1)^2 + (x_4 - 1)^4 + (x_5 - 1)^6$
Restrições	$f_1(x) = x_1^2 x_4 + \sin(x_4 - x_5) - 2\sqrt{2}$ $f_2(x) = x_2 + x_3^4 x_4^2 - 8 - \sqrt{2}$
Ponto inicial	$x_0 = [0 \ 0 \ 0 \ 0 \ 0]$
Ponto ótimo	$x^* = [1.1662 \ 1.1821 \ 1.3803 \ 1.5060 \ 0.6109]$

7.1.1 Resultados Obtidos

A tabela 7.1 fornece os parâmetros usados pelo método DSG para resolver os problemas teste propostos aqui. Para o método IED, usamos a mesma aproximação inicial, x_0 , c_0 , e u_0 usado pelo DSG, e os parâmetros, $B = I$, $\lambda = 0.5$, $T = 10^{-6}$ e η_0 dado aleatoriamente, em cada problema.

Para resolver o subproblema nos passos k.2 e k.3.b do algoritmo IED usamos a rotina `fminsearch` do Matlab que se baseia no método Nelder-Mead e também o algoritmo de Shor conhecido como `SolvOpt`.

Em cada problema substituímos a restrição de desigualdade da forma

$$g(x) \leq 0,$$

pela restrição equivalente de igualdade não suave

$$\max\{g(x), 0\} = 0,$$

de tal forma que todo problema com restrições de desigualdade é transformado em um problema com restrição de igualdade.

Os problemas teste resolvidos pelo IED foram extraído de Hock e Swittkowski [11]. Na seção 7.2, apresentaremos várias formulações do bem conhecido Kissing Number Problem que também são resolvidos pelo IED como mostrado.

Procedemos os experimentos numéricos considerando os casos :*(i)* $A = I$, isto é, empregando uma lagrangeana aumentada, e *(ii)* $A = 0$, isto é, empregando apenas uma função de penalidade. Sob cada caso informamos o número de iterações e o número de vezes que a função dual foi calculada pelos métodos `fminsearch` e `SolvOpt`. Os casos em que uma solução nem sempre pode ser encontrada serão indicadas por um traço. Observaremos que os problemas

Problema	x_0	c_0	u_0	\hat{H}	\hat{c}	α	δ
PQR-T1-7	$-(5, 5, 5)$	1.5	$(1, 1, 1, 1, 1, 1, 1)$	0	10	1	0.1
SQR-P1-1	$(0, 0)$	0.2	$(0, 0, 0)$	2	5	1	0.1
GQR-P1-1	$(90, 10)$	1	$(1, 1, 1, 1, 1, 1, 1)$	0	5	1	0.1
QQR-P1-1	$(0, 0, 0)$	0.1	$(1, 1)$	0	5	1	0.1
GLR-P1-1	$(0, 0, 0)$	50	$(0, 0, 0, 0)$	-20000	10000	1	0.5
QQR-P1-2	$(2, 2, 2)$	0.1	$(1, 1, 1, 1, 1)$	1000	5	1	0.1
PPR-P1-2	$(1, 1, 1)$	0.1	$(1, 1, 1, 1)$	10000	200	1	0.1
LGR-P1-1	$(1, 1, 1)$	0.1	$(0, 0, 0, 0, 0, 0, 0, 0)$	1	2	1	0.1
PPR-P1-3	$(5, 5, 5, 5)$	0.5	$(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$	20	5	1	0.1
LPR-P1-1	$(1, 1, 1, 1)$	1	$(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$	1000	50000	1	0.1
PGR-P1-3	$(0, 0, 0, 0, 0)$	0.0001	$(0, 0)$	0.5	1	1	0.1
GQR-T1-5	$(1, 1, 1)$	1.5	$(0, 0)$	2	20	1	0.1

Tabela 7.1: Parâmetros e pontos iniciais usados pelo método DSG.

resolvidos pelo método SolvOpt quase sempre apresentam melhores resultados do que quando resolvidos pelo método fmisearch. Para os problemas PQR-T1-7, SQR-P1-1 e QQR-P1-1 (tabelas 7.2), QQR-P1-2 e LGR-P1-1 (tabela 7.3) e PRR-P1-3, LPR-P1-1 e GQR-T1-5 (tabela 7.4) observamos que o método SolvOpt apresenta melhores resultados enquanto que os problemas GLR-P1-1 e PPR-P1-2 não puderem ser resolvidos por este método.

A	$\sigma(\cdot)$	PQR-T1-7				SQR-P1-1				GQR-P1-1				QQR-P1-1			
		fmin		Solv		fmin		Solv		fmin		Solv		fmin		Solv	
		n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L
I	$\ \cdot\ _2$	-	-	-	-	1	668	1	463	1	775	-	-	1	1103	2	808
	$\ \cdot\ _1$	-	-	-	-	1	762	1	527	-	-	-	-	1	1162	2	887
	$\sqrt{m}\ \cdot\ _\infty$	4	7861	2	2521	1	643	1	480	1	334	-	-	1	1193	2	800
	$e^{\ \cdot\ _2} - 1$	1	1764	1	501	1	638	1	447	1	775	1	437	1	1311	2	813
	$e^{\ \cdot\ _1} - 1$	1	4774	1	480	1	702	1	517	1	775	-	-	1	1429	2	767
	$\sqrt{m}(e^{\ \cdot\ _\infty} - 1)$	1	446	1	232	1	639	1	468	1	486	1	879	2	2035	2	776
$\max(\ \cdot\ _2, \ \cdot\ _2^2)$	3	2664	1	559	1	668	1	463	1	1084	1	400	1	1084	1	468	
0	$\ \cdot\ _2$	7	4526	6	1655	1	660	1	468	1	577	-	-	4	2605	5	1563
	$\ \cdot\ _1$	3	4387	-	-	1	762	1	504	1	450	-	-	4	3049	5	1581
	$\sqrt{m}\ \cdot\ _\infty$	2	3300	2	1877	1	641	1	475	1	453	-	-	4	3347	5	1712
	$e^{\ \cdot\ _2} - 1$	1	645	1	233	1	683	1	447	1	436	1	197	2	1631	3	1027
	$e^{\ \cdot\ _1} - 1$	1	473	1	231	1	804	1	474	1	436	-	-	2	2726	3	993
	$\sqrt{m}(e^{\ \cdot\ _\infty} - 1)$	1	571	1	323	1	647	1	474	1	438	1	348	2	2101	3	1188
$\max(\ \cdot\ _2, \ \cdot\ _2^2)$	1	653	1	399	1	660	1	468	1	445	1	195	2	1548	3	988	

Tabela 7.2: n_{its} =número de iterações, n_L =número de vezes que a função lagrangeana é calculada

A	$\sigma(\cdot)$	GLR-P1-1				QQR-P1-2				PPR-P1-2				LGR-P1-1			
		fmin		Solv		fmin		Solv		fmin		Solv		fmin		Solv	
		n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L
I	$\ \cdot\ _2$	3	3129	-	-	-	-	-	-	4	2859	-	-	-	-	-	-
	$\ \cdot\ _1$	2	1849	-	-	-	-	-	-	4	2859	-	-	5	4132	5	2189
	$\sqrt{m}\ \cdot\ _\infty$	2	2165	-	-	-	-	-	-	4	2726	-	-	6	7428	3	1440
	$e^{\ \cdot\ _2} - 1$	2	2394	-	-	7	4564	4	1193	-	-	-	-	5	2953	5	1495
	$e^{\ \cdot\ _1} - 1$	2	1550	-	-	17	6550	4	1797	-	-	-	-	5	4793	5	1789
	$\sqrt{m}(e^{\ \cdot\ _\infty} - 1)$	14	5500	-	-	3	2894	4	1572	-	-	-	-	3	3318	3	1306
	$\max(\ \cdot\ _2, \ \cdot\ _2^2)$	7	4564	-	-	4	2568	4	1297	4	2649	-	-	5	3466	6	2125
0	$\ \cdot\ _2$	3	2753	-	-	-	-	-	-	5	3347	-	-	-	-	-	-
	$\ \cdot\ _1$	2	1813	-	-	-	-	-	-	5	3347	-	-	-	-	-	-
	$\sqrt{m}\ \cdot\ _\infty$	3	2218	-	-	-	-	-	-	5	3370	-	-	3	33841	3	1473
	$e^{\ \cdot\ _2} - 1$	3	2433	-	-	2	1475	4	1895	-	-	-	-	6	3432	6	1825
	$e^{\ \cdot\ _1} - 1$	2	1644	-	-	2	2285	4	1986	-	-	-	-	6	5106	6	2340
	$\sqrt{m}(e^{\ \cdot\ _\infty} - 1)$	3	2252	-	-	2	2637	4	2362	-	-	-	-	3	3192	3	1296
	$\max(\ \cdot\ _2, \ \cdot\ _2^2)$	7	4564	-	-	3	2135	4	1538	5	3394	-	-	7	4599	7	2565

Tabela 7.3: n_{its} =número de iterações, n_L =número de vezes que a função lagrangeana é calculada

A	$\sigma(\cdot)$	PPR-P1-3				LPR-P1-1				PGR-P1-3				GQR-T1-5			
		fmin		Solv		fmin		Solv		fmin		Solv		fmin		Solv	
		n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L	n_{its}	n_L
I	$\ \cdot\ _2$	-	-	-	-	8	7621	8	2805	2	4838	8	6468	-	-	-	-
	$\ \cdot\ _1$	-	-	-	-	7	8107	-	-	3	7014	8	5706	-	-	-	-
	$\sqrt{m}\ \cdot\ _\infty$	-	-	-	-	6	7794	-	-	3	6671	8	5702	1	726	-	-
	$e^{\ \cdot\ _2} - 1$	5	4897	6	2556	8	6888	-	-	2	4500	8	6282	6	1387	3	1182
	$e^{\ \cdot\ _1} - 1$	4	8824	4	2772	8	7828	-	-	3	7630	8	7216	2	1422	3	1184
	$\sqrt{m}(e^{\ \cdot\ _\infty} - 1)$	5	22156	4	6370	6	7762	6	3131	3	5972	8	5870	1	934	2	757
	$\max(\ \cdot\ _2, \ \cdot\ _2^2)$	6	7090	7	3270	8	7479	-	-	2	4546	8	5256	-	-	-	-
0	$\ \cdot\ _2$	-	-	-	-	1	660	11	3920	3	4462	8	6366	-	-	-	-
	$\ \cdot\ _1$	-	-	-	-	9	10061	11	4273	3	6128	3	2132	-	-	-	-
	$\sqrt{m}\ \cdot\ _\infty$	-	-	3	4297	7	9329	-	-	3	8151	8	6733	1	726	-	-
	$e^{\ \cdot\ _2} - 1$	2	2090	4	7669	11	8837	-	-	2	3791	8	5969	8	1692	3	962
	$e^{\ \cdot\ _1} - 1$	2	3819	3	2384	9	9773	-	-	2	5578	8	8789	2	1433	3	1408
	$\sqrt{m}(e^{\ \cdot\ _\infty} - 1)$	3	10319	3	3328	7	16579	6	3091	2	4817	8	8314	1	939	2	888
	$\max(\ \cdot\ _2, \ \cdot\ _2^2)$	2	4364	4	3573	11	9655	-	-	3	4183	8	5250	-	-	-	-

Tabela 7.4: n_{its} =número de iterações, n_L =número de vezes que a função lagrangeana é calculada

n	k_n
1	2
2	6
3	12
4	24
5	40-44
6	72-78
7	126-134
8	240
9	306-364
10	500-554
11	582-870
12	840-1357

Tabela 7.5: Alguns valores conhecidos e limitantes para k_n

7.2 APLICAÇÃO

7.2.1 Kissing Number Problem

O Kissing Number Problem é definido como o número de esferas de raio r que tocam outra esfera central dada de mesmo raio, sem interseção. O Kissing Number Problem tenta encontrar o número máximo k_n de esferas de raio $r \in \mathbb{R}_+$ que simultaneamente tocam a uma esfera central sem interseções.

O Kissing Number Problem pode ser formulado como um problema de otimização não convexo, onde a distância mínima entre pares de centros das esferas que tocam a esfera central em \mathbb{R}^n será maximizado.

Assim, uma formulação possível é:

$$(P1) \quad \begin{cases} \max \min_{i \neq j} \|y^i - y^j\| \\ \|y^k\| = 1, k = 1, \dots, p \end{cases}$$

Aqui, $y^k \in \mathbb{R}^n$ e, sem perda de generalidade, o raio das esferas é tomado igual a $1/2$. O menor valor de p para o qual o valor ótimo do problema (P1) é menor ou igual a 1 é nada mais que k_n . O problema (P1) tem na prática um número muito grande de pontos estacionários, o que faz com que a tarefa de encontrar k_n seja muito complicada mesmo para problemas com pequenas dimensões.

Tabela 7.5 fornece valores já conhecidos para algumas instâncias do Kissing Number Problem e também alguns limitantes para alguns caso em que não se tem ainda o valor exato de k_n .

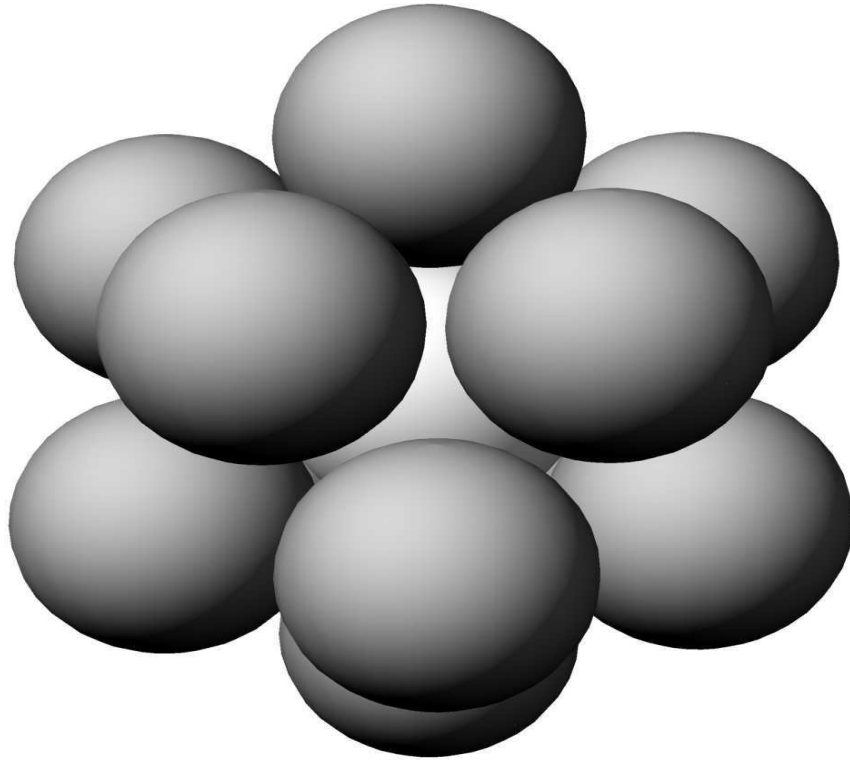


Figura 7.1: Kissing Number Problem para dimensão 3

Resolvemos o problema ($P1$) usando o método IED e comparamos os resultados obtidos com aqueles obtidos pelos métodos AUGPEN, DSG, Ipopt, ALBOX e LANCELOT já existentes e amplamente testados.

Os resultados numéricos são visualizados na tabela 7.6, que apresenta os valores mínimo, máximo e médio obtido por cada método bem como a dimensão dos problemas que foram resolvidos 50 vezes cada.

Em cada iteração do método IED escolhemos $c_0 = 0.5$ e as componentes do vetor $u_0 \in \mathbb{R}^p$ são geradas aleatoriamente entre 0 e 0.1. Ao invés de $\max_{i \neq j} (\|y^i - y^j\|)$ usamos $\min \left[-\min_{i \neq j} (\|y^i - y^j\|) \right]$. Além disso, os parâmetros $B = I$, $\eta_0 = 10^{-6}$ são usados em todos os problemas. Para os termos de penalidade da lagrangeana aumentada, escolhemos $\sigma(\cdot) = \exp(\|\cdot\|_1) - 1$ e como critério de parada usamos

$$\|(\|y^1\| - 1, \dots, \|y^p\| - 1)\|_1 \leq 10^{-6}.$$

Para resolver o subproblema empregamos SolvOpt [12].

Os resultados na tabela 7.6 mostram que o método IED é competitivo quando comparado a outros métodos já consagrados.

Dimensão do problema			Valores min, max e a médio			
$\begin{bmatrix} n \\ p \end{bmatrix}$	var.	restr.	Método	min.	max.	média
$\begin{bmatrix} 3 \\ 11 \end{bmatrix}$	33	11	IED	0.9759717	1.0514620	1.0385263
	33	11	AUGPEN	1.0513841	1.0514622	1.0514485
	33	11	DSG	1.0514112	1.0514622	1.0514497
	33	66	Ipopt	1.0514622	1.0514622	1.0514622
	89	66	ALBOX	1.0514622	1.0514622	1.0514622
	89	66	LANCELOT	1.0514656	1.0514656	1.0514656
$\begin{bmatrix} 3 \\ 12 \end{bmatrix}$	36	12	IED	0.8825956	1.0514614	1.0279810
	36	12	AUGPEN	1.0513831	1.0514622	1.0514451
	36	12	DSG	1.0513618	1.0514622	1.0514446
	37	78	Ipopt	1.0514622	1.0514622	1.0514622
	103	78	ALBOX	0.9463817	1.0514622	1.0451574
	103	78	LANCELOT	0.9447856	1.0514656	1.0430604
$\begin{bmatrix} 3 \\ 13 \end{bmatrix}$	39	13	IED	0.8677158	0.9564116	0.9374836
	39	13	AUGPEN	0.9442759	0.9535789	0.9482584
	39	13	DSG	0.9420563	0.9564074	0.9483538
	40	91	Ipopt	0.9443564	0.9564136	0.9508438
	118	91	ALBOX	0.9281797	0.9564136	0.9488794
	118	91	LANCELOT	0.9443516	0.9564099	0.9512710
$\begin{bmatrix} 4 \\ 23 \end{bmatrix}$	92	23	IED	0.9625933	0.9918578	0.9813914
	92	23	AUGPEN	0.9694194	0.9999848	0.9809890
	92	23	DSG	0.9695109	0.9918115	0.9792563
	93	276	Ipopt	0.9709733	1.0000000	0.9850461
	346	276	ALBOX	0.9723134	0.9999999	0.9862923
	346	276	LANCELOT	0.9740944	0.9918568	0.9847650
$\begin{bmatrix} 4 \\ 24 \end{bmatrix}$	96	24	IED	0.9574948	0.9999860	0.9697619
	96	24	AUGPEN	0.9539599	0.9999352	0.9724289
	96	24	DSG	0.9582749	0.9999896	0.9723225
	97	300	Ipopt	0.9635972	1.0000000	0.9773164
	373	300	ALBOX	0.9630174	0.9999999	0.9769849
	373	300	LANCELOT	0.9580083	0.9828733	0.9751985
$\begin{bmatrix} 4 \\ 25 \end{bmatrix}$	100	25	IED	0.8321989	0.9605343	0.9495874
	100	25	AUGPEN	0.9483343	0.9613067	0.9558045
	100	25	DSG	0.9467224	0.9616356	0.9458296
	101	325	Ipopt	0.9511129	0.9619604	0.9593897
	401	325	ALBOX	0.9529038	0.9619429	0.9580977
	401	325	LANCELOT	0.9465833	0.9619563	0.9574963
$\begin{bmatrix} 5 \\ 39 \end{bmatrix}$	195	39	IED	0.9673393	0.9885168	0.9781089
	195	39	AUGPEN	0.9702280	0.9877491	0.9796894
	195	39	DSG	0.9698423	0.987433	0.9805071
	937	780	Ipopt	0.9826355	0.9949991	0.9888488
	937	780	ALBOX	0.9758879	0.9926700	0.9879840
	937	780	LANCELOT	0.9808159	0.9920876	0.9881179
$\begin{bmatrix} 5 \\ 40 \end{bmatrix}$	200	40	IED	0.9605262	0.9819962	0.9702498
	200	40	AUGPEN	0.9595603	0.9827173	0.9740579
	200	40	DSG	0.9640701	0.9834953	0.9747307
	201	820	Ipopt	0.9726550	0.9923987	0.9842284
	981	820	ALBOX	0.9750429	0.9883235	0.9817903
	981	820	LANCELOT	0.9701958	0.9920282	0.9810864

Tabela 7.6: IED frente a AUGPEN, DSG, Ipopt, ALBOX, LANCELOT

8 *CONCLUSÕES*

Neste trabalho apresentamos o algoritmo IED desenvolvido para minimização de funções não-convexas e não-diferenciáveis sujeitas à restrições de igualdade e/ou desigualdade. O método utiliza dualidade lagrangeana cuja função lagrangeana possui a propriedade de dualidade forte, isto é, os valores ótimos primais e duais são iguais. Apresentamos também os resultados obtidos pelo algoritmo IED quando aplicado a alguns problemas teste extraídos da literatura bem como ao Kissing Number Problem, um problema desafiador que em certas instâncias, envolve um grande número de variáveis.

Foi possível ver que o algoritmo é eficiente, mas depende fortemente da solução do subproblema gerado pela dualidade. Isto consiste em uma fraqueza do método pois a solução de tal problema depende por sua vez de outros algoritmos pré-existentes.

Acreditamos que o algoritmo possa ser melhorado com emprego de técnicas eficientes para a solução do subproblema mencionado e também com a criação de regras para escolha da matriz B bem como escolhas apropriadas da matriz A e da função $\sigma(\cdot)$.

Estes são os objetivos para trabalhos futuros.

REFERÊNCIA

- [1] A. IZMAILOV, M. SOLODOV, **Condições de Otimilidade Elementos de Análisis Convexa e de Dualidade**, Otimização- Vol. 1. Rio de Janeiro. IMPA, 2005.
- [2] A. IZMAILOV, M. SOLODOV, **Métodos Computacionais**, Otimização- Vol. 2. Rio de Janeiro. IMPA, 2005.
- [3] BAZARAA, M. S., SHERALI, H. D., SHETTY, C. M., **Nonlinear Programming, Theory and Algorithms**, 3 ed, John Wiley and Sons, Inc, Hobeken, New Jersey, 2006.
- [4] BURACHIK, R. S., KAYA, C. Y., **O the absence of duality gap for Lagrangiano-type functions**, Journal of Industrial Management and Optimization, 33-38(2005).
- [5] BURACHIK, R. S., KAYA, C. Y., **An augmented penalty function method with penalty parameter updates for nonconvex optimization**. Nonlinear Analysis: Theory Meth.Appl. 75, 1158–1167, 2012.
- [6] BURACHIK, R. S., KAYA, C. Y., **A Deflected Subgradient Method Using a General Augmented Lagrangian Duality With Implications on Penalty Methods**. Variational Analysis and Generalized Differentiation in Optimization and Control Springer Optimization and Its Applications Volume 47, 2010, pp 109-132.
- [7] C. BELTRAN, F. J. HEREDIA, **An Effective Line Search for the Subgradient Method**, Polytechnical University of Catalonia, Barcelona, Spain. November 2004.
- [8] ELON, L. L., **Curso de Análisi vol. 2**. 11.ed. Rio de Janeiro, IMPA, 2010.
- [9] HERSKOVITS, J., **View on Nonlinear Optimization**, ed. Herskovits, J., Advances in Structural Optimization, Kluwer Academic Publishers, pp. 71-116, 1995.
- [10] HERSKOVITS, J., **A Feasible Directions Interior Point Technique for Nonlinear Optimization**. JOTA, Journal of Optimization Theory and Applications, vol. 99, pp. 121-146, 1998.
- [11] HOCK, W., SCHITTKOWSKI, K., **Test Examples for Nonlinear Programming Codes**. Lecture Notes in Economics and Mathematical Systems, Vol. 187, Springer, 1981.
- [12] KAPPEL, F., KUNTSEVICH A. V., **An implementation of Shor's r-algorithm**, Comput. Optim. Appl. 15, 193-205, 2000.
- [13] KREJIC, N., MARTÍNEZ, J. M., MELLO, M., PILOTTA, E. A., **Validation of an augmented Lagrangian algorithm with a Gauss-Newton Hessian approximation using a set of hard-spheres problems**, Comput. Optim. Appl. 16, 247-263, 2000.
- [14] R. M. FREUND, **Subgradient Optimization, Generalized Programming**. Massachusetts Institute of Technology, May, 2004.

- [15] LUENBERGER, D. G., **Linear and Nonlinear Programming**, 2 ed., Addison-Wesley Publishing Company, 1984.
- [16] LEMARÉCHAL, C., **Nondifferentiable Optimization**, in: handbooks in ORandMS, eds. Menhauser, G.L., et al., Elsevier Publishers, North-Holland, pp. 529-572, 1989.
- [17] LIMA A. L. D. S., **Uma Estratégia de Descomposição por Relaxação Lagrangeana para 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.
- [18] HERSKOVITS, J., FREIRE, W. P., TANAKA, M., CANALES, A., **A feasible directions method for nonsmooth convex optimization**. Structural and Multidisciplinary Optimization, v. 44 n. 3, p. 363-377, 2011.
- [19] JOSÉ, M. M., SANDRA, A. S., **Métodos Computacionais de Otimização**. Departamento de Matemática Aplicada.IMECC-UNICAMP 1995.
- [20] REGINA, S. B., RAFAIL, N. G., NERGIZ, A. I., C. YALÇIN K, **On a Modified Subgradient Algorithm for Dual Problems via Sharp Augmented Lagrangian**. Journal of Global Optimization (2006)34:55-78, Springer 2006. DOI 10.1007/s10898-005-3270-5.
- [21] W. P. FREIRE, R. S. BURACHIK, C. Y. KAYA, **Interior Epigraph Directions Method for Nonsmooth and Nonconvex Optimization via Generalized Augmented Lagrangian Duality**, Nonsmooth optimization May 26, 2012.
- [22] ROCKAFELLAR, R. T., ROGER, J-B. W., **Variational Analysis**, Springer Dordrecht Heidelberg London New York, 2009.
- [23] S. BOYD, A. MUTAPCIC, **Subgradient Methods**, Notes for EE364b, Stanford University, Winter 2006-07. January 23, 2007.