

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Programa de Pós-Graduação em Matemática

Daniel Rodrigues Pereira

Equivalência entre dois algoritmos de pontos interiores FDIPA e FDA-NCP

Juiz de Fora

2016

Daniel Rodrigues Pereira

Equivalência entre dois algoritmos de pontos interiores FDIPA e FDA-NCP

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal de Juiz de Fora, na área de concentração em Matemática Aplicada, como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador: Sandro Rodrigues Mazorche

Juiz de Fora

2016

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Rodrigues Pereira, Daniel.

Equivalência entre dois algoritmos de pontos interiores FDIPA e FDA-NCP / Daniel Rodrigues Pereira. – 2016.

51 f. : il.

Orientador: Sandro Rodrigues Mazorche

Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Matemática, 2016.

1. Problema de otimização. 2. Algoritmo FDIPA. 3. Problema de complementaridade. 4. Algoritmo FDA. I. Rodrigues Mazorche, Sandro. Equivalência entre dois algoritmos de pontos interiores FDIPA e FDA-NCP.

Daniel Rodrigues Pereira

Equivalência entre dois algoritmos de pontos interiores FDIPA e FDA-NCP

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal de Juiz de Fora, na área de concentração em Matemática Aplicada, como requisito parcial para obtenção do título de Mestre em Matemática.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Sandro Rodrigues Mazorche - Orientador
Universidade Federal de Juiz de Fora

Professor Dr. Grigori Chapiro
Universidade Federal de Juiz de Fora

Professor Dr. Dênis Emanuel da Costa Vargas
Instituto Federal Sudeste de Minas Gerais

AGRADECIMENTOS

Agradeço a todos que de forma relevante contribuíram para o desenvolvimento deste trabalho.

Aos meus pais, Albenir Pereira e Ludmila Coelho Rodrigues Pereira, legítimos responsáveis por minha realização.

Ao meu orientador, Prof. Dr. Sandro Rodrigues Mazorche, pela paciência e ajuda no desenvolvimento deste trabalho.

A todos mios estimados amigos de la maestria: Oscar, Vladimir, Julio Cesar, Pacheco, Alberth, Hugo, Jorge Luis, Santiago, Mario, Rossini Cachorro y Eduardo Tintas Medias. Viva Perú, tierra del Depredador. Todavía me faltan muchos por mencionar, mas pido me disculpen por no nombrarlos, a excepcion de Miguel Cutipa (matemático brillante que jamas sere) y Giovanna Arelis - abajo de Dios fueron ellos quienes me ayudaran.

Aos professores participantes da banca examinadora, Dr. Grigori Chapiro e Dr. Dênis Emanuel da Costa Vargas.

À CAPES, pelo apoio financeiro.

Envio aqui esta mensagem a todos os citados e espero que os encontre contentes, gozando saúde e felizes em seus amores.

RESUMO

Apresentamos neste trabalho o algoritmo de pontos interiores e direções viáveis denominado FDIPA para resolução de problemas de otimização definido por uma função diferenciável e por restrições de desigualdades. O algoritmo gera uma sequência de pontos interiores a partir de um dado ponto inicial também de interior e converge globalmente com ordem superlinear para um par Karush-Kuhn-Tucker do problema. A cada iteração uma direção de descida da função potencial é calculada inicialmente pela resolução de um sistema nas variáveis dual e primal. Apresentamos também o algoritmo FDA para resolução de problemas de complementaridade definido por uma função diferenciável e não linear. Mostramos a equivalência entre os dois métodos no sentido de gerarem as mesmas direções de descida, viável e de restauração a partir de uma atualização dos multiplicadores de Lagrange do problema de otimização. Realizamos uma comparação entre os métodos em uma coletânea de problemas de complementaridade.

Palavras-chave: Problema de otimização. Algoritmo FDIPA. Problema de complementaridade. Algoritmo FDA.

ABSTRACT

In this work we present the algorithm of internal points and viable directions denominated FDIPA to solve optimization problems defined by a differentiable function and by inequalities restrictions. The algorithm generates a sequence of interior points from a given interior starting point and converges globally with superlinear order to a Karush-Kuhn-Tucker pair of the problem. At each iteration a descent direction of the potential function is calculated initially by the solution of a system in the dual and primal variables. We also present the FDA algorithm to solve complementarity problems defined by a non-linear differentiable function. We show the equivalence between the two methods in the sense that they generate the same descent, feasible and restoring directions from an update to the Lagrange multipliers of the optimization problem. We perform a comparison between the two methods in a collection of complementarity problems.

Keywords: Optimization Problem. FDIPA algorithm. Complementarity problem. FDA algorithm.

LISTA DE ILUSTRAÇÕES

Figura 1 – Mínimos globais e locais	12
Figura 2 – Conjunto convexo	12
Figura 3 – Conjunto não convexo	12
Figura 4 – Direção de descida	15
Figura 5 – Direções viáveis	15
Figura 6 – Método de Newton	19
Figura 7 – Busca linear inexata de Armijo	21
Figura 8 – Busca linear inexata de Wolfe	22
Figura 9 – Algoritmo FDA	36
Figura 10 – Problema da Meia Lua	43
Figura 11 – Problema do Peixe	44

LISTA DE ABREVIATURAS E SIGLAS

FDIPA	Feasible Direct Interior Point Algorithm
FDA	Feasible Direct Algorithm
LCP	Problema de Complementaridade Linear
NCP	Problema de Complementaridade Não Linear
KKT	Karush-Kuhn-Tucker
BFGS	Broyden-Fletcher-Goldfarb-Shano

SUMÁRIO

1	INTRODUÇÃO	9
2	Conceitos e Resultados Preliminares	11
2.1	Problema de Minimização	11
2.1.1	Condições de Otimalidade	16
2.1.2	Método de Newton para equações	18
2.1.3	Buscas Lineares	20
2.2	Problema de Complementaridade	23
3	FDIPA	25
3.1	Descrição do Método FDIPA	26
3.1.1	Buscas Lineares no FDIPA	28
3.2	Convergência Global do algoritmo FDIPA	30
3.3	Atualizações para a matriz B_k	33
3.3.1	$B = I$	33
3.3.2	Atualização BFGS	33
3.3.3	Outras atualizações	34
4	FDA	35
4.1	Descrição do Método FDA	35
5	NOVA ATUALIZAÇÃO	38
5.1	Introdução	38
5.2	Reescrevendo o NCP como um problema de minimização	38
5.3	Resolução do NCP modificado via FDIPA	38
5.4	Uma nova atualização para λ	39
6	Testes Numéricos	42
7	Conclusão	49
7.1	Trabalhos Futuros	49
	REFERÊNCIAS	50

1 INTRODUÇÃO

Um problema de otimização (ou minimização) definido por uma região viável e uma função objetivo f consiste em encontrar um ponto desta região que possui o menor valor por f .

Problemas de otimização surgem, por exemplo, na engenharia de design quando o objetivo é encontrar o melhor design que um determinado objeto deve assumir de modo que minimize o custo. Problemas físicos também podem ser modelados como problemas de minimização - por exemplo, quando o objetivo é determinar o equilíbrio de uma dada função de energia.

Denotando $[x_1, x_2, \dots, x_n]$ como as variáveis de design, f a função objetivo (ou potencial) e Ω a região viável (ou factível), o problema de otimização pode ser formulado por

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && x \in \Omega. \end{aligned} \tag{1.1}$$

O problema é dito restritivo quando Ω é definido por uma classe de inequações e irrestrito se $\Omega = \mathbb{R}^n$.

Algoritmos de direções viáveis constituem uma importante classe de métodos para a resolução do problema de minimização restritivo. No FDIPA, em particular, a direção de busca é uma direção viável das restrições de desigualdade e também de descida da função objetivo.

O algoritmo é bastante eficiente no sentido de que pode ser interrompido sempre que um ponto da iteração estiver suficientemente próximo da solução pois o algoritmo gera a cada iteração um ponto dentro da região viável.

O algoritmo de pontos interiores e direções viáveis conhecido como FDIPA (*Feasible Direct Interior Point Algorithm*) foi inicialmente proposto por José Herskovits em 1982 para resolução do problema de minimização (3.1). Deriva do método de Newton aplicado às condições de Karush-Kuhn-Tucker do problema de minimização restritivo. As direções de descida e viável são obtidas de uma mesma matriz, onde em seguida uma busca linear inexata é implementada a fim de que um ponto dentro da região viável seja obtido de modo que proporcione uma redução significativa da função objetivo.

O problema de complementaridade não linear (NCP) foi inicialmente proposto por R. W. Cottle em sua dissertação de Ph.D. na data de 1968 como uma generalização do problema de complementaridade linear (LCP).

Condições de otimalidade de programas lineares, quadráticos, não lineares e problemas de equilíbrio podem ser modelados como problemas de complementaridade. Vários algoritmos para resolução deste problema (mediante métodos de ponto fixo, projeção,

homotopia, Newton, otimização linear e não linear, equações lineares e não lineares) tornaram-se populares a partir do fim da década de 70 e início dos anos 80.

Joseph [8] apresentou em 1979 um método de Newton generalizado para resolução de inequações variacionais onde o qual tem o NCP como caso particular. O método de Joseph-Newton possui convergência local com ordem quadrática.

Pang e Gabriel [17] combinaram o método de sistema de equações não diferenciáveis com programação quadrática para desenvolver o algoritmo conhecido como NE/SQP com convergência global e ordem quadrática.

Podemos notar assim que a proposta de resolução do NCP mediante métodos de sistemas de equações não é atual. Mangasarian [12, 20] propôs em 1976 a resolução via equações diferenciáveis mas a singularidade do sistema restringe o método somente a alguns casos específicos.

É possível ainda a reformulação do NCP através de um problema de otimização diferenciável sem restrições. Mangasarian e Solodov [11] utilizaram uma função diferenciável de tal modo que todo mínimo global do problema irrestrito seja uma solução do NCP. Yamashita e Fukushima [21] provaram que todo ponto estacionário do problema de otimização é uma solução do NCP se a função do problema é continuamente diferenciável e fortemente monótona em \mathbb{R}^n . A partir deste resultado ficou estabelecido que métodos para resolução de problemas de otimização podem ser aplicados aos problemas de complementaridade.

O método Feasible Direct Algorithm (FDA) proposto por Sandro R. Mazorche [14] não envolve inequações variacionais nem tampouco reformula o problema de complementaridade como um sistema de equações ou problema de otimização. Trata-se de um método direto de pontos interiores dado por uma modificação ao problema via método de Newton.

O Capítulo 2 contém as definições e resultados básicos acerca da teoria de otimização e complementaridade. O Capítulo 3 trata do algoritmo FDIPA para resolução do problema de otimização dado por restrições de desigualdade e o Capítulo 4 trata do algoritmo FDA para resolução do problema de complementaridade. É proposto no Capítulo 5 uma nova regra de atualização para o multiplicador de Lagrange do algoritmo FDIPA. O último capítulo contém uma coletânea de problemas de complementaridade onde é feita uma comparação entre os métodos FDIPA e FDA.

2 Conceitos e Resultados Preliminares

Apresentamos neste capítulo os conceitos e resultados relativos aos problemas de minimização e complementaridade necessários para o desenvolvimento deste trabalho - consoantes as principais referências da teoria fundamental de otimização como Bazaraa [1], Luenberger [10], Nocedal [15] e Herskovits [3].

2.1 Problema de Minimização

Em linhas gerais, otimização consiste em encontrar máximos e mínimos que uma função real de n variáveis assume em uma determinada região do \mathbb{R}^n . Formalizamos esse conceito no seguinte problema de minimização

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && x \in \Omega \subseteq \mathbb{R}^n. \end{aligned} \tag{2.1}$$

onde $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é chamada de *função objetivo* e Ω de *conjunto viável* ou *factível*.

O conjunto Ω é frequentemente definido por igualdades e/ou desigualdades, e.g., $\Omega = \{x \in \mathbb{R}^n; g(x) \leq 0 \text{ e } h(x) = 0, \text{ com } g: \mathbb{R}^n \rightarrow \mathbb{R}^m \text{ e } h: \mathbb{R}^n \rightarrow \mathbb{R}^p\}$.

Seja o problema de otimização não linear dado por

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && g_i(x) \leq 0; \quad i = 1, 2, \dots, m. \\ &&& h_i(x) = 0; \quad i = 1, 2, \dots, p. \end{aligned} \tag{2.2}$$

onde $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ e $h: \mathbb{R}^n \rightarrow \mathbb{R}^p$ são funções diferenciáveis e pelo menos uma destas é não linear.

Definição 2.1. *Uma restrição de desigualdade é dita ativa se $g_i(x) = 0$ e inativa se $g_i(x) < 0$.*

Denotando $g(x) = [g_1(x), g_2(x), \dots, g_m(x)]^t$ e $h(x) = [h_1(x), h_2(x), \dots, h_p(x)]^t$, o problema acima assume a seguinte forma

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && g(x) \leq 0 \\ &&& h(x) = 0. \end{aligned} \tag{2.3}$$

As duas definições a seguir são relativas ao problema de otimização (2.3).

Definição 2.2. *Um ponto $x^* \in \Omega$ é um Mínimo Global (ou Mínimo Absoluto) de f sobre Ω se $f(x) \geq f(x^*)$ para qualquer $x \in \Omega$. Se $f(x) \geq f(x^*)$ então x^* é um Mínimo Global Estrito (ou Mínimo Absoluto Estrito).*

Definição 2.3. Um ponto $x^* \in \Omega$ é um *Mínimo Local* (ou *Mínimo Relativo*) de f sobre Ω se existe uma vizinhança $V = \{x \in \Omega; \|x - x^*\| \leq \delta\}$ tal que $f(x) \geq f(x^*)$ para qualquer $x \in V$. Se $f(x) > f(x^*)$ para qualquer $x \in V$ então x^* é um *Mínimo Local Estrito* (ou *Mínimo Relativo Estrito*).

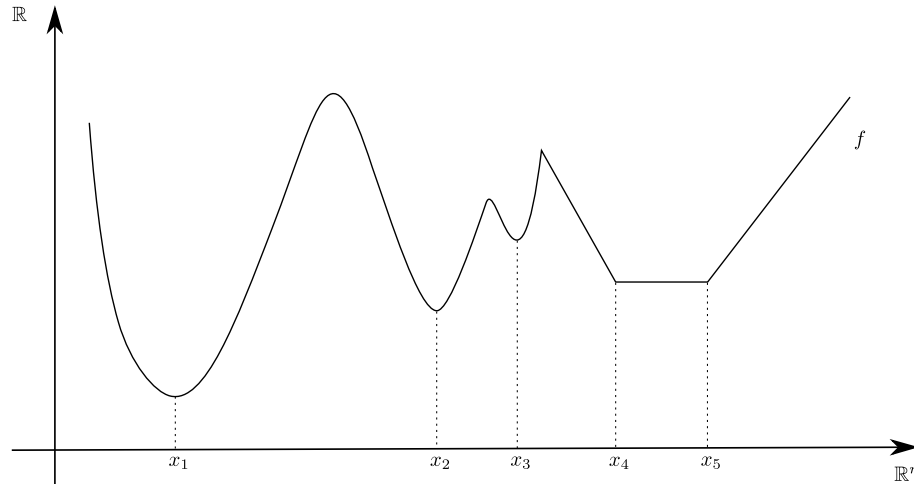


Figura 1 – x_1 é um mínimo global, x_1 , x_2 e x_3 são mínimos locais estritos e o intervalo $[x_4, x_5]$ é um conjunto de mínimos locais.

Por definição, todo mínimo global é também local. Num problema de programação convexa (quando f e Ω são convexas) vale também a recíproca (Teorema 2.1).

Definição 2.4. Um subconjunto $K \subset \mathbb{R}^n$ chama-se *convexo* se para todos $x, y \in K$ e $\lambda \in [0, 1]$ verifica-se $\lambda x + (1 - \lambda)y \in K$, isto é, o segmento de reta que liga quaisquer dois pontos de K está contido em K .

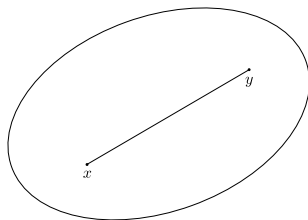


Figura 2 – Conjunto convexo

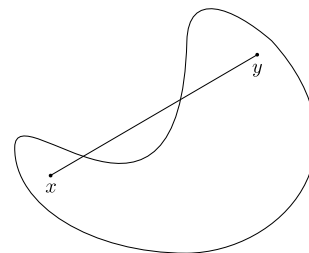


Figura 3 – Conjunto não convexo

Definição 2.5. Uma função real definida em um conjunto convexo K é *convexa* se para todos $x, y \in K$ e $\lambda \in [0, 1]$ verifica-se $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$. Se para todos $x \neq y$ e $\lambda \in [0, 1]$ vale a desigualdade estrita então f é dita *estritamente convexa*.

Teorema 2.1. Em um problema de programação convexa todo mínimo local é global.

Vale ainda num problema de programação convexa que o conjunto dos mínimos globais (ou locais) é convexo. Ainda, se f é estritamente convexa então não pode haver

mais de um mínimo. A prova destas afirmações e do teorema 2.1 pode ser encontrada em [1].

Métodos numéricos para resolução do problema (2.1) são, sobretudo, iterativos. Dado um ponto inicial x^0 na região factível, uma sequência $\{x^k\}$ é obtida de modo recursivo e amiúde através de uma regra do tipo

$$x^{k+1} := \Psi_k(x^k), \quad k = 0, 1, \dots \quad (2.4)$$

onde Ψ é um operador em \mathbb{R}^n . Isto significa que o ponto x^{k+1} é uma melhor aproximação da solução do problema obtida a partir de x^k .

Se x^k é solução do problema para algum k finito então dizemos que o método tem convergência finita. Dizemos que a convergência é assintótica quando a solução não é atingida antes de um número finito de iterações. Neste último caso, não há como garantir que um elemento x^k da sequência é uma solução exata do problema. Resta portanto provar que x^k é de algum modo uma boa aproximação da solução exata, para um k suficientemente grande.

Definição 2.6. *Um método iterativo é globalmente convergente se a sequência $\{x^k\}$ gerada a partir de qualquer ponto inicial $x^0 \in \mathbb{R}^n$ (ou $x^0 \in \Omega$) converge para a solução do problema.*

Eventualmente, algum tipo de convergência pode ser garantida somente a partir de pontos suficientemente próximos da solução x^* do problema. Daí, o ensejo para a

Definição 2.7. *Um método iterativo é localmente convergente se existe um $\rho > 0$ tal que a sequência $\{x^k\}$ gerada a partir de um ponto inicial x^0 satisfazendo $\|x^0 - x^*\| < \rho$ é convergente.*

Uma análise acerca da taxa de convergência de um método com convergência assintótica pode ser feita através da definição de ordem de convergência que se segue.

Definição 2.8. *A ordem de convergência de uma sequência $\{x^k\} \rightarrow x^*$ é o maior valor p positivo que satisfaz*

$$\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^p} = \beta < \infty \quad (2.5)$$

Dizemos que o método possui *convergência linear* se $p = 1$ e *superlinear* se ainda $\beta = 0$. A convergência é *quadrática* quando $p = 2$. Ao valor $\beta < 1$ chamamos de *raio de convergência*.

Os valores de p e β fornecem uma garantia de velocidade na redução da distância de $\|x^{k+1} - x^*\|$ em relação a $\|x^k - x^*\|$ quando k é suficientemente grande. Somente uma boa ordem de convergência não garante a rapidez de um método. A convergência é rápida quando p é grande e β pequeno.

Seja agora $x^k \in \mathbb{R}^n$ uma aproximação da solução do problema de otimização não linear irrestrito definido por

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeito a} && x \in \mathbb{R}^n. \end{aligned} \tag{2.6}$$

onde $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é não linear.

O objetivo num método iterativo para resolução deste problema consiste na atualização de um ponto x^{k+1} de modo que este seja uma melhor aproximação da solução e onde, a priori, deve satisfazer

$$f(x^{k+1}) < f(x^k). \tag{2.7}$$

A estratégia adotada na maior parte dos métodos de programação não linear consiste na busca de uma direção d^k onde f seja decrescente (pelo menos para passos curtos) a partir de x^k nessa direção.

Deste modo, o problema se reduz a calcular um comprimento de passo $\alpha_k > 0$ que determina o ponto x^{k+1} na direção d^k que será implementado. Resolve-se mediante busca linear (exata ou inexata) o subproblema

$$\text{minimizar} \quad f(x^k + \alpha_k d^k) \tag{2.8}$$

na variável α_k . Em seguida, toma-se a atualização $x^{k+1} := x^k + \alpha_k d^k$ satisfazendo (2.7).

Estas direções também são implementadas em métodos restritivos e possuem a propriedade da definição abaixo.

Definição 2.9. *Um vetor $d \in \mathbb{R}^n$ é uma direção de descida de uma função $f: \mathbb{R}^n \rightarrow \mathbb{R}$ em $x \in \Omega$ se existe $\theta > 0$ tal que $f(x + td) \leq f(x)$, $\forall t \in (0, \theta)$.*

O teorema a seguir estabelece uma condição suficiente para que uma direção seja de descida e permite uma interpretação geométrica desta definição ilustrada na figura 2.1.

Teorema 2.2. *Se f é diferenciável em x e $d^t \nabla f(x) < 0$ então d é uma direção de descida para f em x .*

Demonstração. A prova encontra-se em [1]. □

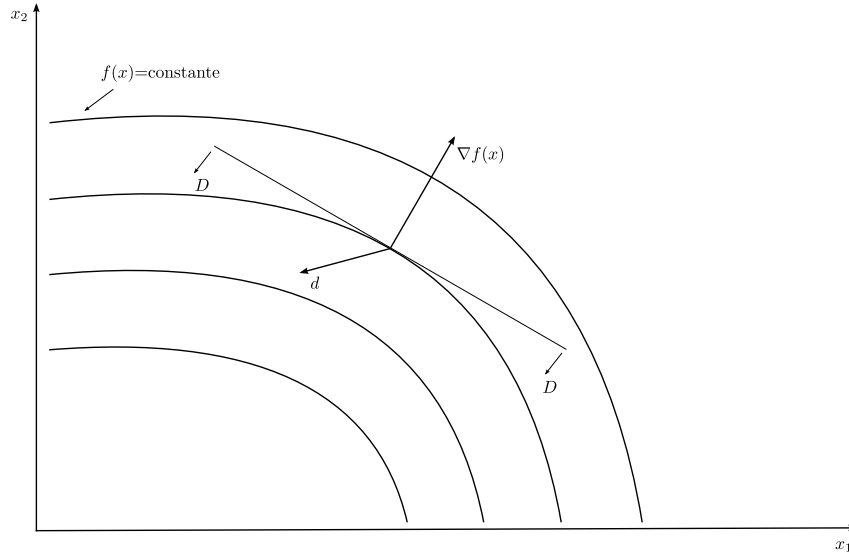


Figura 4 – Uma direção d é de descida se satisfaz $|\angle(\nabla f(x), d)| > \pi/2$. O conjunto de todas as direções de descida constituem o semi-espaço D .

Interessa-nos saber ao passo da k -ésima iteração de um método de pontos interiores para resolução do problema restrito (2.1) se a atualização $x^{k+1} := x^k + \alpha_k d^k$ pertence ao conjunto viável.

Definição 2.10. Um vetor $d \in \mathbb{R}^n$ é uma direção viável em $x \in \Omega$ se existe $\theta > 0$ tal que $x + td \in \Omega, \forall t \in [0, \theta]$.

A figura 5 retrata um problema de otimização com restrições de desigualdade. Note pela definição que a direção d em x e quaisquer outras em Ω^0 são viáveis.

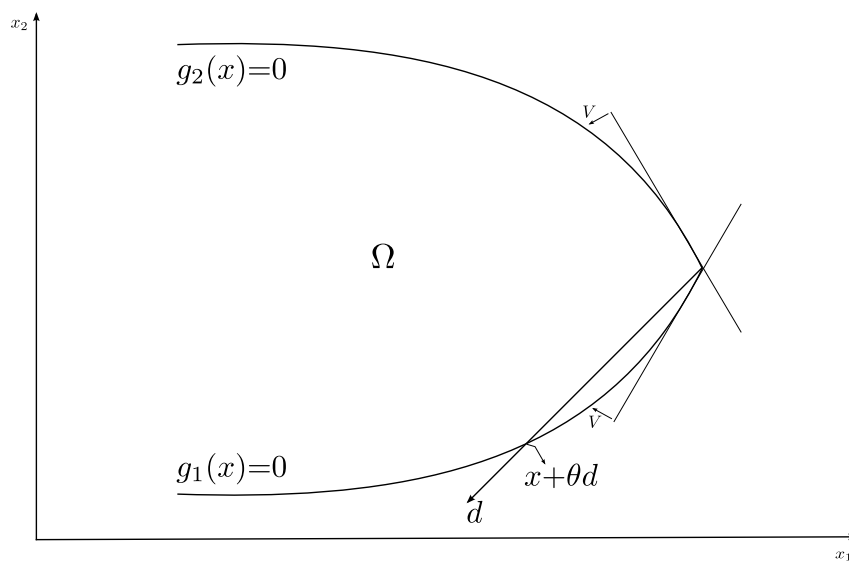


Figura 5 – Direções viáveis

2.1.1 Condições de Otimalidade

Nosso objetivo nesta seção é caracterizar mínimos locais de problemas irrestritos e restritos do tipo (2.6). Os resultados a seguir fornecem condições necessárias e suficientes que tais pontos devem satisfazer.

Teorema 2.3 (Condição necessária de primeira ordem). *Sejam $f: \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^1 e x^* um mínimo local de f sobre Ω . Qualquer direção viável d em x^* é tal que $d^t \nabla f(x^*) \geq 0$.*

Demonstração. Veja [10]. □

O teorema estabelece que uma direção viável num mínimo local não é uma direção de descida. Note ainda que se $x^* \in \Omega^0$ então toda direção em x^* é viável e pelo teorema acima temos $d^t \nabla f(x^*) \geq 0, \forall d \in \mathbb{R}^n$. Em particular, $d = -\nabla f(x^*) \Rightarrow -\|\nabla f(x^*)\|^2 \geq 0$ e, conseqüentemente, $\nabla f(x^*) = 0$. Formalizamos este fato no corolário a seguir.

Corolário 2.1 (Condição necessária de primeira ordem - caso irrestrito). *Sejam $f: \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^1 e x^* um mínimo local de f sobre Ω . Se $x^* \in \Omega^0$ então $\nabla f(x^*) = 0$.*

As condições necessárias de segunda ordem determinadas em termos da matriz Hessiana de f definida por

$$\nabla^2 f(x) = \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right]_{n \times n} \quad (2.9)$$

fornecem mais informações sobre mínimos locais.

Teorema 2.4 (Condições necessárias de segunda ordem). *Sejam $f: \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^2 e x^* um mínimo local de f sobre Ω . Se $d \in \mathbb{R}^n$ é uma direção viável em x^* tal que $d^t \nabla f(x^*) = 0$ então $d^t \nabla^2 f(x^*) d \geq 0$.*

Demonstração. Veja [10]. □

Expressamos na proposição seguinte o caso particular em que $x^* \in \Omega^0$. Isto ocorre por exemplo se $\Omega = \mathbb{R}^n$.

Proposição 2.1 (Condições necessárias de segunda ordem - caso irrestrito). *Se $x^* \in \Omega^0$ é um mínimo local de $f \in C^2$ definida sobre Ω então qualquer $d \in \mathbb{R}^n$ satisfaz $d^t \nabla^2 f(x^*) d \geq 0$. Isto é, $\nabla^2 f(x^*)$ é semi-definida positiva.*

A proposição seguinte aplica-se a problemas sem restrições ou ainda aos quais o mínimo local encontra-se no interior da região viável.

Proposição 2.2 (Condições suficientes de segunda ordem - caso irrestrito). *Suponhamos que $f: \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ é de classe C^2 e $x^* \in \Omega^0$ são tais que*

- (i) $\nabla f(x^*) = 0$;
- (ii) $\nabla^2 f(x^*)$ é definida positiva.

Então x^* é um mínimo local estrito de f .

Demonstração. Veja [10]. □

Consideremos agora o problema de otimização dado por restrições de igualdades e desigualdades (2.3).

Definição 2.11. Um ponto $x \in \Omega$ é um ponto regular das restrições se os vetores $\nabla h_i(x)$ para $i = 1, \dots, p$ e $\nabla g_i(x)$ para $i \in I(x)$ são linearmente independentes, onde $I(x) = \{i \mid g_i(x) = 0\}$ é o conjunto das restrições ativas de x .

As condições necessárias de primeira ordem para este problema são chamadas de condições de Karush-Kuhn-Tucker.

Teorema 2.5 (Condições de Karush-Kuhn-Tucker). Se um ponto x^* é regular das restrições e de mínimo local do problema (2.3) então existem $\lambda^* \in \mathbb{R}^m$ e $\mu^* \in \mathbb{R}^p$ tais que

$$\nabla f(x^*) + \nabla g(x^*)\lambda^* + \nabla h(x^*)\mu^* = 0, \quad (2.10)$$

$$G(x^*)\lambda^* = 0, \quad (2.11)$$

$$h(x^*) = 0, \quad (2.12)$$

$$g(x^*) \leq 0, \quad (2.13)$$

$$\lambda \geq 0, \quad (2.14)$$

onde $G(x^*)$ é a matriz diagonal $m \times m$ tal que $[G(x^*)]_{ii} = g_i(x^*)$.

Demonstração. Veja [10]. □

A condição (2.11) é chamada condição de complementaridade e pode ser expressa também por $\lambda_i^* g_i(x^*) = 0, \forall i \in \{1, \dots, m\}$. Noutros termos, esta determina que

$$\lambda_i^* = 0, \forall i \in \{1, \dots, m\} \setminus I(x^*).$$

Definição 2.12. Definimos a função Lagrangiana $L: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ associada ao problema (2.3) por

$$L(x, \lambda, \mu) = f(x) + \lambda^t g(x) + \mu^t h(x), \quad (2.15)$$

onde $\lambda \in \mathbb{R}^m$ e $\mu \in \mathbb{R}^p$ são os multiplicadores de Lagrange das restrições g e h , respectivamente.

A partir desta definição a equação (2.10) pode ser expressa de forma mais compacta por

$$\nabla_x L(x, \lambda, \mu) = 0.$$

Definição 2.13. A hessiana da função lagrangiana para o problema (2.3) define-se pela função $H: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}^{n \times n}$ dada por

$$H(x, \lambda, \mu) = \nabla_{xx}^2 L(x, \lambda, \mu) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x) + \sum_{i=1}^p \mu_i \nabla^2 h_i(x)$$

Definição 2.14. Definimos o espaço tangente em x regular por

$$T = \{d \mid \nabla g_i^t(x)d = 0 \text{ para } i \in I(x) \text{ e } \nabla h_i^t(x)d = 0 \text{ para } i = 1, \dots, p\}.$$

Teorema 2.6 (Condições necessárias de segunda ordem). *Suponhamos que as funções f , g e h do problema (2.3) são de classe C^2 e x^* é um ponto regular das restrições e mínimo local deste problema. Então existem $\lambda^* \in \mathbb{R}^m$ e $\mu^* \in \mathbb{R}^p$ que satisfazem (2.10) - (2.14) e ainda tais que a matriz $H(x^*, \lambda^*, \mu^*)$ é semidefinida positiva no espaço tangente, ou seja,*

$$d^t H(x^*, \lambda^*, \mu^*) d \geq 0$$

para todo $d \in T$.

Demonstração. Veja [10]. □

Teorema 2.7 (Condições suficientes de segunda ordem). *Sejam f , g e h de classe C^2 e x^* um ponto tal que $h(x^*) = 0$ e $g(x^*) \leq 0$. Suponhamos que existem $\lambda^* \in \mathbb{R}^m$, $\lambda^* \geq 0$, e $\mu^* \in \mathbb{R}^p$ tais que*

$$\nabla f(x^*) + \nabla g(x^*)\lambda^* + \nabla h(x^*)\mu^* = 0$$

e $H(x^, \lambda^*, \mu^*)$ definida positiva no espaço tangente. Então x^* é um mínimo local estrito do problema (2.3).*

Demonstração. Veja [10]. □

2.1.2 Método de Newton para equações

Os algoritmos FDIPA e FDA descritos neste trabalho são modificações do método de Newton para resolução de sistemas não lineares do tipo

$$\Phi(x) = 0 \tag{2.16}$$

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

Seja $x^k \in \mathbb{R}^n$ suficientemente próximo a uma solução de (2.16). O método de Newton deriva da aproximação linear de Φ em torno a x^k dada por

$$M_k(x) = \Phi(x^k) + \nabla \Phi(x^k)(x - x^k),$$

aonde tomamos o próximo ponto x^{k+1} como a raiz de M_k . Isto significa que a atualização x^{k+1} tal que $M_k(x^{k+1}) = 0$ determina um ponto mais perto da solução de (2.16).

A partir destas considerações, de modo mais preciso, definimos o método através da direção de Newton d^k e pela atualização $x^{k+1} := x^k + d^k$, aonde

$$\nabla\Phi(x^k)d^k = -\Phi(x^k). \quad (2.17)$$

O algoritmo básico do Método de Newton é apresentado a seguir. Um teste de convergência deve ser implementado como critério de parada [9].

Algoritmo 2.1 (Método de Newton).

Parâmetros: $\epsilon > 0$.

Dados Iniciais: $x^0 \in \mathbb{R}^n$ e $k := 0$.

Início

Passo 1. Cálculo da direção de Newton.

Resolva o sistema $\nabla\Phi(x^k)d^k = -\Phi(x^k)$.

Passo 2. Atualização.

Tomar $x^{k+1} := x^k + d^k$.

Passo 3. Critério de parada.

Fim

A figura abaixo ilustra um exemplo das iterações do método para um caso bidimensional.

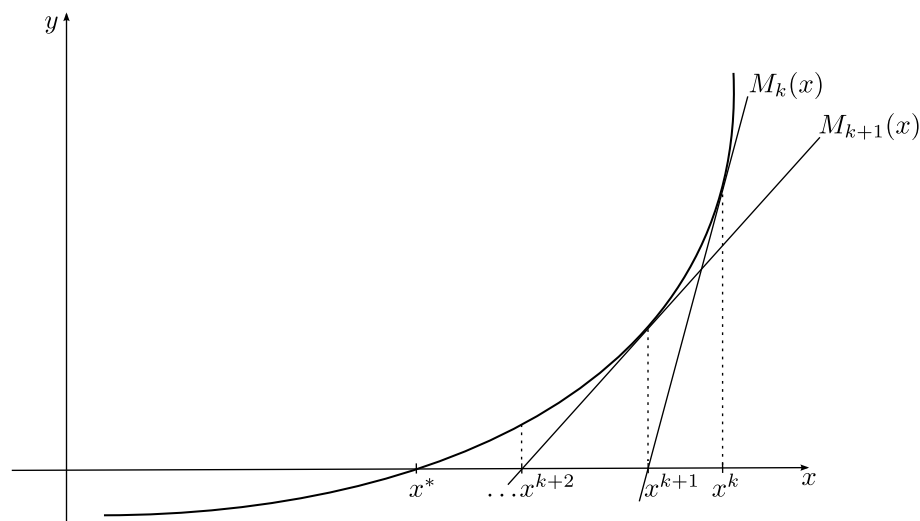


Figura 6 – Método de Newton

Se a matriz $\nabla\Phi(x^k)$ é não singular, a sequência do método pode ser obtida pela atualização $x^{k+1} := x^k + d^k$, aonde

$$d^k = -(\nabla\Phi(x^k))^{-1}\Phi(x^k), \quad k = 0, 1, \dots \quad (2.18)$$

Dizemos que a sequência de Newton está bem definida se a matriz $\nabla\Phi(x^k)$ é não singular para todo $k \geq 0$. O próximo teorema cuja prova encontra-se em [2, 6] estabelece condições para a convergência local do método.

Teorema 2.8 (Convergência local do método de Newton). *Sejam x^* uma solução de (2.16) e $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$ diferenciável numa vizinhança de x^* com derivada contínua neste ponto e ainda tal que a matriz $\nabla\Phi(x^*)$ é não singular. Se x^0 é um ponto suficientemente próximo a x^* então o algoritmo (2.1) gera uma sequência bem definida $\{x^k\}$ que converge a x^* com ordem superlinear. Além disso, se Φ é localmente lipschitziana em x^* então a ordem de convergência é quadrática.*

A hipótese de Φ ser localmente lipschitziana ao redor da solução de $\Phi(x) = 0$ garante a convergência quadrática do algoritmo no sentido definido em (2.8), i.e., existe $K > 0$ tal que

$$\|x^{k+1} - x^*\| \leq K\|x^k - x^*\|^2,$$

para todo k suficientemente grande.

2.1.3 Buscas Lineares

Num problema de minimização sem restrições definido por

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && x \in \mathbb{R}^n. \end{aligned} \quad (2.19)$$

uma vez obtida a direção de descida, diferentes buscas lineares inexatas podem ser implementadas a fim de obter um novo ponto $x + \alpha d$ que proporciona uma boa redução da função potencial f e que ainda assevera a convergência (global se possível) do método para a solução do problema.

Tratamos aqui duas destas buscas - Armijo e Wolfe. Na primeira, definimos o comprimento de passo α como o primeiro valor da sequência $\{1, \nu, \nu^2, \nu^3, \dots\}$ que satisfaz

$$f(x + \alpha d) \leq f(x) + \alpha\eta_1 \nabla f(x)^t d, \quad (2.20)$$

aonde $\nu \in (0, 1)$ e $\eta_1 \in (0, 1)$ são parâmetros fixados. Chamamos a desigualdade (2.20) de condição de Armijo.

A principal vantagem ao se optar por essa busca é seu baixo custo computacional à medida que não há necessidade do cálculo de derivadas a cada iteração.

Destacamos também sobre a busca de Armijo, o fato de que não é sempre possível obter valores α próximos de 1 quando a iteração se aproxima da fronteira do problema em questão - podendo inclusive ocorrer o conhecido “efeito Maratos” [13].

Note na figura abaixo que um passo aceitável para α encontra-se sob o gráfico de $p(\alpha) = f(x) + \alpha\eta_1\nabla f(x)^t d$.

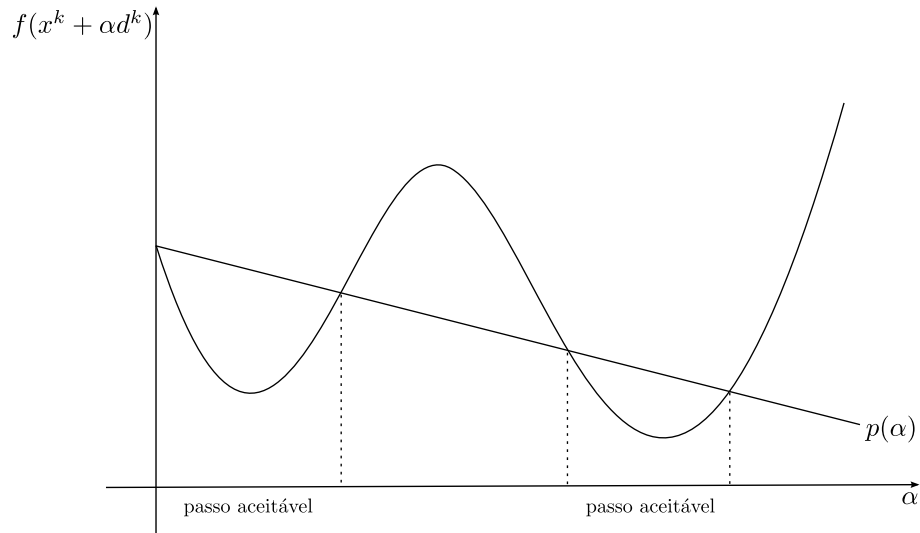


Figura 7 – Busca linear inexata de Armijo

Se f é diferenciável em x e d é uma direção de descida em x então a busca de Armijo está bem definida, isto é, existe um $\alpha \in \{1, \nu, \nu^2, \dots\}$ que satisfaz a condição (2.20). Este resultado é demonstrado em [6, 10]. Uma demonstração também encontra-se em [4], para o caso em que f é de classe C^1 e $\nabla f(x)$ satisfaz a condição de Lipschitz num conjunto compacto $\Omega_a = \{x \in \Omega \mid f(x) \leq a\}$ de interior não vazio.

Segue agora o algoritmo da busca linear inexata de Armijo para um problema de minimização sem restrições. O algoritmo retorna um valor de α satisfazendo a condição (2.20).

Algoritmo 2.2 (Busca de Armijo).

Parâmetros: $\eta_1 \in (0, 1)$, $\nu \in (0, 1)$

Dados Iniciais: x , $f(x)$, $\nabla f(x)$, d , $\alpha_1 := 1$, $i := 1$.

Início

1. Se $f(x + \alpha_i d) \leq f(x) + \eta_1 \alpha_i \nabla f(x)^t d$, passar ao item 4.
2. Tomar $\alpha_{i+1} := \nu \alpha_i$.
3. Fazer $i := i + 1$ e passar ao item 1.
4. Tomar $\alpha = \alpha_i$.

Fim

Definimos na busca de Wolfe o comprimento de passo α como o primeiro valor da sequência $\{1, \nu, \nu^2, \nu^3, \dots\}$ satisfazendo (2.20) e também a seguinte condição

$$\nabla f(x + \alpha d)d \geq \eta_2 \nabla f(x)d \quad (2.21)$$

para alguma constante $\eta_2 \in (\eta_1, 1)$ fixa. Chamamos a desigualdade (2.21) de condição de Wolfe.

Exige-se nesta segunda condição que a inclinação de ∇f em $x + \alpha d$ não seja menor que uma contante da inclinação em x e deste modo impõe-se um limite inferior na escolha de α . Isto impede que passos “muito curtos” sejam obtidos caso somente a condição de Armijo é exigida.

Note na figura 8 que quanto menor o valor de α (próximo a um mínimo local) menor a inclinação de ∇f em $x + \alpha d$.

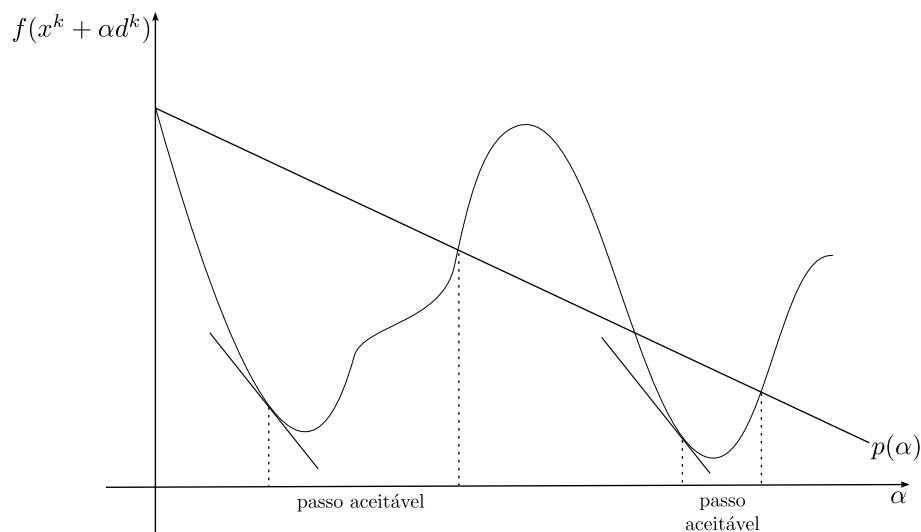


Figura 8 – Busca linear inexata de Wolfe

É provado em [6] que se f é de classe C^1 e limitada inferiormente então a busca de Wolfe está bem definida, ou seja, existe um $\alpha > 0$ que satisfaz as condições (2.20) e (2.21).

O algoritmo da busca de Wolfe para um problema irrestrito é descrito na sequência. O algoritmo retorna um valor α_k satisfazendo as condições exigidas.

Algoritmo 2.3 (Busca de Wolfe).

Parâmetros: $\eta_1 \in (0, 1)$, $\eta_2 \in (\eta_1, 1)$

Dados Iniciais: x , $f(x)$, $\nabla f(x)$, d , $\check{\alpha} := 0$, $\hat{\alpha} := 0$ e $\alpha := 1$.

Início

1. Se $f(x + \alpha d) \leq f(x) + \eta_1 \alpha \nabla f(x)d$ e $\nabla f(x + \alpha d)d \geq \eta_2 \nabla f(x)d$, passar ao item 6.
2. Se $f(x + \alpha d) > f(x) + \eta_1 \alpha \nabla f(x)d$, tomar $\hat{\alpha} := \alpha$ e passar ao item 5.
3. Se $\nabla f(x + \alpha d)d < \eta_2 \nabla f(x)d$, tomar $\check{\alpha} := \alpha$.
4. Se $\hat{\alpha} = 0$, escolher um novo valor $\alpha > 0$ por extrapolação e passar ao item 1.
5. Escolher um novo valor $\alpha \in (\check{\alpha}, \hat{\alpha})$ por interpolação e passar ao item 1.
6. Tomar $\alpha_k := \alpha$.

Fim

A extrapolação do item 4 pode ser implementada fixando $\theta_1 > 0$ e tomando a cada iteração $\alpha := \theta_1 \alpha$.

Já a interpolação do item seguinte, fixando $\theta_2 \in (0, 1)$ e tomando $\alpha := (1 - \theta_2)\check{\alpha} + \theta_2 \hat{\alpha}$. Deste modo, minora-se o intervalo $(\check{\alpha}, \hat{\alpha})$ até que um passo aceitável seja encontrado. Note que o valor $\hat{\alpha}$ aumenta a cada atualização do item 2 mas nunca ultrapassando $\check{\alpha}$.

Interpolações mais sofisticadas como a quadrática ou cúbica também podem ser efetuadas [15].

O algoritmo termina com um valor α_k no item 6 quando as condições de Armijo (2.20) e Wolfe (2.21) são satisfeitas no item 1.

Nas buscas lineares implementadas em algoritmos como o FDIPA e FDA para resolução de problemas restritivos, condições de viabilidade também devem ser exigidas.

2.2 Problema de Complementaridade

Um problema de complementaridade não linear (NCP, do inglês Nonlinear Complementarity Problem) definido por uma aplicação $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ de classe C^1 consiste em encontrar um vetor $x \in \mathbb{R}^n$ que satisfaz as condições

$$x \geq 0, \quad F(x) \geq 0, \quad x^T F(x) = 0. \quad (2.22)$$

Definição 2.15. O conjunto $\Omega = \{x \in \mathbb{R}^n \mid x \geq 0 \text{ e } F(x) \geq 0\}$ é chamado de conjunto dos pontos viáveis do problema de complementaridade (2.22).

Definição 2.16. Denota-se por Ω^0 ao conjuntos dos pontos no interior de Ω . Isto é, $\Omega^0 = \{x \in \mathbb{R}^n \mid x > 0 \text{ e } F(x) > 0\}$ e denomina-se conjunto dos pontos estritamente viáveis de (2.22).

Definição 2.17. Uma solução x^* do problema (2.22) é dita degenerada se $x_i^* = 0$ e $F_i(x^*) = 0$ para alguma índice $i \in \{1, \dots, n\}$ e não degenerada se para todo $i \in \{1, \dots, n\}$ tem-se $x_i F_i(x) \neq 0$.

O NCP acima pode ser resolvido numericamente por meio de vários métodos existentes. Abordamos neste trabalho sobre dois destes. O primeiro denomina-se FDA e foi desenvolvido por Sandro R. Mazorche [14]. O outro consiste em reescrever (2.1) como um problema de otimização sem restrições equivalente e obter uma solução via FDIPA.

Proposição 2.3. *Se $d \in \mathbb{R}^n$ e $x \in \Omega$ satisfazem as condições*

(i) $d_i > 0$ sempre que $x_i = 0$

(ii) $d^t \nabla F_i(x) > 0$ sempre que $F_i(x) = 0$

então d é uma direção viável do NCP, em x .

A teoria fundamental de otimização referente aos problemas de minimização e complementaridade está fundamentada. Nosso objetivo nos capítulos 3 e 4 será descrever os métodos FDIPA e FDA.

3 FDIPA

Consideremos o problema de minimização dado unicamente por restrições de desigualdades

$$\begin{aligned} & \text{minimizar } f(x) \\ & \text{sujeito a } g(x) \leq 0. \end{aligned} \tag{3.1}$$

onde g é diferenciável e o conjunto viável é dado por $\Omega = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$.

Uma condição mais forte do que a viabilidade de uma direção em um ponto $x \in \Omega$ é dada na definição abaixo.

Definição 3.1. *Um campo vetorial $d(x)$ definido em Ω é dito campo uniforme de direções viáveis do problema (3.1), se existe $\tau > 0$ tal $x + td \in \Omega$ para todo $t \in [0, \tau]$ e para todo $x \in \Omega$.*

A direção do algoritmo FDIPA constitui um campo uniforme de direções viáveis. Caso esta condição não seja exigida, comprimentos de passos muito baixos podem ser obtidos para pontos perto da fronteira de Ω e conseqüentemente forçar a convergência do algoritmo para pontos que não são de Karush-Kuhn-Tucker (definição 3.6).

Os conceitos de função Lagrangiana, matriz hessiana, conjunto de restrições ativas e ponto regular para este problema são definidos adiante.

Definição 3.2. *Definimos a função Lagrangiana $L: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ associada ao problema (3.1) por*

$$L(x, \lambda) = f(x) + \lambda^t g(x) \tag{3.2}$$

onde $\lambda \in \mathbb{R}^m$ é o multiplicadores de Lagrange da restrição g .

Definição 3.3. *A Hessiana da função lagrangiana para o problema (3.1) define-se pela função $H: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n \times n}$ dada por*

$$H(x, \lambda) = \nabla_{xx}^2 L(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x)$$

Definição 3.4. *Chamamos o conjunto $I(x) = \{i \mid g_i(x) = 0, 1 \leq i \leq m\}$ de conjunto das restrições ativas em x e dizemos que x é um ponto regular se os vetores $\nabla g_i(x)$ para $i \in I(x)$ são linearmente independentes.*

Como g é diferenciável, decorre da fórmula de Taylor que se d satisfaz $d^t \nabla g(x) < 0$ para todo $i \in I(x)$, então d é uma direção viável em x . Isto é, $g(x + td) < 0$ para todo t suficientemente pequeno.

Um caso particular das condições necessárias de primeira ordem estabelecidas para o problema de minimização mais geral dado por restrições de igualdades e desigualdades também podem ser formuladas para o problema (3.1).

Teorema 3.1 (Condições de Karush-Kuhn-Tucker). *Seja x^* um ponto regular das restrições $g(x) \leq 0$ e mínimo local do problema (3.1). Existe um vetor $\lambda^* \in \mathbb{R}^m$ tal que*

$$\nabla f(x^*) + \nabla g(x^*)\lambda^* = 0, \quad (3.3)$$

$$G(x^*)\lambda^* = 0, \quad (3.4)$$

$$\lambda^* \geq 0, \quad (3.5)$$

$$g(x^*) \leq 0. \quad (3.6)$$

Definição 3.5. *Dizemos que $x \in \mathbb{R}^n$ é um ponto estacionário do problema (3.1) se existe um multiplicador de Lagrange $\lambda \in \mathbb{R}^m$ que satisfaz as condições (3.3) e (3.4).*

Definição 3.6. *Dizemos que $x \in \mathbb{R}^n$ é um ponto de Karush-Kuhn-Tucker (ou ponto de KKT) do problema (3.1) se é estacionário e o multiplicador de Lagrange $\lambda \in \mathbb{R}^m$ satisfaz $\lambda \geq 0$. Chamamos (x, λ) de par KKT, aonde x é um ponto de KKT e $\lambda \geq 0$ é o multiplicador de Lagrange associado .*

3.1 Descrição do Método FDIPA

As seguintes suposições acerca do problema (3.1) serão consideradas.

Suposição 3.1. Existe um número real a tal que o conjunto $\Omega_a = \{x \in \Omega \mid f(x) \leq a\}$ é compacto e tem interior Ω_a^0 não vazio.

Suposição 3.2. Todo ponto $x \in \Omega_a^0$ satisfaz $g(x) < 0$.

Suposição 3.3. As funções f e g são de classe C^1 em Ω_a^0 e suas derivadas satisfazem a condição de Lipschitz.

Suposição 3.4. Todo ponto $x \in \Omega_a$ é regular.

O algoritmo FDIPA gera uma sequência de pontos interiores a partir de um dado ponto inicial também de interior e converge globalmente com ordem superlinear para um par KKT do problema após um número finito de iterações.

Em cada iteração uma direção de descida da função potencial é calculada inicialmente pela resolução de um sistema nas variáveis dual e primal. Em seguida, uma direção viável é obtida pela resolução desse mesmo sistema modificado por uma perturbação. A última etapa da iteração consiste em uma busca linear inexata na qual é definido um ponto na região viável que decresça o valor da função potencial de modo que assegure a convergência global do algoritmo.

Inicialmente, aplicamos o método de Newton (descrito na seção 2.1.2) às condições (3.3) e (3.4) de Karush-Kuhn-Tucker. Com efeito, tomando x_α^k e λ_α^k como as novas

estimativas de x^k e λ^k , respectivamente, obtemos

$$\begin{bmatrix} B_k & \nabla g(x^k) \\ \Lambda_k \nabla g^t(x^k) & G(x^k) \end{bmatrix} \begin{bmatrix} x_\alpha^k - x^k \\ \lambda_\alpha^k - \lambda^k \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) + \nabla g(x^k) \lambda^k \\ G(x^k) \lambda^k \end{bmatrix} \quad (3.7)$$

aonde Λ_k é a matriz diagonal $m \times m$ com $[\Lambda_k]_{ii} = \lambda_i^k$ e B_k é $n \times n$ e pode ser tomada pela matriz identidade, pela aproximação quasi-Newton da matriz hessiana $H(x^k, \lambda^k)$ ou ainda pela própria $H(x^k, \lambda^k)$ nos casos em que é definida positiva. Cada uma destas regras de atualização será discutida na próxima seção.

Em seguida efetuamos a multiplicação matricial acima e fazemos $d_\alpha^k = x_\alpha^k - x^k$. Daí, segue

$$\begin{aligned} B_k d_\alpha^k + \nabla g(x^k) \lambda_\alpha^k - \nabla g(x^k) \lambda^k &= -(\nabla f(x^k) + \nabla g(x^k) \lambda^k), \\ \Lambda_k \nabla g^t(x^k) d_\alpha^k + G(x^k) \lambda_\alpha^k - G(x^k) \lambda^k &= -G(x^k) \lambda^k. \end{aligned} \quad (3.8)$$

Eliminando os termos $\nabla g(x^k) \lambda^k$ da primeira equação e $G(x^k) \lambda^k$ da segunda, obtemos o sistema escrito novamente na forma matricial

$$\begin{bmatrix} B_k & \nabla g(x^k) \\ \Lambda_k \nabla g^t(x^k) & G(x^k) \end{bmatrix} \begin{bmatrix} d_\alpha^k \\ \lambda_\alpha^k \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) \\ 0 \end{bmatrix} \quad (3.9)$$

Como por hipótese B_k é definida positiva, decorre do lema (3.3) provado na seção seguinte que d_α^k é uma direção de descida da função objetivo.

Não podemos no entanto afirmar que d_α^k constitui um campo uniforme de direções viáveis. De fato, a i -ésima linha de (3.8) é dada por

$$\lambda_{\alpha_i}^k \nabla g_i^t(x^k) d_\alpha^k + g_i(x^k) \lambda_{\alpha_i}^k = 0, \quad i = 1, \dots, m \quad (3.10)$$

e logo $\nabla g_i^t(x^k) d_\alpha^k = 0$ sempre que $g_i(x^k) = 0$. Isto significa que d_α^k tangencia o conjunto viável à medida que $g_i(x^k) \rightarrow 0$ e portanto não é um campo uniforme de direções viáveis.

Por esta razão devemos obter um novo sistema cujo vetor solução seja uma direção de busca que possua ambas as propriedades de viabilidade e descida.

Consideremos agora o sistema

$$\begin{bmatrix} B_k & \nabla g(x^k) \\ \Lambda_k \nabla g^t(x^k) d^k & G(x^k) \end{bmatrix} \begin{bmatrix} d^k \\ \bar{\lambda}_\alpha^k \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) \\ \rho^k \Lambda_k \omega \end{bmatrix} \quad (3.11)$$

obtido a partir de (3.9) por uma perturbação, onde $\rho^k \in \mathbb{R}$ e $\omega \in \mathbb{R}^n$ são ambos positivos.

Um vetor solução qualquer desse novo sistema será uma direção de busca com a primeira das propriedades discutidas acima. Adiante obtemos uma solução particular deste que por fim será a direção de busca que goza de ambas as propriedades.

Para que d^k seja uma direção viável é suficiente que se tenha $\nabla g_i^t(x^k)d^k < 0$. Ora, o sistema agora considerado corresponde a

$$\begin{aligned} B_k d^k + \nabla g(x^k) \bar{\lambda}_\alpha^k &= -\nabla f(x^k), \\ \Lambda_k \nabla g^t(x^k) d^k + G(x^k) \bar{\lambda}_\alpha^k &= -\rho^k \Lambda_k \omega, \end{aligned} \quad (3.12)$$

onde a i -ésima linha de (3.12) satisfaz

$$\lambda_i^k \nabla g_i^t(x^k) d^k + g_i(x^k) \bar{\lambda}_{\alpha_i}^k = -\rho^k \lambda_i^k \omega_i, \quad i = 1, \dots, m.$$

Este que por sua vez nos dá se $g_i(x^k) = 0$ que $\nabla g_i^t(x^k) d^k = -\rho^k \omega_i < 0$. Daí vemos que d^k é uma direção viável. Todavia não podemos afirmar que toda solução d^k é também direção de descida.

Ademais, observe agora que como $(d_\alpha^k)^t \nabla f(x^k) < 0$ então uma direção d^k que em particular satisfaz a condição, $\forall \xi \in (0, 1)$,

$$(d^k)^t \nabla f(x^k) < \xi (d_\alpha^k)^t \nabla f(x^k), \quad (3.13)$$

é também uma direção de descida. Num sentido que ficará evidente logo adiante, um d^k que cumpre a relação acima será obtido impondo-se um limite adequado para ρ^k .

Motivados por este fato, devemos obter uma direção de restauração que nos assegure obter um d^k viável e de descida na qual baseia-se o algoritmo. Denotemos tal direção por d_β^k e como segue mostramos que é possível obtê-la e que deveras restaura d_α^k .

Ainda, como d^k é uma deflexão de d_α^k proporcional a ρ^k , façamos $d^k = d_\alpha^k + \rho^k d_\beta^k$. Desta forma, o problema se reduz a encontrar um ρ^k adequado. É claro que d_β^k e ρ^k devem ser escolhidos de tal forma que d^k ainda seja solução de (3.11). Fazemos valer esta condição obtendo d_β^k do sistema

$$\begin{bmatrix} B_k & \nabla g(x^k) \\ \Lambda_k \nabla g(x^k) & G(x^k) \end{bmatrix} \begin{bmatrix} d_\beta^k \\ \lambda_\beta^k \end{bmatrix} = \begin{bmatrix} 0 \\ -\Lambda_k \omega \end{bmatrix} \quad (3.14)$$

onde $\lambda^k = \lambda_\alpha^k + \rho^k \lambda_\beta^k$. É imediato verificar que $d^k = d_\alpha^k + \rho^k d_\beta^k$ é uma solução de (3.11).

De resto, segue que um ρ^k tal que

$$\rho^k < \frac{(\xi - 1)(d_\alpha^k)^t \nabla f(x^k)}{(d_\beta^k)^t \nabla f(x^k)}.$$

é suficiente para que se tenha (3.13) e finalmente podemos afirmar que $d^k = d_\alpha^k + \rho^k d_\beta^k$ é uma direção viável e de descida.

3.1.1 Buscas Lineares no FDIPA

Obtemos até aqui uma direção de descida e viável da função objetivo f . Diferentes métodos de busca inexata podem ser implementadas a fim de obter um tamanho de passo

α_k de modo que a atualização $x^{k+1} := x^k + \alpha_k d^k$ minimize f e ainda ateste a convergência do método FDIPA a um par KKT do problema.

Diferente de quando tratamos de buscas lineares para um problema irrestrito, exigimos para o caso restrito também condições de viabilidade.

Na busca de Armijo, escolhemos o tamanho de passo α_k como o primeiro valor da sequência $\{1, \nu, \nu^2, \nu^3, \dots\}$ que satisfaz as condições

$$f(x^k + \alpha d^k) \leq f(x^k) + \eta_1 \alpha_k \nabla f(x^k) d^k \quad (3.15)$$

e ainda as condições denominadas uniforme de viabilidade, dadas por

$$g_i(x^k + \alpha d^k) \leq 0, \quad (3.16)$$

se $\lambda_i \geq 0$ e, caso contrário, por

$$g_i(x^k + \alpha d^k) \leq g_i(x). \quad (3.17)$$

Exigimos na busca de Wolfe além das condições de Armijo e uniforme de viabilidade que ao menos uma das $m + 1$ condições

$$\begin{aligned} \nabla f(x^k + \alpha d^k) d^k &\geq \eta_2 \nabla f(x^k) d^k \\ g_i(x^k + \alpha d^k) &\geq \gamma g_i(x^k) \end{aligned}$$

seja satisfeita, aonde $i = 1, \dots, m$, $0 < \eta_1 < \eta_2 < 1$ e $\gamma \in (0, 1)$.

Segue agora a descrição do algoritmo simplificado do método FDIPA apresentada na seção anterior.

Algoritmo 3.1 (FDIPA).

Parâmetros: $\epsilon, \varphi > 0$; $\xi, \nu, \eta \in (0, 1)$

Dados Iniciais: $x^0 \in \Omega^0$, $\lambda_0 > 0$, $0 < \omega \in \mathbb{R}^n$ e $B^0 \in \mathbb{R}^{n \times n}$ simétrica e definida positiva.

Início

Passo 1: Direção de Busca.

1. Obter $(d_\alpha^k, \lambda_\alpha^k)$ através do sistema (3.9).

Se $d_\alpha^k = 0$, fim.

2. Calcular $(d_\beta^k, \lambda_\beta^k)$ através do sistema (3.14).

3. Se $\nabla f(x^k) d_\beta^k > 0$, definir $\rho^k = \min \left\{ \varphi \|d_\alpha^k\|^2; (\xi - 1) \frac{(d_\alpha^k)^t \nabla f(x^k)}{(d_\beta^k)^t \nabla f(x^k)} \right\}$.

Senão, defina $\rho^k = \varphi \|d_\alpha^k\|^2$.

4. Tomar $d^k := d_\alpha^k + \rho^k d_\beta^k$ e $\lambda^k := \lambda_\alpha^k + \rho^k \lambda_\beta^k$.

Passo 2: Busca Linear.

Encontrar um passo aceitável α_k satisfazendo as buscas de Armijo ou Wolfe e as condições de viabilidade.

Passo 3: Atualização.

Fazer $x^{k+1} = x^k + t^k d^k$ e $k := k + 1$.

Passo 4: Critério de Parada.

Se $f(x^k) < \epsilon$, fim. Senão retornar ao passo 1.

Fim

3.2 Convergência Global do algoritmo FDIPA

Uma das características do algoritmo FDIPA consiste na sua flexibilidade em relação às atualizações dos parâmetros de entrada λ , B e w . Nesta seção provamos a convergência global do algoritmo quaisquer que sejam as atualizações implementadas sobre os mesmos. As hipóteses necessárias para tal fim estão estabelecidas nas suposições abaixo.

Suposição 3.5 (Sobre λ). Existem números positivos λ^I , λ^S e β tais que $0 < \lambda_i \leq \lambda^S$, $\forall i = 1, \dots, m$. Ainda, $\lambda_i \geq \lambda^I$ para todo i tal que $g_i(x) \geq -\beta$.

Suposição 3.6 (Sobre B). Existem números positivos σ_1 e σ_2 tais que

$$\sigma_1 \|d\|^2 \leq d^T B d \leq \sigma_2 \|d\|^2, \quad \forall d \in \mathbb{R}^n.$$

Suposição 3.7 (Sobre ω). Existem números positivos ω_1 e ω_2 tais que $\omega_1 \leq \omega \leq \omega_2$.

Nosso objetivo agora é provar a convergência global do algoritmo a um par de KKT do problema (3.1) qual sejam as atualizações sobre os parâmetros satisfazendo as suposições acima. Todos os resultados aqui expostos foram obtidos de [4, 16].

Iniciamos com um importante lema cujo resultado estabelece que o algoritmo FDIPA nunca falha. Isto é, as soluções d_α , λ_α , d_β e λ_β obtidas pelo algoritmo são determinados de maneira única. Destacamos antes que todas essas soluções foram obtidas de uma mesma matriz $(m+n) \times (m+n)$ dada por

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

chamada de matriz do sistema FDIPA.

Lema 3.1. *Sejam dados $x \in \Omega_\alpha$, $B \in \mathbb{R}^{n \times n}$ simétrica definida positiva e $\lambda \in \mathbb{R}^m$ não negativo tal que $\lambda_i > 0$ sempre que $g_i(x) = 0$. A matriz $M(x, \lambda, B)$ do sistema FDIPA é não singular.*

Demonstração. É suficiente provar que $(d, \mu) = (0, 0)$ é a única solução do sistema homogêneo

$$\begin{bmatrix} B & \nabla g(x) \\ \Lambda \nabla g^T(x) & G(x) \end{bmatrix} \begin{bmatrix} d \\ \mu \end{bmatrix} = 0, \quad (3.18)$$

aonde $d \in \mathbb{R}^n$ e $\mu \in \mathbb{R}^m$. Mostraremos primeiro que $d = 0$ e em seguida $\mu_i = 0$, $\forall i \in \{1, \dots, m\}$. Com efeito, o sistema corresponde a

$$Bd + \sum_{i=1}^m \mu_i \nabla g_i(x) = 0, \quad (3.19)$$

$$\lambda_i d \nabla g_i(x) + g_i(x) \mu_i = 0, \quad i = 1, \dots, m. \quad (3.20)$$

Como por hipótese $\lambda_i > 0$ quando $g_i(x) = 0$, podemos afirmar a partir da equação (3.20) que

- (a) $\nabla g_i(x)d = 0$, para todo $i \in I(x)$;
- (b) $\mu_i = 0$, para todo i tal que $\lambda_i = 0$.

Fazendo agora o produto escalar por d em ambos os lados de (3.19), obtemos a seguinte equação

$$d^t B d + \sum_{i=1}^m \mu_i d \nabla g_i(x) = 0. \quad (3.21)$$

Seja $\hat{I}(x)$ o conjunto formado pelas restrições $i \notin I(x)$ tais que $\lambda_i > 0$. Em virtude dos itens (a) e (b), a equação (3.21) pode ser expressa por

$$d^t B d + \sum_{i \in \hat{I}(x)} \mu_i d \nabla g_i(x) = 0. \quad (3.22)$$

e conjuntamente com (3.20), por

$$d^t B d - \sum_{i \in \hat{I}(x)} \frac{\mu_i^2}{\lambda_i} g_i(x) = 0. \quad (3.23)$$

Como B é definida positiva e $g(x) \leq 0$ então $d = 0$ e $\mu_i = 0$ para todo $i \in \hat{I}(x)$. Visto também que $\mu_i = 0$ quando $\lambda_i = 0$, resta então concluir que $\mu_i = 0$ para o último caso em que $i \in I(x)$. Considerando a suposição 3.4 e a equação (3.19) temos $\mu_i = 0$ para este caso e, portanto, $\mu_i = 0$ para todo $i \in \{1, \dots, m\}$.

□

Uma vez que x pertence a um conjunto compacto e ambos λ e B são limitados (suposições 3.5 e 3.6) então a matriz M é limitada e distante de zero. Isto implica que

as soluções d_α , λ_α , d_β e λ_β são limitadas inferiormente. Este fato será utilizado ao longo desta seção.

Deve-se ao lema a seguir uma justificativa para o caso particular em que o algoritmo 3.1 termina ao primeiro item do passo 1.

Lema 3.2. *Se $d_\alpha^k = 0$ então $\nabla f(x^k) = 0$ e x^k é um ponto de KKT do problema (3.1).*

Demonstração. Tomando $d_\alpha^k = 0$ na equação (3.9) resulta $G(x^k)\lambda_\alpha^k = 0$. Sendo x^k um ponto estritamente viável então $\lambda_\alpha^k = 0$ e obtemos juntamente com a equação (3.3) que $\nabla f(x^k) = 0$. Assim, as duas primeiras condições que caracterizam um ponto de KKT são satisfeitas. A última condição decorre da suposição 3.5. \square

Consideramos de agora em diante somente o caso em que $d_\alpha^k \neq 0$ e o algoritmo avança após o passo 1. Nestes casos, esta direção é de descida da função potencial f conforme o resultado do próximo lema.

Lema 3.3. $(d_\alpha^k)^t \nabla f(x^k) \leq -(d_\alpha^k)^t B_k d_\alpha^k$.

Demonstração. O sistema (3.9) corresponde a

$$\begin{aligned} -d_\alpha^k \nabla f(x^k) &= -(d_\alpha^k)^t B_k d_\alpha^k - d_\alpha^k \nabla g(x^k) \lambda_\alpha^k \\ \lambda_\alpha^k \nabla g(x^k) d_\alpha^k &= -\lambda_\alpha^k \Lambda_k^{-1} G(x^k) \lambda_\alpha^k \end{aligned}$$

onde que por substituição direta, $(d_\alpha^k)^t \nabla f(x^k) = -(d_\alpha^k)^t B_k d_\alpha^k + (\lambda_\alpha^k)^t \Lambda_k^{-1} G(x^k) \lambda_\alpha^k$ e ainda como $G(x) \leq 0$, o lema segue. \square

Lema 3.4. *A direção d^k satisfaz $(d^k)^t \nabla f(x^k) \leq \varepsilon (d_\alpha^k)^t \nabla f(x^k)$.*

Demonstração. Como $d^k = d_\alpha^k + \rho^k d_\beta^k$ e por conseguinte $(d^k)^t \nabla f(x^k) = (d_\alpha^k)^t \nabla f(x^k) + \rho^k (d_\beta^k)^t \nabla f(x^k)$ então obtemos a prova do lema pela implementação ρ^k dada no passo 1 do algoritmo 3.1. \square

Proposição 3.1. *Seja $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^1 tal que $\nabla \phi$ satisfaz a condição de Lipschitz em algum conjunto $\Gamma \subset \mathbb{R}^n$. Se $[x, y] \subset \Gamma$ então $\phi(y) \leq \phi(x) + (y - x)^t \nabla \phi(x) + k \|y - x\|^2$.*

Demonstração. A demonstração desta proposição decorre diretamente do teorema do valor médio e pode ser encontrada em [4]. \square

O importante lema abaixo garante a existência de um t satisfazendo as condições (3.15) - (3.15).

Lema 3.5. *Para quaisquer $x \in \Omega_a$ e d do algoritmo FDIPA, existe um $\tau > 0$ tal que as condições (3.15) - (3.15) são satisfeitas para todo $t \in [0, \tau]$.*

Lema 3.6. *Qualquer ponto de acumulação x^* da sequência gerada pelo algoritmo é um ponto estacionário do problema. Além disto, $(x^*, \lambda_\alpha(x^*))$ constitui um par estacionário, aonde $\lambda_\alpha^* = \lambda_\alpha(x^*)$.*

Teorema 3.2. *Qualquer ponto de acumulação x^* da sequência gerada pelo algoritmo é um ponto de Karush-Kuhn-Tucker do problema.*

3.3 Atualizações para a matriz B_k

Diferentes atualizações para a matriz B_k do algoritmo FDIPA podem ser implementadas. Indicamos nesta seção quatro destas que fornecem uma matriz simétrica definida positiva e, à exceção da atualização BFGS, todas satisfazendo a suposição 3.6.

3.3.1 $B = I$

A primeira é de ordem 1 e constante para todo k , na qual implementa-se a cada iteração do algoritmo a atualização $B_k := I$, onde I é a matriz identidade $n \times n$.

Não há estudos teóricos em relação a ordem de convergência para o FDIPA munida desta atualização - sendo provavelmente, não mais do que linear.

Esta atualização é indicada principalmente a problemas que não necessitam de grande precisão. A principal vantagem consiste no baixo custo computacional uma vez que não necessita do cálculo de derivadas de nenhuma ordem.

3.3.2 Atualização BFGS

A atualização BFGS (*Broyden-Fletcher-Goldfarb-Shano*) consiste na aproximação quase-Newton da matriz hessiana $H(x^k, \lambda^k)$, definida pela regra de atualização

$$B^{k+1} := B^k + \frac{\gamma\gamma^t}{\delta^t\gamma} - \frac{B^k\delta^t\delta B^k}{\delta^t B^k \delta}, \quad (3.24)$$

aonde

$$\delta = x^{k+1} - x^k \quad \text{e} \quad \gamma = \nabla_x L(x^{k+1}, \lambda^k) - \nabla_x L(x^k, \lambda^k).$$

Como não há garantia de que a matriz hessiana é em geral definida positiva num ponto de KKT (condição esta exigida pelo algoritmo FDIPA), aplica-se à atualização a seguinte modificação proposta por Powell [18]:

Se $\delta^t\gamma < 0, 2\delta^t B^k \delta$, toma-se

$$\phi = \frac{0,8\delta^t B^k \delta}{\delta^t B^k \delta - \delta^t\gamma}$$

e implementa-se

$$\gamma = \phi\gamma + (1 - \phi)B^k\delta.$$

Em seguida, aplica-se a atualização inicial (3.24).

Powell [19] prova que métodos quase-Newton possuem ordem de convergência superlinear de dois passos, i.e.,

$$\lim_{k \rightarrow \infty} \frac{x^{k+2} - x^*}{x^k - x^*} = 0$$

e dependendo da atualização para λ segue que FDIPA com a atualização BFGS também possui esta ordem de convergência. A atualização específica para λ é dada em [4].

A matriz B^{k+1} da atualização com a modificação de Powell é definida positiva para todo k . No entanto, não é claro quando esta matriz satisfaz a suposição 3.6.

3.3.3 Outras atualizações

Propomos também a atualização de segunda ordem $B_k = H(x^k, \lambda^k)$. Esta deve ser implementada somente aos problemas em que $H(x^k, \lambda^k)$ é definida positiva para todo k .

A principal atualização proposta neste texto e que comprova a flexibilidade do FDIPA no sentido de ser eficiente para resolução de problemas de complementaridade é dada por

$$B_k = \nabla f(x^k) + (\nabla f(x^k))^T$$

e deriva da regra de atualização para os multiplicadores de Lagrange detalhada no capítulo 5. A importância desta atualização consiste no fato que possibilita mostrar a equivalência entre os métodos FDIPA e FDA, isto é, as direções de busca dos métodos são iguais.

4 FDA

4.1 Descrição do Método FDA

O algoritmo FDA-NCP que tratamos neste capítulo é um método iterativo para resolução do sistema (2.22) via sequência de pontos interiores e direções viáveis.

Resolve-se de modo recursivo um sistema obtido de (2.22) por uma perturbação ao passo que em cada iteração uma direção viável e de descida da função potencial

$$\phi(x) = x^t F(x),$$

é calculada. Em seguida, efetua-se uma busca linear inexata sobre essa direção com o objetivo de encontrar um comprimento de passo α_k de modo que a atualização $x^{k+1} := x^k + \alpha_k d^k$ seja um ponto viável e mais próximo da solução.

Para fixar as ideias, veremos que a direção obtida diretamente do sistema inicial é de descida mas não necessariamente viável. Daí a necessidade de modificá-lo para que uma direção com ambas as propriedades seja alcançada.

Antes de descrevermos o algoritmo, por razão de ordem prática, consideremos inicialmente o NCP como em (2.22) reescrito abaixo com a seguinte notação:

$$H(x) = 0, \tag{4.1}$$

onde $H(x) = x \bullet F(x)$ e $x \in \Omega$. Em decorrência desta notação, a matriz jacobiana de $H \in \mathbb{R}^{n \times 1}$ é dada por

$$\nabla H(x) = \text{diag}(F(x)) + \text{diag}(x) \nabla F(x),$$

onde que dado $v \in \mathbb{R}^n$, $\text{diag}(v) \in \mathbb{R}^{n \times n}$ denota a matriz diagonal tal que $[\text{diag}(v)]_{ii} = v_i$ e

$$\nabla F(x) = \begin{bmatrix} \frac{\partial F_i(x)}{\partial x_j} \end{bmatrix}_{n \times n}.$$

Iniciamos a descrição do algoritmo aplicando o método de Newton ao sistema (4.1), de onde vem

$$\nabla H(x^k) [x^{k+1} - x^k] = -H(x^k) \tag{4.2}$$

Primeiro, afirmamos que pondo-se $d_1^k = x^{k+1} - x^k$ então esta deve ser uma direção de descida da função potencial $\phi(x) = x^t F(x)$. Em virtude da proposição 2.3 basta mostrarmos que $\nabla \phi(x^k) d_1^k < 0$. Ora, como $\nabla \phi(x) = E^t \nabla H(x^k)$ onde $E = (1, \dots, 1) \in \mathbb{R}^n$ e d_1^k é solução do sistema acima, então

$$\nabla \phi(x^k) d_1^k = [E^t \nabla H(x^k)] [\nabla H^{-1}(x^k)(-H(x^k))] = -E^t H(x^k)$$

e sendo os vetores E^t e $H(x)$ ambos positivos, a afirmação procede.

Muito embora d_1^k seja um vetor de descida, não podemos afirmar que seja viável. Por certo, como (4.2) é equivalente a $[F_i(x^k)e_i + x_i^k \nabla F_i(x^k)] d_1^k = -x_i^k F_i(x^k)$, $i = 1, \dots, n$, e uma vez que

- (i) $F_i(x^k)d_{1i}^k = 0$ quando $x_i^k = 0$ e
- (ii) $\nabla F_i(x^k)d_1^k = 0$ quando $F_i(x^k) = 0$,

então d_1^k não satisfaz as condições dadas pela proposição (2.3) e assim é inconteste afirmar que é viável.

Uma perturbação adequada ao sistema inicial pode ser feita a fim de obter uma outra direção de busca que seja de descida e também viável. A saber, provemos que o sistema modificado dado por

$$\nabla H(x^k)d^k = -H(x^k) + \rho^k E \quad (4.3)$$

em que $E = (1, \dots, 1) \in \mathbb{R}^n$ e ρ^k convenientemente escolhido é tal que um vetor solução possui ambas as propriedades.

Mostremos que um $\rho^k > 0$ implica a viabilidade de d^k . Com efeito, (4.3) é equivalente a $[F_i(x^k)e_i + x_i^k \nabla F_i(x^k)] d^k = -x_i^k F_i(x^k) + \rho^k$, $i = 1, \dots, n$. E, daí

- (i) Se $x_i^k = 0$ então $d_i^k = \frac{\rho^k}{F_i(x^k)} > 0$,
- (ii) Se $F_i(x^k) = 0$ então $\nabla F_i(x^k)d^k = \frac{\rho^k}{x_i^k} > 0$.

Portanto, pela proposição (2.3) segue que d^k é viável em Ω como queríamos mostrar e ilustrado na figura a seguir em que x^* é a solução de um exemplo específico e $i = 2$.

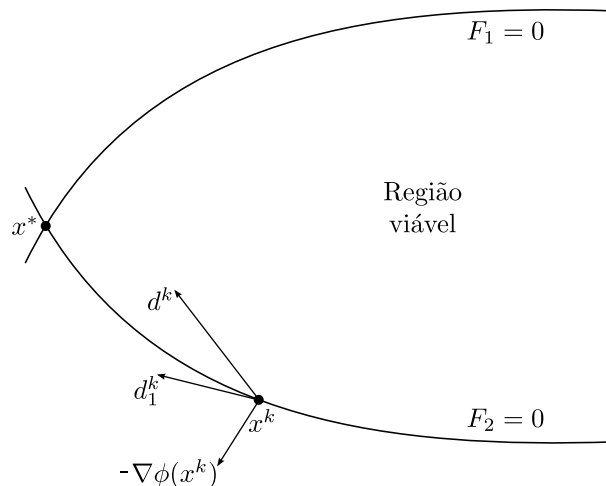


Figura 9 – Algoritmo FDA

Resta ainda mostrar que d^k é também direção de descida. Tendo em vista que d_1^k é uma direção de descida e d^k é uma deflexão deste e proporcional a ρ^k então podemos impor um limite adequado a ρ^k a fim de garantir que d^k também seja de descida. Mostremos que isto de fato é possível como segue. Para tanto, notemos que como $\nabla\phi(x^k)d^k = [E^t\nabla H(x^k)]d^k$, então

$$\nabla\phi(x^k)d^k = E^t[-x^k \bullet F(x^k) + \rho^k E] = [-\phi(x^k) + n\rho^k] = -\phi(x^k) \left[1 - \frac{n\rho^k}{\phi(x^k)} \right].$$

Em seguida, fazendo a identificação $\rho^k = \rho_0[\phi(x^k)^\beta/n]$, obtemos

$$\nabla\phi(x^k)d^k = -\phi(x^k) \left[1 - \rho_0\phi(x^k)^{\beta-1} \right] < 0,$$

onde a desigualdade é válida caso $\rho_0\phi(x^k)^{\beta-1} < 1$, com $\beta \in (1, 2]$ e $\rho_0 \in (0, 1)$.

Por fim, concluímos que d^k é uma direção viável em Ω e de descida da função potencial desde que tenhamos $\rho_0\phi(x^k)^{\beta-1} < 1$ e $\nabla H(x^k)$ não singular em Ω^0 .

Segue agora a descrição simplificada do algoritmo FDA baseada nas ideias discutidas acima.

Algoritmo 4.1 (Algoritmo FDA).

Parâmetros: $\epsilon > 0$; α, η e $\nu \in (0, 1)$; $\beta \in (1, 2]$ e $\rho_0 < \min\{1; 1/(c^{\beta-1})\}$.

Dados Iniciais: $x^0 \in \Omega^0$ tal que $\phi(x^0) < c$ e $k = 0$.

Início

Passo 1: Direção de Busca.

Faça $\rho^k := \rho_0[\phi(x^k)^\beta/n]$.

Resolva o sistema: $\nabla H(x^k)d^k = -H(x^k) + \rho^k E$.

Passo 2: Busca Linear.

Escolha o tamanho do passo t^k como sendo o primeiro valor da sequência $\{1, \nu, \nu^2, \nu^3, \dots\}$ que satisfaz:

i. $x^k + t^k d^k > 0$;

ii. $F(x^k + t^k d^k) > 0$;

iii. $\phi(x^k + t^k d^k) < \phi(x^k) + t^k \eta \nabla\phi(x^k)d^k$.

Passo 3: Atualização dos Dados.

Faça $x^{k+1} := x^k + t^k d^k$ e $k := k + 1$.

Passo 4: Critério de Parada.

Se $\phi(x^k) < \epsilon$, pare. Senão, volte ao passo 1.

Fim

5 NOVA ATUALIZAÇÃO

5.1 Introdução

Nosso objetivo nesta seção é reescrever o problema de complementaridade não linear como um problema de minimização com restrições.

5.2 Reescrevendo o NCP como um problema de minimização

Reescrevemos o NCP como um problema de minimização com restrições por

$$\begin{aligned} &\text{minimizar } f(x) \\ &\text{sujeito a } x \in \Omega \end{aligned} \tag{5.1}$$

onde $f: \mathbb{R}^n \rightarrow \mathbb{R}$ e $\Omega \subseteq \mathbb{R}^n$ são dados por

$$f(x) = \sum_{i=1}^n x_i F_i(x) \quad \text{e} \quad \Omega = \{x \in \mathbb{R}^n; -F(x) \leq 0 \text{ e } -x \leq 0\}.$$

Podemos também definir o conjunto viável Ω pela desigualdade $g(x) \leq 0$, sendo $g: \mathbb{R}^n \rightarrow \mathbb{R}^{2n}$ definida por $g(x) = (-F(x), -x)$.

As condições de otimalidade de primeira ordem de Karush-Kuhn-Tucker (KKT) correspondentes ao problema (5.1) são expressas desta vez por:

$$\nabla f(x) - (\nabla F(x))^t \lambda^1 - \Lambda^2 = 0, \tag{5.2}$$

$$-F(x) \lambda^1 = 0, \tag{5.3}$$

$$-X \lambda^2 = 0, \tag{5.4}$$

$$-F(x) \leq 0, \tag{5.5}$$

$$-X \leq 0, \tag{5.6}$$

$$\lambda^1 \geq 0, \tag{5.7}$$

$$\lambda^2 \geq 0, \tag{5.8}$$

aonde λ^1 e λ^2 são os os multiplicadores de Lagrange associados às restrições $-F(x)$ e $-x$, respectivamente.

5.3 Resolução do NCP modificado via FDIPA

Resolve-se o problema (5.1) pelo método FDIPA como descrito em (3.1). A particularidade neste caso é que $f(x) = \sum_i^n x_i F_i(x)$ e $g(x) = (-x, -F(x))$.

O último capítulo deste trabalho será dedicado à comparação numérica entre as resoluções do NCP - direta mediante FDA, e indireta como um problema de minimização via FDIPA.

5.4 Uma nova atualização para λ

Definição 5.1. Dizemos que uma função $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ é monótona em \mathbb{R}^n quando para quaisquer $x, y \in \mathbb{R}^n$, vale

$$(x - y)^t(F(x) - F(y)) \geq 0.$$

O próximo teorema estabelece condições necessárias e suficientes sob F que um dado ponto $(x^*, \lambda^{1*}, \lambda^{2*})$ satisfazendo as condições de KKT do problema de minimização (5.1) deve cumprir para que x^* seja uma solução do problema de complementaridade associado.

Teorema 5.1. Seja $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ diferenciável e monótona em algum intervalo aberto contendo \mathbb{R}_+^n . Se um ponto $(x^*, \lambda^{1*}, \lambda^{2*})$ satisfaz as condições (5.2) - (5.8) então x^* é solução do problema de complementaridade. Reciprocamente, se x^* é solução do problema de complementaridade então $x^*, \lambda^{1*} = x^*$ e $\lambda^{2*} = F(x^*)$ formam uma tríplice que satisfaz as condições (5.2) - (5.8).

Demonstração. (\Rightarrow). Devemos provar que se um ponto $(x^*, \lambda^{1*}, \lambda^{2*})$ é de KKT então $x^* \bullet F(x^*) = 0$. Com efeito, tendo em vista que $\nabla f(x) = F(x) + (\nabla F(x))^T x$, segue da equação (5.2),

$$F(x^*) + (\nabla F(x^*))^T(x^* - \lambda^{1*}) - \lambda^{2*} = 0.$$

Agora, multiplicando por $(x^* - \lambda^{1*})^T$, obtemos

$$(x^* - \lambda^{1*})^T F(x^*) + (x^* - \lambda^{1*})^T (\nabla F(x^*))^T(x^* - \lambda^{1*}) - (x^* - \lambda^{1*})^T \lambda^{2*} = 0.$$

Ainda, uma vez que $-F^t(x)\lambda^1 = 0$ e $-x^t\lambda^2 = 0$, então

$$x^{*T} F(x^*) + (x^* - \lambda^{1*})^T (\nabla F(x^*))^T(x^* - \lambda^{1*}) + (\lambda^{1*})^T \lambda^{2*} = 0.$$

Notemos que das condições KKT e também como F por hipótese é monótona então esta é uma soma de termos não negativos, donde concluímos que $x^{*T} F(x^*) = 0$. Daí, obtemos o resultado esperado.

(\Leftarrow). Supondo $\lambda^{1*} = x^*$, $\lambda^{2*} = F(x^*)$ e x^* uma solução do NCP, obtemos de imediato as relações (5.2) - (5.8). \square

Por conseguinte, o teorema garante a convergência global da sequência gerada pelo algoritmos FDIPA aplicados ao problema (5.1) para uma solução do NCP (2.22).

Além disso, se um ponto x^* é uma solução do NCP, a recíproca nos fornece ainda uma família de atualizações para o multiplicadores de Lagrange λ^1 e λ^2 do método FDIPA.

Inicialmente, notemos que a matriz hessiana B do problema de minimização (5.1) é dada por

$$B = \nabla F(x) + (\nabla F(x))^T + \sum_{i=1}^n (x_i - \lambda_i^1) \nabla^2 F_i(x).$$

Considerando a atualização $\lambda^1 = x$ indicada pelo teorema (5.1), a hessiana assume portanto a seguinte forma:

$$B = \nabla F(x) + (\nabla F(x))^T. \quad (5.9)$$

Note que esta matriz modificada é simétrica e se ainda F é monótona (portanto ∇F definida positiva) segue que B satisfaz as condições exigidas pelo algoritmo FDIPA.

Considerando ambas as atualizações $\lambda^1 = x$ e $\lambda^2 = F(x)$, a matriz do método FDIPA fica da seguinte forma:

$$\begin{bmatrix} \nabla F(x) + (\nabla F(x))^T & -(\nabla F(x))^T & -I \\ X\nabla F(x) & -\mathbb{F} & 0 \\ -\mathbb{F} & 0 & -X \end{bmatrix}, \quad (5.10)$$

onde I é a matriz identidade de ordem n , \mathbb{F} e X são matrizes diagonais tais que $[\mathbb{F}]_{ii} = F_i$ e $[X]_{ii} = x_i$, para $i = 1, \dots, n$.

Consideremos agora a matriz B modifica dada em (5.9). Afirmamos que as direções de descida e restauração dos métodos FDIPA e FDA são as mesmas. Isto é, as direções obtidas pelo método FDA aplicado a um problema de complementaridade e as direções do método FDIPA (tomando as atualizações sugeridas anteriormente) aplicado ao mesmo problema adaptado como um de minimização coincidem.

Provemos este fato como segue. Inicialmente, obtemos a direção de descida do FDIPA de forma análoga a (3.9) através do sistema:

$$\begin{bmatrix} \nabla F(x) + (\nabla F(x))^T & -(\nabla F(x))^T & -I \\ X\nabla F(x) & -\mathbb{F} & 0 \\ -\mathbb{F} & 0 & -X \end{bmatrix} \begin{bmatrix} d_\alpha \\ \lambda_\alpha^1 \\ \lambda_\alpha^2 \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ 0 \\ 0 \end{bmatrix}.$$

Obtemos em seguida, efetuando a multiplicação acima que

$$\begin{aligned} (\nabla F(x) + (\nabla F(x))^T)d_\alpha - (\nabla F(x))^T\lambda_\alpha^1 - I\lambda_\alpha^2 &= -\nabla f(x), \\ X\nabla F(x)d_\alpha - \mathbb{F}\lambda_\alpha^1 &= 0, \\ -\mathbb{F}d_\alpha - X\lambda_\alpha^2 &= 0. \end{aligned}$$

Ainda, multiplicando a primeira linha do sistema acima pela matriz X e em seguida substituindo os termos $\mathbb{F}\lambda_\alpha^1$ e $X\lambda_\alpha^2$ obtidos da segunda e terceira linhas, obtemos

$$\left[X\nabla F(x) + X(\nabla F(x))^T + X(\nabla F(x))^T\mathbb{F}^{-1}X\nabla F(x) + \mathbb{F} \right] d_\alpha = -X\nabla f(x). \quad (5.11)$$

Lembrando que nosso objetivo é mostrar que d_α é também uma direção de descida do método FDA. Mais precisamente, fazendo a identificação $M = \mathbb{F} + X\nabla F$, devemos mostrar que $Md_\alpha = -x \bullet F$. Pois bem, substituindo M na equação (5.11), resulta

$$\left[M + X\nabla F^T \mathbb{F}^{-1} M \right] d_\alpha = -XM^T E,$$

aonde $M^T = \mathbb{F} + (\nabla F)^T X$ e $M^T E = \nabla f(x)$. Por sua vez, substituindo M^T na equação anterior, chegamos a

$$\left[\mathbb{F} + X\nabla F^T \right] \mathbb{F}^{-1} M d_\alpha = - \left[\mathbb{F} + X\nabla F^T \right] X E.$$

Finalmente, decorre da equação anterior que $\mathbb{F}^{-1} M d_\alpha = -X E$ e, portanto, daí concluímos que $M d_\alpha = -x \bullet F$ como havíamos afirmado anteriormente.

Resta agora mostrar que as direções de restauração também coincidem. A direção neste caso agora é obtido através do sistema

$$\begin{bmatrix} \nabla F(x) + (\nabla F(x))^T & -(\nabla F(x))^T & -I \\ X\nabla F(x) & -\mathbb{F} & 0 \\ -\mathbb{F} & 0 & -X \end{bmatrix} \begin{bmatrix} d_\beta \\ \lambda_\beta^1 \\ \lambda_\beta^2 \end{bmatrix} = - \begin{bmatrix} 0 \\ X\omega_1 \\ \mathbb{F}\omega_2 \end{bmatrix}$$

onde tomamos as atualização para ω_1 e ω_2 por $X\omega_1 = \rho^k E$ e $\mathbb{F}\omega_2 = \rho^k E$, respectivamente. De modo análogo ao caso anterior chegamos a $Md_\beta = -x \bullet F$. Provamos assim a afirmação anterior de que as direções de fato coincidem.

6 Testes Numéricos

Neste último capítulo aplicamos os métodos até aqui estudados numa coletânea de problemas de complementaridade. Resolvemos via FDIPA cada um destes adaptados como problemas de minimização com as diferentes atualizações para a matriz B_k discutidas neste trabalho.

Apresentamos uma tabela com os resultados obtidos no MatLab onde estão discriminados a quantidade de iterações necessárias para a convergência (caso ocorra) e o número total de buscas lineares realizadas.

1. Problema da Meia Lua:

Seja $F: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definida por

$$F(x_1, x_2) = \begin{bmatrix} 1 - \frac{(x_1 - 1, 5)^2}{(1, 5)^2} - (x_2 - 1, 5)^2 \\ -1 + \frac{(x_1 - 3)^2}{(1, 5)^2} + (x_2 - 1, 5)^2 \end{bmatrix}.$$

- Conjunto dos pontos viáveis:

$$\Omega = \left\{ (x_1, x_2) \in \mathbb{R}_+^2; 1 - \frac{(x_1 - 3)^2}{(1, 5)^2} \leq (x_2 - 1, 5)^2 \leq 1 - \frac{(x_1 - 1, 5)^2}{(1, 5)^2} \right\}$$

- Pontos Iniciais: $x_1^0 = [1.5; 2.2] \in \Omega_0$ e $x_2^0 = [1.1; 1.1] \in \Omega_0$

- Soluções: $x^* = [2.25; 2.36]$ e $x^{**} = [2.25; 0.36]$ com $x_1^0 \rightarrow x^*$ e $x_2^0 \rightarrow x^{**}$.

- Tabela com os resultados obtidos:

Tabela 1 – Problema da Meia Lua

Ponto Inicial	x_1^0		x_2^0	
Algoritmo	It	BL	It	BL
FDA	13	14	13	19
FDIPA-H	8	10	15	21
FDIPA-I	7	7	11	13
FDIPA-B	8	10	18	40
FDIPA-C	7	9	10	16

It - Número de iterações para a convergência; Bl - Número total de buscas lineares realizadas pelo algoritmo.

onde FDIPA-H denota o método FDIPA com a atualização para a matriz B_k dada pela hessiana, FDIPA-I com a atualização dada pela identidade, FDIPA-B pela atualização BFGS e FDIPA-C pela atualização de complementaridade.

- Gráfico com as iterações dos métodos FDA e FDIPA-C:

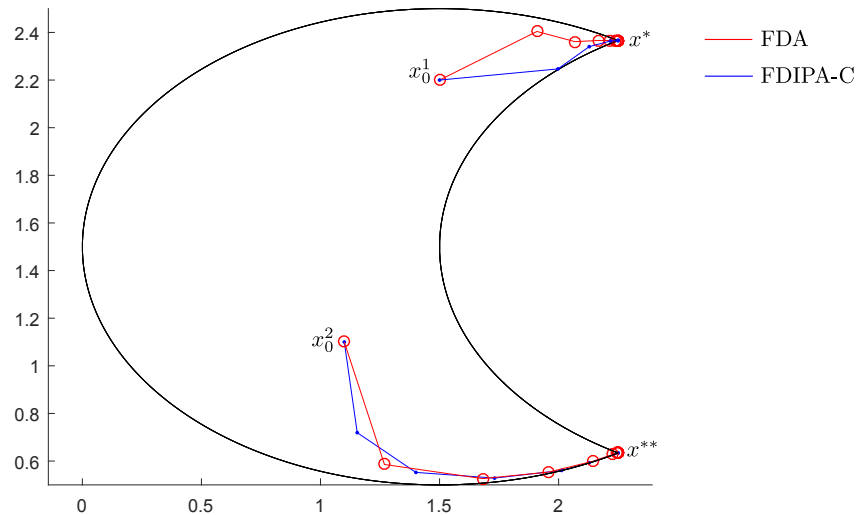


Figura 10 – Problema da Meia Lua

2. Problema do Peixe.

Seja $F: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definida por

$$F(x_1, x_2) = \begin{bmatrix} x_2 - 2(x_1 - 1)^2 \\ -x_1 - x_2^2 + 1 \end{bmatrix}.$$

- Conjunto dos pontos viáveis:

$$\Omega = \{(x_1, x_2) \in \mathbb{R}_+^2; x_1 \leq -x_2^2 + 1 \text{ e } x_2 \geq 2(x_1 - 1)^2\}$$

- Pontos Iniciais: $x_1^0 = [0.6; 0.6] \in \Omega_0$ e $x_2^0 = [0.7; 0.4] \in \Omega_0$.
- Soluções: $x^* = [0.37; 0.79]$ e $x^{**} = [1, 0]$ com $x_1^0 \rightarrow x^*$ e $x_2^0 \rightarrow x^{**}$.
- Tabela com os resultados obtidos:

Tabela 2 – Problema do Peixe

Ponto Inicial	x_1^0		x_2^0	
Algoritmo	It	BL	It	BL
FDA	10	10	452	9287
FDIPA-H	6	6	61	549
FDIPA-I	6	6	68	510
FDIPA-B	13	35	-	-
FDIPA-C	8	10	70	616

Note que o método FDIPA com a atualização BFGS não convergiu tomando como ponto inicial x_2^0 .

- Gráfico com as iterações dos métodos FDA e FDIPA-C:

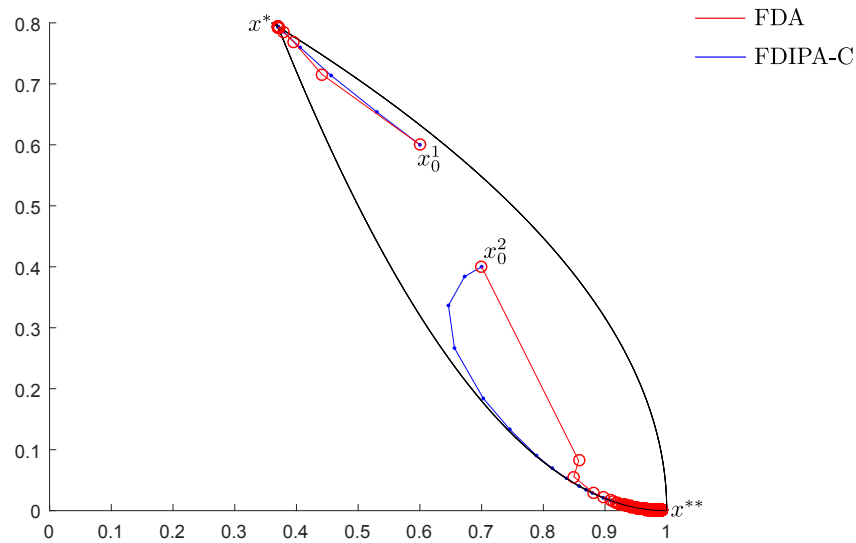


Figura 11 – Problema do Peixe

3. O problema de Kojima-Josephy.

Seja $F: \mathbb{R}^4 \rightarrow \mathbb{R}^4$ definida por

$$F(x) = \begin{bmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_2^2 + x_1 + 3x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 3x_4 - 9 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{bmatrix}$$

- Ponto Inicial: $x^0 = [1; 1; 1; 1]$.
- Solução: $x^* = [1.22; 0; 0; 0.5]$ com $F(x^*) = [0; 3.22; 5; 0]$.
- Tabela com os resultados obtidos:

Tabela 3 – Problema de Kojima-Josephy

Ponto Inicial	x^0	
Algoritmo	It	BL
FDA	19	19
FDIPA-H	8	270
FDIPA-I	9	273
FDIPA-B	9	279
FDIPA-C	3	27

4. O Problema de Kojima-Shindo.

Seja $F: \mathbb{R}^4 \rightarrow \mathbb{R}^4$ definida por

$$F(x) = \begin{bmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_2^2 + x_1 + 10x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{bmatrix}$$

- Ponto Inicial: $x^0 = [1; 0.01; 3; 0.01]$.
- Soluções: $x^* = [1.22; 0; 0; 0.5]$ com $F(x^*) = [0; 3.22; 0; 0]$ e $x^{**} = [1; 0; 3; 0]$ com $F(x^{**}) = [0; 31; 0; 4]$.
- Tabela com os resultados obtidos:

Tabela 4 – Problema de Kojima-Shindo

Ponto Inicial	x_1^0	
Algoritmo	It	BL
FDA	12	12
FDIPA-H	2	19
FDIPA-I	5	219
FDIPA-B	8	382
FDIPA-C	2	19

5. Problema de Mathiesen modificado [7].

Seja $\mathbb{R}^4 \rightarrow \mathbb{R}^4$ definida por

$$F(x) = \begin{bmatrix} -x_2 + x_3 + x_4 \\ x_1 - \frac{4.5x_3 + 2.7x_4}{x_2 + 1} \\ 5 - x_1 - \frac{0.5x_3 + 0.3x_4}{x_3 + 1} \\ 3 - x_1 \end{bmatrix}$$

- Ponto Inicial: $x^0 = [2.9; 2; 0.01; 3]$.
- Soluções: $x^* = [x_1^*; 0; 0; 0]$, onde $x_1^* \in [0, 3]$
e $x^0 \rightarrow x^* = [2.508; 0; 0; 0]$.
- Tabela com os resultados obtidos:

Tabela 5 – Problema de Mathiesen modificado

Ponto Inicial	x_1^0	
Algoritmo	It	BL
FDA	14	15
FDIPA-H	11	287
FDIPA-I	10	266
FDIPA-B	9	225
FDIPA-C	9	259

6. Problema de complementaridade não linear número 9 do Artigo [22].

A função F é definida por

$$F(x) = \begin{bmatrix} x_1 - 2 \\ x_2^3 + x_2 - x_3 + 3 \\ x_2 + 2x_3^3 + x_3 - 3 \end{bmatrix}$$

- Ponto Inicial: $x^0 = [3; 3; 3]$.
- Solução: $x^* = [2; 0; 1]$ com $F(x^*) = [0; 2; 0]$.
- Tabela com os resultados obtidos:

Tabela 6 – Problema de complementaridade não linear

Ponto Inicial	x_1^0	
	It	BL
FDA	34	54
FDIPA-H	13	339
FDIPA-I	12	370
FDIPA-B	12	282
FDIPA-C	10	256

7. Problema de Complementaridade não linear de número 12 do Artigo [22].

O NPC é definido pela função $F: \mathbb{R}^4 \rightarrow \mathbb{R}^4$ não linear dada por

$$F(x) = \begin{bmatrix} x_1^3 - 8 \\ x_2 + x_2^3 - x_3 + 3 \\ x_2 + 2x_3^3 + x_3 - 3 \\ x_4 + 2x_4^3 \end{bmatrix}$$

- Ponto Inicial: $x^0 = [3; 3; 3; 3]$.
- Solução: $x^* = [2; 0; 1; 0]$ com $F(x^*) = [0; 2; 0; 0]$.
- Tabela com os resultados obtidos:

Tabela 7 – Problema de complementaridade não linear

Ponto Inicial	x_1^0	
Algoritmo	It	BL
FDA	40	42
FDIPA-H	9	201
FDIPA-I	9	271
FDIPA-B	13	317
FDIPA-C	11	287

8. Problema de complementaridade linear com a matriz B singular.

O LPC é definido pela função linear dada por $F(x) = Bx + q$, com

$$B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad q = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- Ponto Inicial: $x^0 = [1; 1; 1]$.
- Soluções: $x^* = [0; \lambda^*; 0]$ com $\lambda^* \in [0, 1]$ e $x^{**} = [\lambda^{**}; 0; 0]$ com $\lambda^{**} \geq 0$, onde $x^0 \rightarrow [0; 0.5; 0]$.
- Tabela com os resultados obtidos:

Tabela 8 – NPC com a matriz B singular

Ponto Inicial	x_1^0	
Algoritmo	It	BL
FDA	14	14
FDIPA-H	10	338
FDIPA-I	12	346
FDIPA-B	12	368
FDIPA-C	9	317

7 Conclusão

Realizamos neste trabalho a descrição dos algoritmos FDIPA para resolução do problemas de otimização e FDA para problemas de complementaridade.

Propomos a reformulação de um problema de complementaridade em problema de otimização e a resolução através do FDIPA.

O FDIPA embora seja um algoritmo de otimização, mostrou ser eficiente também para resolução de problemas de complementaridade. Comprovamos este fato através de exemplos propostos no capítulo 5 onde realizamos a comparação entre os dois métodos.

Observamos ainda que o FDIPA com a atualização proposta para a matriz $B_k = \nabla f(x^k) + \nabla f(x^k)^T$ necessita de um número menor de iterações para convergência a uma solução do que o próprio FDA. Por exemplo, o Problema de Kojima-Shindo (número 4) com o FDIPA munido desta atualização necessita de duas iterações para a convergência enquanto que o FDA necessita de doze. Este exemplo e cada um dos exemplos resolvidos com ambos os métodos comprovam a flexibilidade do FDIPA para a resolução de problemas de complementaridade.

7.1 Trabalhos Futuros

- Fazer uma análise de sensibilidade de parâmetros do algoritmo FDIPA e FDA;
- Estudar a convergência dos algoritmos utilizando a busca linear de Wolfe;
- Estender o estudo de problema de complementaridade para complementaridade mista e aplicar em problemas da engenharia;
- Estudar o caso em que a matriz $\mathbb{F} + X\nabla F^T$ (Página 40) é inversível.

REFERÊNCIAS

- [1] Bazaraa, M. S. Shetty, C. M., “Nonlinear Programming, Theory and Algorithms”, John Wiley e Sons, New York, 1979.
- [2] Dennis, J. E., Schnabel R.B., “Numerical Methods for Unconstrained Optimization and Nonlinear Equations”. Prentice-Hall, 1983.
- [3] Herskovits, J., “A View on Nonlinear Optimization”, ed. Herskovits, J., Advances in Structural Optimization, Kluwer Academic Publishers, p. 71-116, 1995.
- [4] Herskovits, J., “Feasible direction interior-point technique for nonlinear optimization”, J. Optim. Theory Appl., v. 99, n. 1, p. 121–146, 1998.
- [5] Herskovits, J. Falcon, G. S. “On The Computer Implementation Of Feasible Direction Interior Point Algorithms For Nonlinear Optimization”. Structural and Multidisciplinary Optimization, Germany, v. 14, n.2-3, p. 165-172, 1997.
- [6] Izmailov, A. F., Solodov, M. V., “Condições de Otimalidade. Elementos de Análise Convexa e de Dualidade”. IMPA, 2007.
- [7] Jiang, H., Qi, L., “A new nonsmooth equations approach to nonlinear complementarity problems”. SIAM J. Control Optim, 35(1):178-193, 1997.
- [8] Josephy, N.H.: 1979, “Newton’s method for generalized equations”, Technical report 1965, Mathematics Research Center, University of Wisconsin, Madison.
- [9] Kelley, C. T., “Iterative Methods for Linear and Nonlinear Equations”. North Carolina State University. Society for Industrial and Applied Mathematics, SIAM. Philadelphia, 1995.
- [10] Luenberger, D. G., “Linear and Nonlinear Programming”, 2nd. Edition, Addison-Wesley, Reading, 1984.
- [11] Mangasarian, O. L. e Solodov, J. V., “Nonlinear Complementarity as Unconstrained and Constrained Minimization”, Mathematical Programming, Vol 62B, p. 277–297, 1993.
- [12] Mangasarian, O. L., “Equivalence of the Complementarity Problem to a System of Nonlinear Equations”, SIAM Journal on Applied Mathematics, Vol. 31, p. 89–92, 1976.
- [13] Maratos, N., “Exact penalty function algorithms for finite dimensional optimization problems with equality and inequality constraints”, Ph.D. Thesis, Imperial College of Science and Technology, London, 1978.
- [14] Mazorche, S. R., “Algoritmos Para Problemas de Complementaridade Não Linear”, Tese de Doutorado. Universidade Federal do Rio de Janeiro, 2007.
- [15] Nocedal, J., Wright, S. J., “Numerical optimization”. Springer Series in Operations Research, Springer-Verlag: New York, 1999.

- [16] Panier, Eliane R., Andre L., Tits, Herskovits J. (1988) “A QP-free, Globally Convergent, Locally Super-linearly Convergent Algorithm for Inequality Constrained Optimization”, *SIAM Journal of Control and Optimization*, 26(4), 788–811.
- [17] Pang, J. S. Gabriel, S. A., “NE/SQP: A robust algorithm for the nonlinear complementarity problem”, *Mathematical Programming*, 60 (1993), p. 295-337.
- [18] Powell, M. J. D., “Variable Metric Methods for Constrained Optimization”, in *Mathematical Programming - The State of the Art*, Edited by A. Bachem, M. Grottschet and B. Korte, Springer-Verlag, Berlin, 1983.
- [19] Powell, M. J. D., “The Convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Calculations”, in *Nonlinear Programming 3*, edited by O.L. Mangasarian, R. R. Meyer and S. M. Robinson, Academic Press, London, 1978.
- [20] Subramanian, P. K., “Gauss-Newton Methods for the Complementarity Problem”, *Journal of Optimization Theory and Applications*, Vol. 77, p. 467–482, 1993.
- [21] Yamashita, N., Fukushima, M., “On Stationary Points of the Implicit Lagrangian for Nonlinear complementary Problems”, *Journal of Optimization Theory and Applications*, Vol. 84, p. 653–663, 1995.
- [22] Yamashita, N., Dan, H., Fukushima, M., “On the identification of degenerate indices in the nonlinear complementarity problem with the proximal point algorithm”. *Mathematical Programming, Se. A*(99):377-397, 2004.