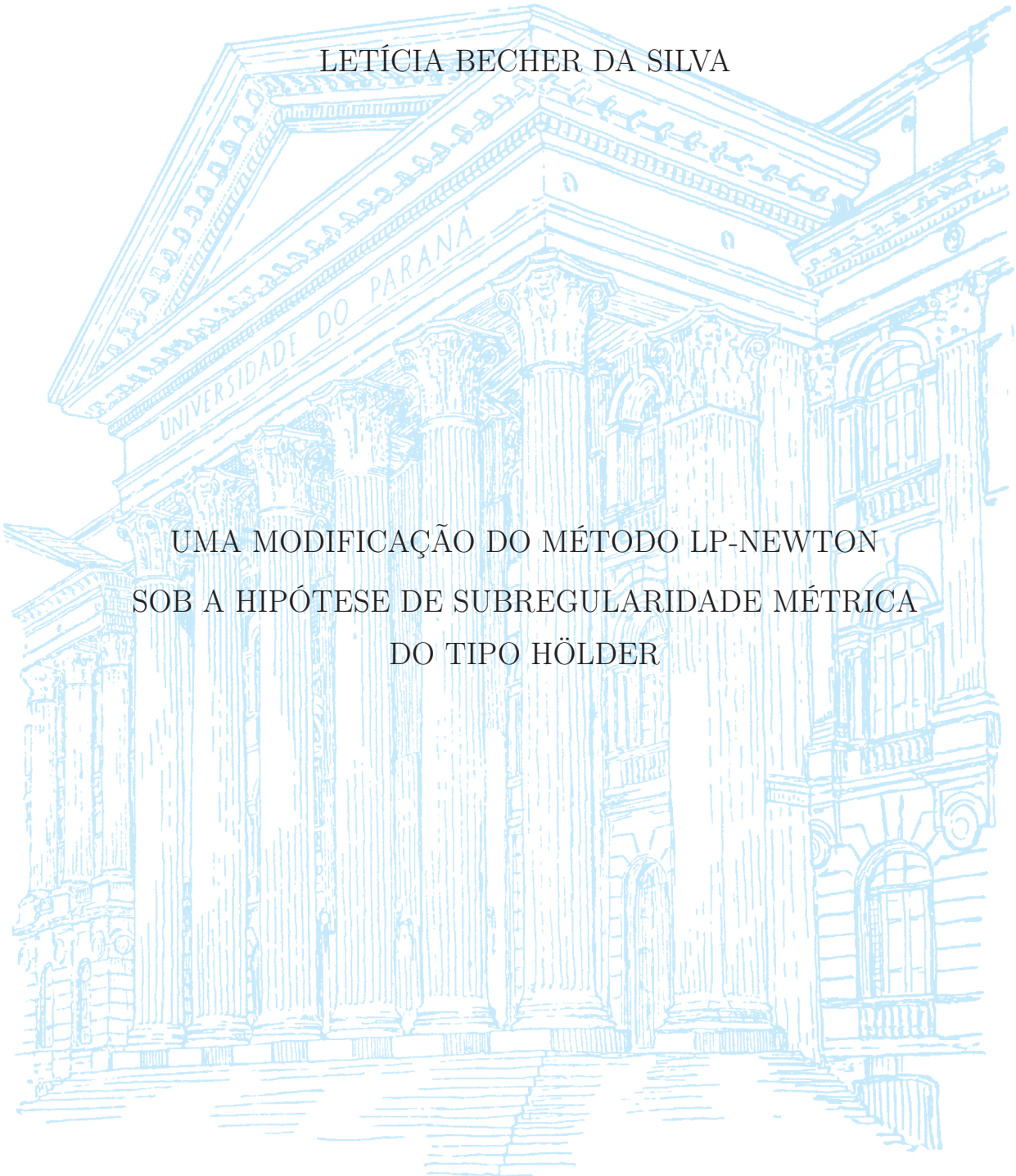
UNIVERSIDADE FEDERAL DO PARANÁ

LETÍCIA BECHER DA SILVA



UMA MODIFICAÇÃO DO MÉTODO LP-NEWTON SOB A HIPÓTESE DE SUBREGULARIDADE MÉTRICA DO TIPO HÖLDER

CURITIBA

2019

LETÍCIA BECHER DA SILVA

UMA MODIFICAÇÃO DO MÉTODO LP-NEWTON SOB A HIPÓTESE DE SUBREGULARIDADE MÉTRICA DO TIPO HÖLDER

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Matemática, no Curso de Pós-Graduação em Matemática, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Orientador: Prof. Dr. Alberto Ramos.

CURITIBA

2019

Catalogação na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

S586m Silva, Letícia Becher da
Uma modificação do método LP-Newton sob a hipótese de subregularidade métrica do tipo Hölder [recurso eletrônico] / Letícia Becher da Silva. – Curitiba, 2019.

Dissertação - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Matemática, 2019.

Orientador: Alberto Ramos.

1. Programação linear. 2. Equações não lineares. 3. convergência . I. Universidade Federal do Paraná. II. Ramos, Alberto. III. Título.

CDD: 519.72

Bibliotecária: Vanusa Maciel CRB- 9/1928

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIENCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO MATEMÁTICA - 40001016041P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MATEMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **LETÍCIA BECHER DA SILVA** intitulada: **UMA MODIFICAÇÃO DO MÉTODO LP-NEWTON SOB A HIPÓTESE DE SUB-REGULARIDADE MÉTRICA DO TIPO HÖLDER**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua aprovação no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 31 de Maio de 2019.

JOSÉ ALBERTO RAMOS FLOR
Presidente da Banca Examinadora (UFPR)

PRISCILA SAVULSKI FERREIRA
Avaliador Externo (UTFPR/PR)

SANDRA AUGUSTA SANTOS
Avaliador Externo (UNICAMP)

ELIZABETH WEGNER KARAS
Avaliador Interno (UFPR)

Agradecimentos

A Deus, pela vida e suas oportunidades.

Ao Rick, por todo o apoio.

Ao professor Alberto, pela disponibilidade e paciência durante a realização deste trabalho.

À minha família e a todos que, de alguma forma, contribuíram para este propósito.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

RESUMO

O método LP-Newton é um método do tipo Newton que permite determinar um zero de um sistema de equações não lineares e resolve, a cada iteração, um problema de programação linear. Uma das características deste método é o fato de que ele apresenta mesma ordem de convergência local que o método de Newton, mesmo sob hipóteses mais fracas. Pouco depois de este método local haver sido proposto, uma versão globalizada deste método foi apresentada, garantindo o critério de Bouligand-estacionaridade para determinada classe de funções. Tendo em vista estes fatos e o objetivo de lidar com hipóteses diferentes, neste trabalho buscamos modificar a versão local do método LP-Newton de modo que, a cada iteração, o novo método ainda consista em resolver um problema de programação linear. Apresentamos um estudo da teoria de convergência local deste método sob hipóteses diferentes das consideradas pelo método LP-Newton, tais como a hipótese de subregularidade métrica do tipo Hölder, garantindo convergência local superlinear. De modo análogo à globalização do método LP-Newton, apresentamos também uma versão globalizada deste novo método utilizando busca linear. Para o algoritmo global, prova-se que o critério de Clarke-estacionaridade é satisfeito, ao considerarmos o problema de minimizar a norma infinito de uma função continuamente diferenciável $F : \mathbb{R}^n \to \mathbb{R}^m$.

Palavras-chave: Método LP-Newton. Subregularidade métrica do tipo Hölder. Sistema de equações não lineares. Otimização restrita. Conjunto poliédrico.

ABSTRACT

The LP-Newton method is a Newton-type method which deals with the problem of find a zero of a system of nonlinear equations and solves, at each iteration, a linear programming problem. One of the features of this method is the fact that it achieves the same local convergence order of Newton's method, even under weaker assumptions, [8]. A global version of this method was presented in [7], which assured the Bouligand-stationarity condition for certain classes of functions. In this work we present both local and global modifications to the LP-Newton method to deal with different assumptions than those considered in the LP-Newton method, as the Hölder metric subregularity hypothesis. The local algorithm of this new method still consists in solving a linear programming problem. A superlinear local convergence order is achieved by the LP-(η, θ) method, for certain class of functions. Likewise the LP-Newton globalization, we also present a globalized version of the LP-(η, θ) method, using line search. For this global algorithm, we prove that the Clarke-stationarity condition is achieved, when we consider the problem of minimizing the infinity norm of a continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^m$.

Keywords: LP-Newton method. Hölder metric subregularity. System of nonlinear equations. Constrained optimization. Polyhedral set.

Lista de Figuras

2.1 Região delimitada pelas condições de Lipschitz continuidade local e *error bound* local de uma função $F : \mathbb{R} \to \mathbb{R}^m$ que apresente um único zero na região $[-1,\ 1]$, de modo que $Z = \{0\}$. A região delimitada por estas duas condições é representada pelas duas desigualdades: $0.5\,|x| \leq \|F(x)\| \leq |x|$, respectivamente. 12

2.2 Região delimitada pelas condições de Hölder continuidade local e subregularidade métrica do tipo Hölder de uma função $F : \mathbb{R} \to \mathbb{R}^m$ que apresente um único zero na região $[-1,\ 1]$, de modo que $Z = \{0\}$. A região delimitada por estas duas condições é representada pelas duas desigualdades: $x^2 \leq \|F(x)\| \leq |x|^{0.5}$, respectivamente. 12

3.1 Passo de Newton para determinar um zero de uma função $H : \mathbb{R} \to \mathbb{R}$. 19

4.1 Gráfico da função $F(x) \coloneqq |x|^{1/2}$, no intervalo $[-1,\ 1]$. A condição de *error bound* local é satisfeita neste intervalo com $\ell = 1$, porém a Hipótese 4.1 não é satisfeita. 40

4.2 Gráfico da função $F(x) \coloneqq |x|^2$, no intervalo $[-1,\ 1]$. A função $F(\cdot)$ é Lipschitz contínua neste intervalo, portanto a Hipótese 4.1 é satisfeita. Entretanto, neste intervalo $F(\cdot)$ não satisfaz a condição de *error bound* local. 41

5.1 Exemplo 5.1. Consideramos a função $F(x) \coloneqq |x|^2$ e as desigualdades: $|x|^4 \leq |x|^2 \leq |x|$ no intervalo $\Omega \coloneqq [0,\ 1]$. Estas desigualdades podem ser interpretadas como: $\text{dist}(x, Z)^4 \leq F(x) \leq \text{dist}(x, Z)$ e representam, respectivamente, a condição de subregularidade métrica do tipo Hölder com constantes $c_1 = 1$ e $q_1 = 1/4$, e a Hipótese 4.1 do método LP-Newton, com constante $L = 1$. 45

5.2 Região viável de q_1 e q_2 para o Teorema 5.5. Consideramos a desigualdade $\dfrac{1}{1+q_2} < q_1$. 57

5.3 Região viável de η para o Teorema 5.5. Consideramos as desigualdades: $\dfrac{1}{q_1} - 1 < \eta \leq q_1\ q_2$. Para cada $q_2 \in (0, 1]$ fixado, a segunda desigualdade determina a região viável para os pares $(q_1,\ \eta)$ pela área sombreada abaixo de cada reta $\eta = q_2\ q_1$. 57

5.4 Região viável de q_1 e q_2 para o Corolário 5.6. Consideramos a desigualdade $\frac{1}{1+q_2} < {q_1}^2$. 58
5.5 Região viável de η para o Corolário 5.6. Consideramos as desigualdades: $\frac{1}{{q_1}^2} - 1 < \eta \leq q_1\, q_2$. Para cada $q_2 \in (0,1]$ fixado, a segunda desigualdade determina a região viável para os pares $(q_1,\ \eta)$ pela área sombreada abaixo de cada reta $\eta = q_2\, q_1$. 58
5.6 Região viável de θ para o Teorema 5.13 e o Lema 5.14. Consideramos as desigualdades: $\frac{1}{1+q_2} < \theta \leq q_1$. Para cada $q_2 \in (0,1]$ fixado, a região viável para os pares $(q_1,\ \theta)$ é a área sombreada acima de cada reta $\theta = \frac{1}{1+q_2}$. 74
5.7 Região viável de η para o Teorema 5.13 e o Lema 5.14. Consideramos as desigualdades: $1 - \frac{1}{1+q_2} < \eta \leq q_1\, q_2$, as quais são mais restritivas do que $1 < \eta + \theta \leq q_1\,(1+q_2)$ com $\frac{1}{1+q_2} < \theta \leq q_1$. Neste gráfico são ilustradas três escolhas para $q_2 \in (0,1]$: $q_2 = 1$ (azul), $q_2 = 0.75$ (rosa) e $q_2 = 0.5$ (verde). A região viável para os pares $(q_2,\ \eta)$ é o triângulo sombreado com as respectivas cores relacionadas a cada escolha de q_2. 75

Sumário

1 Introdução **1**

1.1 Motivação 1

1.2 Resultados 2

1.3 Estrutura da Dissertação 3

2 Notações, definições e teoremas básicos **4**

2.1 Notações 4

2.2 Definições e Teoremas 5

3 Preliminares **18**

3.1 O método de Newton 19

3.2 O método de Gauss-Newton 21

3.3 O método de Levenberg-Marquardt 23

4 O método LP-Newton **31**

4.1 As hipóteses de convergência local 32

4.1.1 Condições suficientes para as hipóteses de convergência . . . 36

4.2 Algoritmo global 41

5 Um novo método: LP-(η, θ) **45**

5.1 Algoritmo e teoria de convergência local 46

5.2 Algoritmo e teoria de convergência global 59

6 Conclusão **76**

Referências **78**

1 Introdução

1.1 Motivação

Seja $F : \mathbb{R}^n \to \mathbb{R}^m$ uma função continuamente diferenciável, consideremos o problema de se determinar um ponto $\bar{x} \in \mathbb{R}^n$ que seja solução do sistema de equações não lineares:

$$F(x) = 0. \tag{1.1}$$

Neste caso, dizemos que $\bar{x}$ é um zero de $F(\cdot)$. Este problema surge, por exemplo, quando precisamos determinar raízes de polinômios e em problemas de minimização irrestrita de funções continuamente diferenciáveis, mas também ocorre em situações menos triviais, a partir de reformulações de problemas de complementaridade, equilíbrios de Nash ou equações KKT [8].

Caso a matriz jacobiana desta função, denotada por $F'(\cdot)$, apresente certas propriedades, o método de Newton pode ser aplicado para resolver o Problema (1.1). O método de Newton é conhecido por garantir ordem de convergência quadrática local (ver [20, Theorem 3.2.2]). Entretanto, uma destas propriedades, a hipótese de não singularidade de $F'(\cdot)$ em uma vizinhança de um zero de $F(\cdot)$, pode restringir a classe de funções contempladas por este método. Algumas implicações decorrentes desta hipótese são o fato de que m deva ser, necessariamente, igual a n e que $\bar{x}$ seja solução isolada do Problema (1.1), ou seja, que em uma vizinhança de $\bar{x}$, ele próprio seja o único zero de $F(\cdot)$. Mais ainda, se quiséssemos resolver o Problema (1.1) restrito a um conjunto $\Omega \subset \mathbb{R}^n$ arbitrário, o método de Newton não estaria bem definido pois, a cada iteração, não haveria garantias de que a iteração seguinte estaria em Ω.

Uma hipótese local aternativa, considerada em certos aspectos mais fraca que a condição de não singularidade da matriz jacobiana de $F(\cdot)$, é a condição de *error bound*. Esta condição não traz implicações diretas sobre a matriz jacobiana de $F(\cdot)$ e nem sequer diz respeito a ela estar ou não definida em $\bar{x}$. A condição de *error bound* também não traz implicações diretas sobre a forma do conjunto Ω, nem exige que uma solução de (1.1) seja isolada. Mesmo assim, deixando de lado a hipótese de não singularidade da matriz jacobiana de $F(\bar{x})$ e utilizando a condição de *error bound*,

em [8] foi garantida a mesma ordem de convergência local do método de Newton a um novo método: o *Linear Programming-Newton*, ou simplesmente LP-Newton.

Tendo em vista obter modificações do método de Newton que garantam, ao menos, convergência local superlinear, muitos outros métodos foram desenvolvidos, inclusive sob hipóteses mais fracas (ver [20]). Um destes métodos é o de Levenberg-Marquardt. Enquanto que o método de Newton inspira-se na aproximação de Taylor de primeira ordem e procura zeros exatos de funções, localmente o método de Levenberg-Marquardt vem da ideia de se determinar um ponto que minimize o resíduo $\frac{1}{2}\|F(\cdot)\|^2$ por meio de uma solução de mínimos quadrados e, portanto, não precisa que a matriz jacobiana de $F(\bar{x})$ seja não singular, nem mesmo quadrada. Ainda assim, ambos os métodos necessitam de que $F(\cdot)$ seja uma função continuamente diferenciável. Considerando a situação em que a matriz jacobiana $F'(x_k)$ possa não estar definida para algum iterando x_k, faz-se interessante considerar uma matriz $G(x_k)$ como uma substituição adequada e é possível desenvolver uma versão do método de Levenberg-Marquardt para problemas com restrições, como foi feito em [12].

Assim como os métodos citados anteriormente, a teoria de convergência local do método LP-Newton está fundamentada sobre determinadas hipóteses, entre elas a condição de *error bound*, satisfeita pela função $F(\cdot)$ (ver [8]). Neste trabalho, é apresentada uma modificação do método LP-Newton e o estudo da convergência local deste método sob uma suposição mais fraca do que a condição de *error bound*: a hipótese de subregularidade métrica do tipo Hölder. Em [10] é apresentada uma estratégia de globalização do método LP-Newton. Tendo em vista esta estratégia, apresentamos também uma modificação deste novo método, bem como um critério de estacionaridade e condições sob as quais este critério pode ser garantido.

1.2 Resultados

Apresentamos neste trabalho uma modificação local do método LP-Newton, à qual chamaremos LP-(η, θ). Mostramos que o algoritmo local relacionado a esta modificação está bem definido e que, sob hipóteses adequadas, as iterações $\{x_k\}$ geradas por este algoritmo local convergem com ordem superlinear para um zero de uma função continuamente diferenciável, $F : \mathbb{R}^n \to \mathbb{R}^m$. Propomos também uma estratégia de globalização deste método, para a qual provamos que, sob determinadas condições, qualquer ponto de acumulação de uma sequência gerada por este algoritmo global satisfaz o critério de Clarke-estacionaridade (ver Definição 2.50) para o problema:

$$\underset{x\in\Omega}{\text{minimizar}} \quad f(x),$$

sendo $f(x) \coloneqq \|F(x)\|_\infty$ e $\Omega \subseteq \mathbb{R}^n$ um conjunto poliédrico (ver Definição 2.24).

1.3 Estrutura da Dissertação

Capítulo 2

Este capítulo foi escrito na intenção de revisar alguns conceitos, definições e teoremas necessários ao entendimento do método apresentado neste trabalho, bem como dos trabalhos anteriores aqui mencionados.

Capítulo 3

Neste capítulo, são apresentados conceitos preliminares relacionados ao método LP-Newton. São discutidos, brevemente, métodos anteriores ao LP-Newton, destacando-se o enfraquecimento de hipóteses e formas de convergência local. Finalmente, é apresentada uma comparação entre o LP-Newton e um algoritmo relacionado ao método de Levenberg-Marquardt, o que leva à motivação de utilizar a mesma estratégia de convergência local de um método de Levenberg-Marquardt para o algoritmo do método LP-(η, θ).

Capítulo 4

Aqui realizamos uma breve revisão do método LP-Newton. As hipóteses que garantem a ordem de convergência quadrática local, bem como situações em que são satisfeitas, o algoritmo local e os lemas nos quais se fundamenta a teoria de convergência local deste método. O algoritmo global do método LP-Newton para uma classe particular de funções e sua estratégia de convergência também são revisados neste capítulo.

Capítulo 5

Apresentamos neste capítulo as versões local e global do novo método, LP-(η, θ). Delimitamos, inicialmente, as hipóteses de convergência local deste método. Para sua versão local, apresentamos um algoritmo, o qual provamos estar bem definido e convergir com ordem superlinear para um zero de uma função F. Apresentamos também uma versão globalizada deste algoritmo para determinada classe de funções, a qual utiliza busca linear. Considerando uma reformulação do problema anterior como um problema de minimização, provamos, sob determinadas condições, que qualquer ponto de acumulação de uma sequência gerada por este algoritmo global satisfaz o critério de Clarke-estacionaridade.

Capítulo 6

Apresentamos neste capítulo um apanhado geral dos resultados deste trabalho, contribuições e possibilidades de investigações futuras.

2 Notações, definições e teoremas básicos

Neste capítulo, revisaremos conceitos, definições e teoremas necessários para a compreensão dos capítulos seguintes. Alguns destes conceitos e teoremas são bem conhecidos na literatura, pois são clássicos em Álgebra e Análise, enquanto outros, relacionados a Análise Variacional, podem ser menos conhecidos.

As principais notações utilizadas ao longo destas definições e teoremas, bem como ao longo dos capítulos seguintes, são apresentadas na seção a seguir. Exceto quando mencionado o contrário, estas notações serão utilizadas durante todo este trabalho conforme definidas aqui.

Durante todo este capítulo, sempre que mencionarmos uma função F estaremos nos referindo a uma função contínua $F : \mathbb{R}^n \to \mathbb{R}^m$. Notações, como a do conjunto Z, utilizada na Definição 3.18, podem ser encontradas na seção correspondente, a seguir.

2.1 Notações

$\|\cdot\|_2$	:	norma euclidiana em $\mathbb{R}^n$ ou $\mathbb{R}^m$
$\|\cdot\|_\infty$	:	norma infinito em $\mathbb{R}^n$ ou $\mathbb{R}^m$
$\|\cdot\|$	:	uma norma qualquer em $\mathbb{R}^n$ ou $\mathbb{R}^m$ (exceto quando for mencionado que esta notação será utilizada para denotar a norma infinito, como no Capítulo 5)
$\mathcal{B}(x,\ r)$	$:=$	$\{x' \in \mathbb{R}^n \mid \|x' - x\|_\infty \leq r\}$, em que $r > 0$
U	:	um subconjunto de $\mathbb{R}^n$
$\text{dist}(x, U)$	$:=$	$\inf_{u \in U} \|u - x\|_\infty$
Ω	$:=$	$\{x \in \mathbb{R}^n \mid A\,x \leq b\}$, em que $A \in \mathbb{R}^{\tilde{n} \times n},\ b \in \mathbb{R}^{\tilde{n}}$
A_i	:	i-ésima linha de uma matriz real A
$T_\Omega(x)$	:	cone tangente ao conjunto Ω em um ponto $x \in \Omega$
$N_\Omega(x)$	:	cone normal ao conjunto Ω em um ponto $x \in \Omega$

F	:	uma função contínua de $\mathbb{R}^n$ em $\mathbb{R}^m$
$f(x)$	$\coloneqq$	$\|F(x)\|$
Z	$\coloneqq$	$\{x \in \Omega \mid F(x) = 0\}$
$\bar{x}$	:	um ponto em Z tal que $\|x - \bar{x}\| = \text{dist}(x, Z)$
$F'(x)$	:	matriz jacobiana de $F(\cdot)$ em x, quando existir
$\mathcal{D}_F$	$\coloneqq$	$\{x \in \mathbb{R}^n \mid F'(x)$ está definida$\}$
$G(x)$	:	substituição para $F'(x)$
$h(x)$	:	uma função contínua de $\mathbb{R}^n$ em $\mathbb{R}$
$\partial_C h(x)$	:	subdiferencial de Clarke de $h(x)$
$\partial_B h(x)$	:	B-subdiferencial de $h(x)$
$O(h(x))$	:	O-grande de $h(x)$
$o(h(x))$	:	o-pequeno de $h(x)$
X^*	$\coloneqq$	$\{x \in \mathbb{R}^n \mid F(x) = 0\}$
$F^i(x)$	:	i-ésima função seleção de $F(x)$
conv U	:	envelope convexo de U
γ_k	:	valor ótimo do Problema (4.3)
$\gamma(x)$	:	valor ótimo do Problema (4.4)
(p_k, γ_k)	:	uma solução ótima do Problema (4.3)
$(p(x), \gamma(x))$	:	uma solução ótima do Problema (4.4)
δ_1 e β_1	:	constantes relacionadas ao Lema 5.3
δ_2 e β_2	:	constantes relacionadas ao Lema 5.4
x^*	:	ponto relacionado às hipóteses de convergência local de um método
1_n	:	matriz $n \times 1$ cujas entradas são todas iguais a 1
I_n	:	Matriz identidade de ordem n

2.2 Definições e Teoremas

Conceitos

Definição 2.1 (Zero)**.** *Dizemos que $x \in \Omega$ é um zero de $F(\cdot)$ quando $x \in Z$, ou seja, quando $F(x) = 0$.*

Definição 2.2 (Solução isolada). *Seja $x^* \in Z$, dizemos que x^* é solução isolada do problema de determinar $x \in Z$ quando existir $r > 0$ tal que:*

$$Z \cap \mathcal{B}(x^*, r) = \{x^*\} \ .$$

Definição 2.3 (Problema de complementaridade). *Um problema de complementaridade consiste em determinar $x \in \mathbb{R}^n$ tal que:*

$$H_1(x) \geq 0 \ , \quad H_2(x) \geq 0 \ \ e \ \ H_1(x)^T \, H_2(x) = 0 \ ,$$

sendo $H_1, H_2 : \mathbb{R}^n \to \mathbb{R}^m$ funções contínuas.

Álgebra, Análise e Matrizes

Definição 2.4 (Norma euclidiana). *Definimos a norma euclidiana de um vetor $x = [x_1 \ \cdots \ x_n]^T$ em $\mathbb{R}^n$ como*

$$\|x\|_2 \coloneqq \sqrt{{x_1}^2 + \cdots {x_n}^2} \ .$$

Definição 2.5 (Norma infinito). *Definimos a norma infinito de um vetor $x = [x_1 \ \cdots \ x_n]^T$ em $\mathbb{R}^n$ como*

$$\|x\| \coloneqq \max \{|x_1|, \ \cdots \ , \ |x_n|\} \ .$$

Proposição 2.6. *Sejam $a, b \in \mathbb{R}$ dois números não negativos. Então, $(a+b)^2 \geq a^2 + b^2$.* □

Corolário 2.7. *Sejam $a, b \in \mathbb{R}$ dois números não negativos. Então, $a + b \geq \sqrt{a^2 + b^2}$.* □

Corolário 2.8. *Sejam $a, b \in \mathbb{R}$ dois números não negativos. Então, $\sqrt{a^2 + b^2} \geq a$ e $\sqrt{a^2 + b^2} \geq b$.* □

Definição 2.9 (Posto de uma matriz). *Denotamos por $posto(A)$ o número de colunas (ou linhas) linearmente independentes de uma matriz $A \in \mathbb{R}^{m \times n}$.*

Definição 2.10 (Matriz não singular). *Dizemos que uma matriz $A \in \mathbb{R}^{m \times n}$ é não singular quando $posto(A) = m = n$.*

Definição 2.11 (Matriz positiva definida). *Dizemos que uma matriz $A \in \mathbb{R}^{n \times n}$ é positiva definida quando*

$$x^T \, A \, x > 0$$

para todo $x \in \mathbb{R}^n \setminus \{0\}$.

Proposição 2.12. *Seja $A \in \mathbb{R}^{n\times n}$ uma matriz positiva definida. Então, A é uma matriz não singular.* □

Sequências

Definição 2.13 (Cota superior ou inferior)**.** *Considere uma sequência $\{\alpha_k\}$ em $\mathbb{R}$. Dizemos que $M_1 \in \mathbb{R}$ é uma cota superior para a sequência $\{\alpha_k\}$ se, para todo k, a seguinte desigualdade for verdadeira:*

$$\alpha_k \leq M_1 \ .$$

Analogamente, dizemos que $M_2 \in \mathbb{R}$ é uma cota inferior para a sequência $\{\alpha_k\}$ se, para todo k, a seguinte desigualdade for verdadeira:

$$M_2 \leq \alpha_k \ .$$

Definição 2.14 (Sequência convergente)**.** *Considere uma sequência $\{x_k\}$ em $\mathbb{R}^n$. Dizemos que esta sequência é convergente, ou que converge para um ponto $\bar{x} \in \mathbb{R}^n$, se para todo $r > 0$ existir $\bar{k} \in \mathbb{N}$ tal que, para todo $k \geq \bar{k}$ a seguinte relação seja satisfeita:*

$$\|x_k - \bar{x}\| \leq r \ .$$

Definição 2.15 (Sequência de Cauchy)**.** *Seja $\{x_k\}$ uma sequência em $\mathbb{R}^n$. Dizemos que esta sequência é de Cauchy se, para todo $r > 0$, existir $\bar{k} \in \mathbb{N}$ tal que, se $k, k' \geq \bar{k}$, então*

$$\|x_k - x_{k'}\| < r \ .$$

Teorema 2.16 (Critério de Convergência de Cauchy)**.** *Uma sequência $\{x_k\}$ em $\mathbb{R}^n$ é convergente se, e somente se, for uma sequência de Cauchy.* □

Teorema 2.17 (Teorema do Confronto)**.** *Sejam $\{y'_k\}$ e $\{y''_k\}$ duas sequências convergentes em $\mathbb{R}$ tais que*

$$\lim_{k\to\infty} y''_k = \lim_{k\to\infty} y'_k = \bar{y} \ .$$

Se uma sequência $\{y_k\}$ em $\mathbb{R}$ for tal que, para k suficientemente grande, tenhamos

$$y''_k \leq y_k \leq y'_k$$

para todo y_k, então

$$\lim_{k\to\infty} y_k = \bar{y} \ .$$ □

Definição 2.18 (Ordem de convergência)**.** *Considere uma sequência $\{x_k\}$ em $\mathbb{R}^n$, convergente para algum ponto $\bar{x} \in \mathbb{R}^n$. Dizemos que esta sequência converge superline-*

armente e tem ordem de convergência $\xi > 1$ se existirem $M \geq 0$, $\bar{k} \in \mathbb{N}$ tais que, para todo $k \geq \bar{k}$, a seguinte desigualdade for verdadeira:

$$\|x_{k+1} - \bar{x}\| \leq M \, \|x_k - \bar{x}\|^{\xi} \, .$$

(Ver [16].)

Definição 2.19 (Convergência quadrática)**.** *Seja $\{x_k\}$ uma sequência convergente em $\mathbb{R}^n$. Dizemos que esta sequência converge quadraticamente ou que apresenta ordem de convergência quadrática quando a Definição 2.18 for satisfeita com $\xi = 2$.*

Definição 2.20 (O-grande e o-pequeno)**.** *Sejam $h, g : \mathbb{R}^n \to \mathbb{R}$ duas funções contínuas e não negativas e $\{x_k\}$ uma sequência convergente em $\mathbb{R}^n$, sendo $x_k \to x$. Então,*

(i) $h(x) = O(g(x))$, se existirem $\bar{k} \in \mathbb{N}$ e $M > 0$ tais que, para todo $k \geq \bar{k}$, tivermos:

$$h(x_k) \leq M \cdot g(x_k).$$

(ii) $h(x) = o(g(x))$, se $g(x_k) > 0$ para todo x_k e

$$\frac{h(x_k)}{g(x_k)} \to 0 \, .$$

Conjuntos

Definição 2.21 (Conjunto convexo)**.** *Dizemos que um conjunto $U \subseteq \mathbb{R}^n$ é convexo quando, para cada par $x, \; x' \in U$ e para todo $t \in [0, 1]$, tivermos*

$$(1 - t) \, x + t \, x' \in U \, .$$

Teorema 2.22 (Distância de um ponto a um conjunto convexo)**.** *Sendo $U \subseteq \mathbb{R}^n$ um conjunto convexo, existirá um único ponto $\bar{x} \in \bar{U}$, o fecho de U, tal que dist$(x, U) = \|x - \bar{x}\|$.* □

Definição 2.23 (Envoltória convexa)**.** *Considere um conjunto $U \subseteq \mathbb{R}^n$. A envoltória convexa de U é definida como a interseção de todos os conjuntos convexos $W \in \mathbb{R}^n$ tais que $U \subseteq W$, ou equivalentemente como o menor conjunto convexo $W \in \mathbb{R}^n$ tal que $U \subseteq W$. A envoltória convexa de U é denotada por conv U .*
(Ver [18, p. 53].)

Definição 2.24 (Conjunto poliédrico)**.** *Dizemos que $U \subseteq \mathbb{R}^n$ é um conjunto poliédrico se existirem $\tilde{n} \in \mathbb{N}$, uma matriz $A \in \mathbb{R}^{\tilde{n} \times n}$ e um vetor $b \in \mathbb{R}^{\tilde{n}}$ tais que o conjunto U possa ser descrito da seguinte forma:*

$$U \coloneqq \{x \in \mathbb{R}^n \mid Ax \leq b\} \, .$$

Definição 2.25 (Cone Tangente). *Seja $U \subseteq \mathbb{R}^n$, $x \in U$. Definimos o cone tangente a U em x como:*

$$T_U(x) \coloneqq \Big\{ d \in \mathbb{R}^n \mid \textit{Existem sequências } \{x_k\} \textit{ em } U, \ \{\eta_k\} \subset \mathbb{R} \textit{ tais que } x_k \to x, \ \eta_k \to 0 \textit{ e } \frac{x_k - x}{\eta_k} \to d \Big\}.$$

(Ver [18, Definição 6.1].)

Definição 2.26 (Cone normal). *Seja $U \subseteq \mathbb{R}^n$, $x \in U$. Definimos o cone normal a U em x como:*

$$N_U(x) \coloneqq \Big\{ d \in \mathbb{R}^n \mid \textit{existem sequências } \{x_k\} \textit{ em } U, \ \{d_k\} \textit{ em } \mathbb{R}^n \textit{ tais que } x_k \to x, \ d_k \to d \textit{ e, para todo } k, \textit{ temos } {d_k}^T (x_k - x) \leq o(\|x_k - x\|) \Big\}.$$

(Ver [18, Definição 6.3].)

Proposição 2.27. *Seja $x \in U$. Temos que $0 \in N_U(x)$.* □

Continuidade de funções

Teorema 2.28 (Weierstrass). *Seja $U \subset \mathbb{R}^n$ um conjunto compacto, $h : \mathbb{R}^n \to \mathbb{R}$ uma função contínua. Então, existem $x', x'' \in U$ tais que*

$$h(x') = \min_{x \in U} h(x) \quad \textit{e} \quad h(x'') = \max_{x \in U} h(x) \, .$$ □

Corolário 2.29. *Seja $U \subseteq \mathbb{R}^n$ compacto, $h_1, h_2 : \mathbb{R}^n \to \mathbb{R}$ duas funções contínuas e não negativas. Então,*

$$\min_{p \, \in \, U} \{h_1(p) + h_2(p)\} \geq \min_{p \, \in \, U} \left\{ \sqrt{h_1(p)^2 + h_2(p)^2} \right\} .$$

Demonstração: Sendo U compacto, $h_1(\cdot)$, $h_2(\cdot)$ funções contínuas, pelo Teorema 2.28 existe $\bar{p} \in U$ tal que:

$$\min_{p \, \in \, U} \{h_1(p) + h_2(p)\} = h_1(\bar{p}) + h_2(\bar{p}) \, .$$

Sendo $h_1(\cdot)$, $h_2(\cdot)$ funções não negativas, pelo Corolário 2.7 temos que:

$$\begin{aligned} \min_{p \, \in \, U} \{h_1(p) + h_2(p)\} &= h_1(\bar{p}) + h_2(\bar{p}) \\ &\geq \sqrt{h_1(\bar{p})^2 + h_2(\bar{p})^2} \\ &\geq \min_{p \, \in \, U} \left\{ \sqrt{h_1(p)^2 + h_2(p)^2} \right\} . \end{aligned}$$

□

Corolário 2.30. *Seja $U \subseteq \mathbb{R}^n$ compacto, $h_1, h_2 : \mathbb{R}^n \to \mathbb{R}$ duas funções contínuas e não negativas. Então,*

$$\min_{p \in U} \left\{ \sqrt{h_1(p)^2 + h_2(p)^2} \right\} \geq \min_{p \in U} \{\max\{h_1(p),\ h_2(p)\}\} \ .$$

Demonstração: Tal como no corolário anterior, sabemos pelo Teorema 2.28 que existe $\bar{p} \in U$ tal que:

$$\min_{p \in U} \left\{ \sqrt{h_1(p)^2 + h_2(p)^2} \right\} = \sqrt{h_1(\bar{p})^2 + h_2(\bar{p})^2} \ .$$

Sendo $h_1(\cdot)$, $h_2(\cdot)$ funções não negativas, pelo Corolário 2.8 temos que:

$$\begin{aligned} \min_{p \in U} \left\{ \sqrt{h_1(p)^2 + h_2(p)^2} \right\} &= \sqrt{h_1(\bar{p})^2 + h_2(\bar{p})^2} \\ &\geq \max\{h_1(\bar{p}),\ h_2(\bar{p})\} \\ &\geq \min_{p \in U} \{\max\{h_1(p),\ h_2(p)\}\} \ . \end{aligned}$$

□

Definição 2.31 (Condição de *error bound*)**.** *Dizemos que a função F satisfaz a condição de error bound local, para o problema de determinar um zero de F em uma vizinhança de um ponto $x^* \in Z$, se existirem $c_1,\ r > 0$ tais que:*

$$c_1\ dist(x, Z) \leq \|F(x)\|_\infty$$

para todo $x \in \mathcal{B}(x^*, r) \cap \Omega$.

A propriedade apresentada a seguir pode ser vista como um enfraquecimento da condição de *error bound* local.

Definição 2.32 (Subregularidade métrica do tipo Hölder)**.** *Dizemos que a função F satisfaz a condição de subregularidade métrica do tipo Hölder, com ordem $q_1 \in (0,\ 1]$, para o problema de determinar um zero de F em uma vizinhança de um ponto $x^* \in Z$, se existirem $c_1,\ r > 0$ tais que:*

$$c_1\ dist(x, Z) \leq \|F(x)\|_\infty^{q_1}$$

para todo $x \in \mathcal{B}(x^*, r) \cap \Omega$.

Note que uma função que satisfaça a condição de *error bound* local também satisfaz a condição de subregularidade métrica do tipo Hölder para qualquer $q_1 \in (0,\ 1]$, quando $r > 0$ é suficientemente pequeno. Além disso, em alguns casos é conveniente escolher $r > 0$ de modo que $\|F(x)\|_\infty < 1$, para todo $x \in \mathcal{B}(x^*, r)$, sendo $x^* \in Z$.

Definição 2.33 (Lipschitz continuidade local)**.** *Dizemos que uma função F é Lipschitz contínua em uma vizinhança de um ponto $x^* \in \mathbb{R}^n$ se existirem $L,\ r > 0$ tais que:*

$$\|F(x') - F(x)\|_\infty \leq L \|x' - x\|_\infty$$

para todo $x,\ x' \in \mathcal{B}(x^*, r)$.

A definição a seguir pode ser vista como um enfraquecimento da condição de Lipschitz continuidade local. Toda função que seja localmente Lipschitz contínua também satisfaz a condição de Hölder continuidade local, para qualquer $q_2 \in (0,\ 1]$, quando $r > 0$ é suficientemente pequeno.

Definição 2.34 (Hölder continuidade)**.** *Dizemos que uma função contínua $H : \mathbb{R}^n \to \mathbb{R}^m$ é Hölder contínua em uma vizinhança de um ponto $x^* \in \mathbb{R}^n$ se existirem $\tilde{L} > 0$ e $q_2 \in (0,\ 1]$ tais que*

$$\|H(x') - H(x)\| \leq \tilde{L} \|x' - x\|^{q_2}$$

para todos $x, x' \in \mathcal{B}(x^*, r)$, para algum $r > 0$.

Observe que, ao considerarmos uma função continuamente diferenciável $F(\cdot)$ e $H(\cdot) = F'(\cdot)$, a condição:

$$\|F(x') - F(x) - H(x)\ (x' - x)\| \leq \tilde{L} \|x' - x\|^{1+q_2}$$

para todo $x, x' \in \mathcal{B}(x^*, r)$ e algum $r > 0$ suficientemente pequeno, relaciona-se à condição de Hölder continuidade local de H e pode ser vista como um enfraquecimento da condição de que a matriz jacobiana de $F(\cdot)$ seja localmente Lipschitz contínua em x^*.

Para demonstrar a ordem de convergência local da sequência $\{\|F(x_k)\|\}$ obtida pelo método LP-Newton, em [8] são utilizados os limitantes obtidos pelas hipóteses de Lipschitz continuidade local e condição de *error bound* local para limitar $\|F(x_k)\|$ superior e inferiormente, respectivamente. Isto é ilustrado pela Figura 2.1, a seguir.

Analogamente, demonstramos a ordem de convergência local de uma sequência $\{\|F(x_k)\|\}$ gerada método LP-(η, θ), apresentado neste trabalho, por meio dos limitantes obtidos das hipóteses de Hölder continuidade local e condição de subregularidade métrica do tipo Hölder. Por estas hipóteses, obtemos limitantes superior e inferior para $\|F(x_k)\|$, respectivamente. Esta diferença é ilustrada pela Figura 2.2, a seguir.

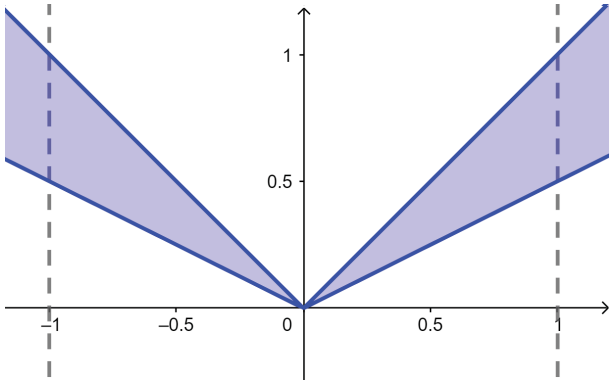


Figura 2.1: Região delimitada pelas condições de Lipschitz continuidade local e *error bound* local de uma função $F : \mathbb{R} \to \mathbb{R}^m$ que apresente um único zero na região $[-1,\ 1]$, de modo que $Z = \{0\}$. A região delimitada por estas duas condições é representada pelas duas desigualdades: $0.5\ |x| \leq \|F(x)\| \leq |x|$, respectivamente.

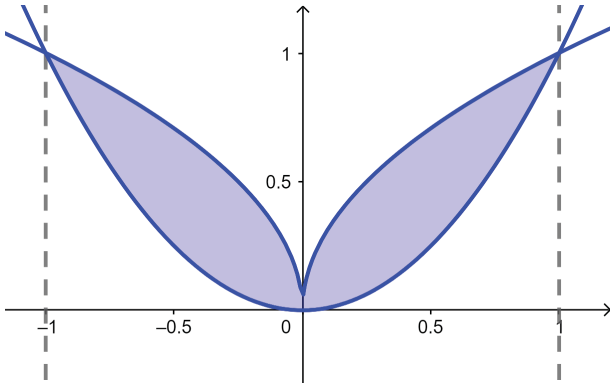


Figura 2.2: Região delimitada pelas condições de Hölder continuidade local e subregularidade métrica do tipo Hölder de uma função $F : \mathbb{R} \to \mathbb{R}^m$ que apresente um único zero na região $[-1,\ 1]$, de modo que $Z = \{0\}$. A região delimitada por estas duas condições é representada pelas duas desigualdades: $x^2 \leq \|F(x)\| \leq |x|^{0.5}$, respectivamente.

Proposição 2.35. *Sejam $h^1, \ldots, h^j : \mathbb{R}^n \to \mathbb{R}$ funções Lipschitz contínuas em uma região $U \subseteq \mathbb{R}^n$, $j \in \mathbb{N}$, e uma função $h : \mathbb{R}^n \to \mathbb{R}$, definida por*

$$h(x) \coloneqq \max\left\{|h^1(x)|,\ \ldots,\ |h^j(x)|\right\}\ .$$

Então, a função h também é Lipschitz contínua em U.

Demonstração: Provaremos o caso em que consideramos a norma infinito, denotada por $\|\cdot\|$. Suponha que para cada $i = 1, \ldots, j$ exista $L_i > 0$ tal que, para todos $x, x' \in U$, tenhamos:

$$|h^i(x') - h^i(x)| \leq L_i\ \|x' - x\|\ .$$

Então, para cada $x, x' \in U$, teremos:

$$|h(x') - h(x)| \;=\; |h^{k_1}(x') - h^{k_2}(x)| \,,$$

para algum par $k_1, k_2 \in \{1, \ldots, j\}$. Suponhamos, s.p.g., que $h^{k_1}(x') \;\geq\; h^{k_2}(x)$. Então,

$$\begin{aligned} |h^{k_1}(x') - h^{k_2}(x)| \;&=\; h^{k_1}(x') - h^{k_2}(x) \;=\; h^{k_1}(x') - \max_{i=1,\ldots,j} h^i(x) \\ &\leq\; h^{k_1}(x') - h^{k_1}(x) \;=\; |h^{k_1}(x') - h^{k_1}(x)| \\ &\leq\; L_{k_1} \,\|x' - x\| \;\leq\; \max_{i=1,\ldots,j} L_i \,\|x' - x\| \,. \end{aligned}$$

Assim, escolhendo $L = \max\limits_{i=1,\ldots,j} L_i$, teremos que

$$|h(x') - h(x)| \;\leq\; L \,\|x' - x\| \,,$$

para cada $x, x' \in U$, portanto a função h é Lipschitz contínua. □

Definição 2.36 (Função de classe C^1)**.** *Dizemos que F é uma função de classe C^1, ou que é continuamente diferenciável, se F é uma função contínua, sua matriz jacobiana $F'(x)$ está definida para todo $x \in \mathbb{R}^n$ e F' é uma função contínua.*

Definição 2.37 (Seleção contínua)**.** *Seja $j \in \mathbb{N}$ e sejam $F^i : \mathbb{R}^n \to \mathbb{R}^m$, $i = 1, \ldots, j$ funções contínuas. Dizemos que F é uma seleção contínua destas funções se F for contínua e, para cada $x \in \mathbb{R}^n$, tivermos $F(x) \in \{F^i(x) \mid i = 1, \ldots, j\}$.*

Definição 2.38 (Função de classe C^1 por partes)**.** *Seja F uma seleção contínua de funções $F^i : \mathbb{R}^n \to \mathbb{R}^m$, $i = 1, \ldots, j$ continuamente diferenciáveis, $j \in \mathbb{N}$. Dizemos, então, que F é uma função de classe C^1 por partes.*

Derivadas

Definição 2.39 (Derivada direcional)**.** *Definimos a derivada direcional da função h em um ponto $x \in \mathbb{R}^n$ na direção $d \in \mathbb{R}^n$, quando existir, por*

$$h'(x; d) \coloneqq \lim_{t \to 0^+} \frac{h(x + t\, d) - h(x)}{t}$$

em que $t \in \mathbb{R}$, $d \in \mathbb{R}^n$.

Definição 2.40 (Derivada de Clarke)**.** *Suponha que a função h seja Lipschitz contínua em uma vizinhança de um ponto $x \in \mathbb{R}^n$. Definimos a derivada de Clarke de $h(x)$ na direção $d \in \mathbb{R}^n$ por*

$$h^o(x;d) \coloneqq \limsup_{x' \to x \ , \ t \to 0^+} \frac{h(x' + t\ d) - h(x')}{t}$$

(Ver [5, p. 10].)

Definição 2.41 (Clarke regularidade)**.** *Suponha que a função h seja Lipschitz contínua em uma vizinhança de um ponto $x \in \mathbb{R}^n$. Dizemos que h é Clarke regular em x se:*

(i) Para todo $d \in \mathbb{R}^n$, a derivada direcional $h'(x;d)$ existe;

(ii) Para todo $d \in \mathbb{R}^n$, temos que $h'(x;d) = h^o(x;d)$.

(Ver [5, p. 39, Definição 2.3.4].)

Proposição 2.42. *Seja $J \coloneqq \{1, \ \ldots, \ j\}$ um conjunto finito de índices e, para cada $i \in J$, seja a função $h_i : \mathbb{R}^n \to \mathbb{R}$ contínua. Considerando a função h definida por $h(x) \coloneqq \max\{h_i(x) \mid i \in J\}$ e, fixado um ponto $\bar{x} \in \mathbb{R}^n$, defina o conjunto $I(\bar{x}) \coloneqq \{i \in J \mid h_i(\bar{x}) = h(\bar{x})\}$. Se, para cada $i \in J$, a função h_i for Lipschitz contínua em uma vizinhança de $\bar{x}$ e Clarke regular em $\bar{x}$, então, a função h é Clarke regular em $\bar{x}$.* □

(Ver [19, Proposição 7.4.7 (Regra do Máximo)].)

Definição 2.43 (Subdiferencial de Clarke)**.** *Considere um conjunto $U \subset \mathbb{R}^n$ e um ponto $x \in \mathbb{R}^n$ tal que a derivada generalizada de Clarke $h^o(x;d)$ esteja definida para todo $d \in T_U(x)$. Definimos o subdiferencial de h em x por:*

$$\partial_C h(x) \coloneqq \left\{ w \in \mathbb{R}^n \mid h^o(x;d) \geq w^T\ d\ , \ \ \forall\ d \in \mathbb{R}^n \right\}.$$

(Ver [5, p. 10].)

Definição 2.44 (B-subdiferencial)**.** *Quando $F : \mathbb{R}^n \to \mathbb{R}^n$ é localmente Lipschitz contínua e diferenciável em todo o $\mathbb{R}^n$, exceto em um número finito de pontos, definimos o B-subdiferencial ou jacobiana limitante de F em um ponto x por:*

$$\partial_B F(x) \coloneqq \left\{ \lim_{k \to \infty} F'(x_k) \mid \lim_{k \to \infty} x_k = x \ , \text{ em que } \{x_k\} \text{ é uma sequência em } \mathcal{D}_F \right\}.$$

(Ver [8].)

Definição 2.45 (Função semi-suave e fortemente semi-suave)**.** *Seja um ponto $x \in \mathbb{R}^n$, uma função contínua h Lipschitz contínua em uma vizinhança de x, tal que exista sua matriz jacobiana em x, $h'(x)$. Então, dizemos que*

(i) h é semi-suave em x se

$$V \cdot d - h'(x) \cdot d = o(\|d\|)$$

para toda sequência $\{d\} \subset \mathbb{R}^n$ tal que $d \to 0$ e todo $V \in conv\ \partial_B\ h(x + d)$.

(ii) h *é fortemente semi-suave em* x *se*

$$V \cdot d - h'(x) \cdot d = O(\|d\|)$$

para toda sequência $\{d\} \subset \mathbb{R}^n$ *tal que* $d \to 0$ *e todo* $V \in$ *conv* $\partial_B\, h(x+d)$.

(Ver [17, p. 2].) Quando F é semi-suave, ou fortemente semi-suave, para todo $x \in \mathbb{R}^n$ dizemos simplesmente que F é semi-suave ou fortemente semi-suave, respectivamente.

Critérios de Estacionaridade

Considerando o problema irrestrito de minimizar uma função continuamente diferenciável $f : \mathbb{R}^n \to \mathbb{R}$, é bem conhecido na literatura o critério de estacionaridade $\nabla f(x) = 0$. Entretanto, quando consideramos problemas de minimização com restrições da forma:

$$\underset{x \in U}{\text{minimizar}}\ h(x) \tag{2.1}$$

em que $U \subseteq \mathbb{R}^n$ e cuja função objetivo $h : \mathbb{R}^n \to \mathbb{R}$ possa apresentar apenas derivadas direcionais em algum ponto ou, ainda, apenas o subdiferencial de Clarke, este critério poderá não ser satisfeito dentro do conjunto U.

Para estes casos, existem outros critérios de estacionaridade, os quais serão aqui descritos. Para os resultados e definições apresentados nesta seção, a única hipótese considerada sobre a função h é a de que esta seja contínua.

Definição 2.46 (A-estacionaridade ou Abadie-estacionaridade)**.** *Seja um conjunto* $U \subseteq \mathbb{R}^n$*, dizemos que um ponto* $\bar{x} \in U$ *é A-estacionário para o Problema (2.1) se, para todo* $d \in T_U(\bar{x})$*, a derivada direcional* $h'(\bar{x}; d)$ *existir e tivermos:*

$$h'(\bar{x}; d) = 0\ .$$

Observe que em um problema irrestrito consideramos $U = \mathbb{R}^n$, logo $T_U(x) = \mathbb{R}^n$, para todo $x \in \mathbb{R}^n$. Neste caso, o critério de A-estacionaridade em um ponto $\bar{x} \in \mathbb{R}^n$ para este problema equivale a: $\nabla h(\bar{x}) = 0$.

Teorema 2.47 (Regra de Fermat)**.** *Seja* h *continuamente diferenciável. Então, toda solução global* $\bar{x}$ *de um problema de otimimização irrestrito :*

$$\min_{x \in \mathbb{R}^n}\ h(x)$$

satisfaz o critério de A-estacionaridade para este problema. □

Definição 2.48 (B-estacionaridade ou Bouligand-estacionaridade)**.** *Seja um conjunto $U \subseteq \mathbb{R}^n$, dizemos que um ponto $\bar{x} \in U$ é B-estacionário para o Problema (2.1) se, para todo $d \in T_U(\bar{x})$, a derivada direcional $h'(\bar{x}; d)$ existir e tivermos:*

$$h'(\bar{x}; d) \geq 0 \, .$$

(Ver [2, Definição 4 (i)].)

Proposição 2.49. *Se o critério de A-estacionaridade é satisfeito em um ponto $\bar{x} \in U$, para o Problema (2.1), então o critério de B-estacionaridade também é satisfeito em $\bar{x}$ para este mesmo problema.* □

Definição 2.50 (C-estacionaridade ou Clarke estacionaridade)**.** *Seja um conjunto $U \subseteq \mathbb{R}^n$, dizemos que um ponto $\bar{x} \in U$ é C-estacionário para o Problema (2.1) se, para todo $d \in T_U(\bar{x})$, tivermos*

$$0 \in \partial_C h(\bar{x}) + N_U(\bar{x}) \, .$$

(Ver [9, p. 634 (7.1.5)].)

Proposição 2.51. *Se o critério de B-estacionaridade é satisfeito em um ponto $\bar{x} \in U$, para o Problema (2.1), então o critério de C-estacionaridade também é satisfeito em $\bar{x}$ para este mesmo problema.* □

Assim, das duas proposições anteriores temos o corolário a seguir.

Corolário 2.52. *Se o critério de A-estacionaridade é satisfeito em um ponto $\bar{x} \in U$, para o Problema (2.1), então o critério de C-estacionaridade também é satisfeito em $\bar{x}$ para este mesmo problema.* □

Destes três resultados, vemos que o critério de A-estacionaridade é uma condição mais forte que o critério de B-estacionaridade o qual, por sua vez, é uma condição mais forte que o critério de C-estacionaridade. Veremos agora exemplos nos quais estes critérios podem não ser satisfeitos simultaneamente.

Exemplo 2.53. *Neste exemplo, mostraremos um caso em que é satisfeito o critério de C-estacionaridade, mas não o de B-estacionaridade. Considere o Problema (2.1), com $U := \{(x_1,\ x_2) \in \mathbb{R}^2 \mid x_1, x_2 \geq 0\}$, $\bar{x} = (\bar{x_1},\ \bar{x_2}) := (0,\ 0)$ e*

$$h(x') := \begin{cases} \dfrac{x_1\, x_2}{x_1^2 + x_2^2} \ , \ \textit{se } (x_1,\ x_2) \neq (0,\ 0) \\ 0 \ , \ \textit{se } (x_1,\ x_2) = (0,\ 0) \ . \end{cases}$$

Vemos que $\bar{x}$ é um ponto C-estacionário, pois o plano identicamente nulo passa por $\bar{x}$ e está abaixo do gráfico da função $h(\cdot)$. Entretanto, o critério de B-estacionaridade não é satisfeito, pois as derivadas direcionais de $h(\cdot)$ não estão definidas no ponto $\bar{x}$. □

Exemplo 2.54. *Neste exemplo, mostraremos um caso em que é satisfeito o critério de B-estacionaridade, mas não o de A-estacionaridade. Considere o problema irrestrito:*

$$\underset{x' \in \mathbb{R}}{\text{minimizar}} \;\; h(x') \;,$$

em que definimos $h(x') := |x'|$ *e o ponto* $x = 0$*. Este ponto é B-estacionário, pois:*

$$h'(x; d) = \begin{cases} d \;, \text{ se } d \geq 0 \\ -d \;, \text{ se } d < 0 \end{cases}$$

ou seja, $h'(x; d) = |d| \geq 0$*, para todo* $d \in \mathbb{R}$*. Entretanto, este ponto não é A-estacionário, pois* $h'(x; d) = |d| \neq 0$*, para todo* $d \neq 0$*.* □

Programação Linear

Definição 2.55 (Primal e dual)**.** *Sejam* $x,\ c \in \mathbb{R}^n$*,* $b \in \mathbb{R}^{\bar{n}}$*,* $A \in \mathbb{R}^{\bar{n} \times n}$*. Considere um problema de programação linear, ao qual denotaremos primal, tal que possa ser descrito da seguinte forma:*

$$\begin{aligned} \text{minimizar} \;\; & c^T x \\ \text{sujeito a:} \;\; & A\, x \leq b \end{aligned} \tag{2.2}$$

Sendo $y \in \mathbb{R}^{\bar{n}}$*, o dual deste problema é definido como:*

$$\begin{aligned} \text{maximizar} \;\; & b^T y \\ \text{sujeito a:} \;\; & A^T y = c \\ & y \geq 0 \;. \end{aligned} \tag{2.3}$$

Definição 2.56 (Condições KKT)**.** *Considere o problema de programação linear definido por (2.2). Dizemos que as condições KKT são satisfeitas para este problema se existirem vetores* $y \geq 0$ *e* $x \in \mathbb{R}^n$*, tais que* $y_i \, (A_i \, x - b_i) = 0$ *, para cada* $i = 1, \ldots, \bar{n}$ *, em que* x *satisfaz as restrições do Problema (2.2) e* y *satisfaz as restrições do Problema (2.3).*

Teorema 2.57 (KKT para programação linear)**.** *Um ponto* $x \in \mathbb{R}^n$ *é minimizador local para o de programação linear (2.2), satisfazendo suas restrições, se, e somente se,* x *é um ponto KKT para este problema.* □

3 Preliminares

Seja $F : \mathbb{R}^n \to \mathbb{R}^m$ uma função continuamente diferenciável, considere o problema de se determinar uma solução para o sistema de equações:

$$F(x) = 0 \, ,$$

cujo conjunto solução denotaremos por $X^* \coloneqq \{x \in \mathbb{R}^n \mid F(x) = 0\}$.

Neste capítulo, abordaremos alguns métodos anteriores ao LP-Newton, os quais podem ser utilizados para se determinar um ponto

$$\bar{x} \in X^* \, . \tag{3.1}$$

No decorrer deste capítulo utilizaremos as notações de normas apresentadas no Capítulo 2. Diferentemente dos capítulos seguintes, utilizaremos as notações:

- $\mathcal{B}_2(x, \, r) \coloneqq \{x' \in \mathbb{R}^n \mid \|x' - x\|_2 \leq r\}$;
- $\mathcal{B}_\infty(x, \, r) \coloneqq \{x' \in \mathbb{R}^n \mid \|x' - x\|_\infty \leq r\}$;

para denotarmos bolas fechadas em $\mathbb{R}^n$, em relação à norma euclidiana e à norma infinito, respectivamente.

Também faremos menção a critérios de estacionaridade. Considerando uma função $f : \mathbb{R}^n \to \mathbb{R}$, continuamente diferenciável, um ponto $x \in \mathbb{R}^n$ e uma direção $d \in \mathbb{R}^n$, é possível mostrar que a derivada direcional de $f(x)$ na direção d é:

$$f'(x; d) = \nabla f(x)^T \, d \, ,$$

sendo $\nabla f(x)$ o vetor gradiente de $f(x)$. Considerando o problema irrestrito de minimizar $f(x)$, vemos pela Definição 2.46, que um ponto $\bar{x}$ é A-estacionário quando $\nabla f(x) = 0$ e, também devido ao fato de este ser um problema irrestrito, pela Definição 2.48 vemos, neste caso, que a B-estacionaridade de um ponto $\bar{x}$ equivale à A-estacionaridade.

Até agora, mencionamos duas funções: $F(\cdot)$ e $f(\cdot)$, continuamente diferenciáveis. Entretanto, nas próximas seções consideraremos hipóteses adicionais a esta. Assim,

faremos uso de três funções: $H_1(\cdot)$, $H_2(\cdot)$ e $H_3(\cdot)$, com suas respectivas hipóteses, na intenção de manter a clareza quando nos referirmos a cada um dos métodos abordados neste capítulo.

3.1 O método de Newton

Seja $H_1 : \mathbb{R}^n \to \mathbb{R}^n$ uma função continuamente diferenciável, consideremos o problema de se determinar um ponto $\bar{x} \in \mathbb{R}^n$ tal que (3.1) seja satisfeita para $F(\cdot) := H_1(\cdot)$. O método de Newton é bastante conhecido para resolver este problema, devido ao fato de apresentar ordem de convergência quadrática local. Sob hipóteses adequadas, dado um iterando $x_k \in \mathbb{R}^n$, o método de Newton calcula o próximo iterando $x_{k+1} := x_k + p_k^N$ por meio do seguinte passo:

$$p_k^N := -H_1'(x_k)^{-1} H_1(x_k) . \tag{3.2}$$

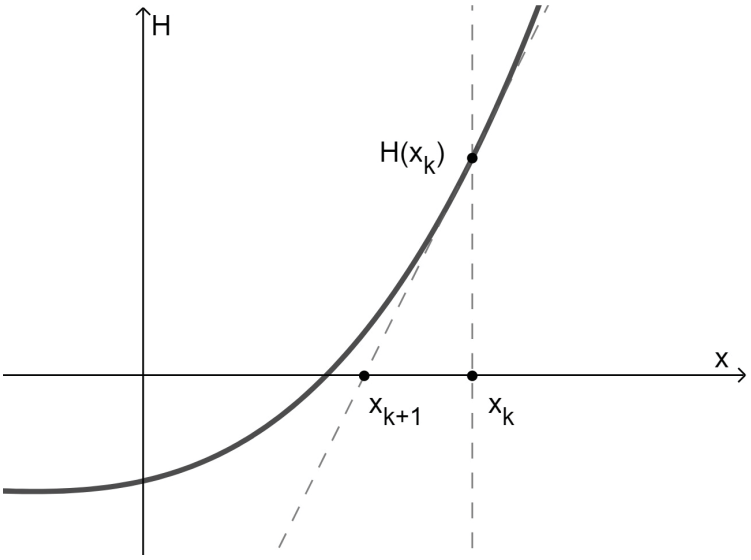


Figura 3.1: Passo de Newton para determinar um zero de uma função $H : \mathbb{R} \to \mathbb{R}$.

Conforme descrito em [20, Teorema 3.2.2], a convergência quadrática local do método de Newton para determinar um zero de uma função pode ser apresentada conforme o teorema enunciado abaixo.

Teorema 3.1. *Seja $H_1 : \mathbb{R}^n \to \mathbb{R}^n$ uma função continuamente diferencíavel, $x^* \in \mathbb{R}^n$ um zero de H_1. Suponhamos que a matriz jacobiana $H_1'(x)$ seja Lipschitz contínua e não singular em uma vizinhança de x^*. Então, considerando x_0 suficientemente próximo de x^*, teremos que:*

(i) para cada x_k, $k = 0, 1, \ldots$, o passo de Newton (3.2) está bem definido.

Além disso, considerando a sequência $\{x_k\}$ gerada pelo método de Newton com ponto inicial x_0, teremos que:

(ii) a sequência $\{x_k\}$ converge para x^*;

(iii) a sequência $\{x_k\}$ apresenta ordem de convergência quadrática;

(iv) a sequência $\{H_1(x_k)\}$ converge quadraticamente para zero. □

O método de Newton também pode ser utilizado para minimizar uma função irrestrita em $\mathbb{R}^n$. Suponha que $h_1 : \mathbb{R}^n \to \mathbb{R}$ seja uma função de classe C^2 e, por simplicidade de notação, denotaremos também por $x^* \in \mathbb{R}^n$ um minimizador de h_1. Sendo este um problema irrestrito, temos pelo critério de A-estacionaridade que $\nabla h_1(x^*) = 0$. Supondo que a matriz hessiana, $\nabla^2 h_1(\cdot)$, seja Lipschitz contínua e definida positiva em uma vizinhança de x^*, conhecido o ponto $x_k \in \mathbb{R}^n$ suficientemente próximo de x^*, o método de Newton para o problema de minimização irrestrita calcula o seguinte passo:

$$p_k^{N_{min}} := -h_1''(x_k)^{-1}\, h_1'(x_k) \tag{3.3}$$

e definimos o iterando seguinte por $x_{k+1} := x_k + p_k^{N_{min}}$. Conforme o Teorema 3.1, para x_0 suficientemente próximo de x^*, esta escolha do passo garante ordem de convergência quadrática da sequência $\{x_k\}$ para x^*. Além disso, devido à não singularidade de $h_1''(\cdot)$ em uma vizinhança de x^*, este ponto será também o único que satisfaz o critério de A-estacionaridade para h_1 nesta região.

Note que o passo de Newton para o problema de determinar um zero de $H_1(\cdot)$ está bem definido devido ao fato de a matriz jacobiana $H_1'(x_k)$ estar definida e ser não singular, para todo iterando x_k. Respectivamente, o passo de Newton para minimização está bem definido devido ao fato de a matriz hessiana $h_1''(x_k)$ estar definida e ser não singular, para todo iterando x_k. Isso limita a classe de funções às quais este método pode ser aplicado, pois ambas as hipóteses de não singularidade excluem funções da forma $H_2 : \mathbb{R}^n \to \mathbb{R}^m$ quando $n \neq m$, por exemplo. Tendo em vista ampliar a classe de funções contempladas por este método, surge a ideia de adaptá-lo, considerando o caso em que $n \neq m$ e a possibilidade de, ao invés de procurar diretamente uma solução do Problema (3.1) com $F(\cdot) := H_2(\cdot)$, buscar um ponto que minimize o quadrado da norma eucidiana do resíduo, $\frac{1}{2}\|H_2(\cdot)\|_2^2$, sendo $H_2(\cdot)$ a função residual.

3.2 O método de Gauss-Newton

Considere uma matriz $M \in \mathbb{R}^{m \times n}$, um vetor $v \in \mathbb{R}^m$ e o problema de determinar $\bar{x} \in \mathbb{R}^n$ tal que $M\bar{x} = v$. Este problema pode ser reformulado, de modo a buscar uma solução que minimize a função $h_2 : \mathbb{R}^n \to \mathbb{R}$ definida por $h_2(x) \coloneqq \frac{1}{2}\|Mx - v\|_2^2$. Sendo esta uma função convexa, a estratégia do método de mínimos quadrados é determinar um ponto $\bar{x} \in \mathbb{R}^n$ tal que satisfaça o critério de A-estacionaridade para a função h_2, ou seja, tal que:

$$M^T M\bar{x} = M^T v \, .$$

A expressão $Mx - v$ é denominada resíduo e esta estratégia leva o nome de método de mínimos quadrados linear devido ao fato de que $Mx = v$ é um sistema de equações linear e minimiza-se o quadrado da norma euclidiana do resíduo. Esta estratégia é bastante útil em casos em que o sistema $Mx = v$ possa não ter solução e, nos casos em que $M^T M$ for uma matriz não singular, determina de forma única a solução $\bar{x}$.

Consideremos agora que $H_2 : \mathbb{R}^n \to \mathbb{R}^m$ seja uma função continuamente diferenciável, cuja matriz jacobiana seja tal que $posto\ (H_2'(x)) = n$. Fixado um ponto $x_k \in \mathbb{R}^n$, buscamos pelo método de Gauss-Newton minimizar o quadrado da norma euclidiana de $H_2(x_k) + H_2'(x_k)\ (x - x_k)$, ou seja, resolver o problema:

$$\min_x h_2(x) \tag{3.4}$$

em que $h_2 : \mathbb{R}^n \to \mathbb{R}$ é definida como

$$h_2(x) \coloneqq \frac{1}{2}\ \|H_2(x_k) + H_2'(x_k)\ (x - x_k)\|_2^2 \, , \tag{3.5}$$

e assim resolver um problema de mínimos quadrados linear. Devemos determinar, então, um ponto em que seja satisfeito o critério de A-estacionaridade para h_2, ou seja, um ponto x tal que:

$$H_2'(x_k)^T\ [H_2(x_k) + H_2'(x_k)\ (x - x_k)] = 0 \, . \tag{3.6}$$

Seguindo este raciocínio, definimos o passo do método de Gauss-Newton, p_k^{GN}, a cada iterando x_k, como solução da seguinte equação:

$$H_2'(x_k)^T H_2'(x_k)\ p_k^{GN} = -H_2'(x_k)^T H_2(x_k) \, . \tag{3.7}$$

Assim, definimos o iterando seguinte por $x_{k+1} \coloneqq x_k + \alpha_k\ p_k^{GN}$, sob a condição de que $\alpha_k > 0$ seja escolhido por meio de uma estratégia de busca linear, por exemplo, as condições de Wolfe (ver Definição 3.2, a seguir). Assim, para cada iterando x_k, sabendo que p_k^{GN} é direção de descida, existirá $\alpha > 0$ suficientemente pequeno de

modo que o critério de busca linear escolhido seja satisfeito. Observe que, para que o passo do método de Gauss-Newton esteja bem definido, não é necessário que $H_2'(x_k)$ seja não singular, basta que $H_2'(x_k)^T H_2'(x_k)$ o seja e isto é equivalente a pedir que $posto(H_2'(x_k)) = n$. Entretanto, se $H_2'(x_k)$ for não singular, é possível mostrar que o passo fornecido por este método será idêntico ao do método de Newton para se determinar um zero de $H_2(\cdot)$. No caso do método de Newton para o problema de minimização, esta modificação do método também pode ser interpretada como a substituição da matriz hessiana $h_1''(x_k)$ pela expressão $h_1'(x_k)^T h_1'(x_k)$, o que reduz o custo computacional do método. Uma das condições de convergência deste método está relacionada à escolha do tamanho do passo realizado em um iterando x_k. Esta hipótese é descrita a seguir.

Definição 3.2 (Condições de Wolfe)**.** *Dada uma função $\tilde{h} : \mathbb{R}^n \to \mathbb{R}$, um ponto $x \in \mathbb{R}^n$, uma direção $p \in \mathbb{R}^n$ e um passo $\alpha > 0$, dizemos que o passo α satisfaz as condições de Wolfe se existirem constantes $0 < c_1 < c_2 < 1$ tais que:*

$$\tilde{h}(x + \alpha\, p) \leq \tilde{h}(x) + c_1\, \alpha\, \nabla\tilde{h}^T(x)\, p \tag{3.8}$$

$$e \quad c_2\, \nabla\tilde{h}(x)^T\, p \leq \nabla\tilde{h}(x + \alpha\, p)^T\, p\,. \tag{3.9}$$

Defina o conjunto $\mathcal{T} := \{x \mid \|H_2(x)\|_2 \leq \|H_2(x_0)\|_2\}$ e suponha que este seja um conjunto limitado. As condições de convergência local do método de Gauss-Newton podem ser enunciadas pelo seguinte teorema:

Teorema 3.3. *Seja $H_2 : \mathbb{R}^n \to \mathbb{R}^m$ uma função continuamente diferenciável tal que, para todo x em uma vizinhança do conjunto $\mathcal{T}$, sua matriz jacobiana seja Lipschitz contínua e satisfaça $posto(H_2'(x)) = n$. Suponhamos também que α_k satisfaça as condições de Wolfe para $\tilde{h}_2(\cdot) := \|H_2(\cdot)\|_2$ e todo $k = 0, 1, \ldots$. Então, toda sequência $\{x_k\}$ gerada pelo método de Gauss-Newton satisfaz:*

$$\lim_{k\to\infty} h_2'(x_k) = 0 \tag{3.10}$$

em que a função $h_2(\cdot)$ é definida conforme (3.5). □

Este teorema é uma versão de [15, Teorema 10.1]. Observe que este teorema não garante que uma sequência gerada pelo método de Gauss-Newton converge. Entretanto, por este teorema podemos afirmar que todo ponto de acumulação desta sequência satisfaz o critério de A-estacionaridade para a função $h_2(\cdot)$. No caso em que H_2 seja uma função de classe C^2 e $m = 1$, a convergência desta sequência está relacionada a quão bem a matriz $H_2'(x_k)^T\, H_2'(x_k)$ substitui a matriz hessiana $H_2''(x_k)$, para cada iterando x_k (ver [20, p. 356]).

Observe que o método de Gauss-Newton precisa que a matriz $H_2'(x_k)^T\ H_2'(x_k)$ seja não singular, para todo iterando x_k. Considerando os casos em que isto possa não ocorrer, faz-se interessante considerar possíveis modificações, de modo a obter um método que permaneça bem definido mesmo em casos nos quais a matriz $H_2'(x_k)^T\ H_2'(x_k)$ seja singular.

3.3 O método de Levenberg-Marquardt

Considere uma função continuamente diferenciável $H_3 : \mathbb{R}^n \to \mathbb{R}^m$. Na intenção de modificar o passo do método de Gauss-Newton para torná-lo bem definido em casos nos quais a matriz $H_3'(x_k)^T H_3'(x_k)$ possa ser singular, surge a ideia de se adicionar um parâmetro positivo $\mu_k > 0$ à equação (3.7). Isto é feito conforme a seguinte equação:

$$[H_3'(x_k)^T H_3'(x_k) + \mu_k\ I]\ p_k^{LM} = -H_3'(x_k)^T H_3(x_k) \tag{3.11}$$

em que I é a matriz identidade de ordem n e p_k^{LM} é definido como passo do método de Levenberg-Marquardt e o próximo iterando é definido como sendo $x_{k+1} \coloneqq x_k + p_k^{LM}$.

Note que o passo p_k^{LM} está bem definido, qualquer que seja a forma da matriz $H_3'(x_k)$, pois a matriz $[H_3'(x_k)^T H_3'(x_k) + \mu_k\ I]$ é sempre não singular. Assim, as propriedades de convergência local do método de Levenberg-Marquardt dependem muito da escolha da sequência $\{\mu_k\}$, sendo $\mu_k,\ k = 1, 2, \ldots, n$, parâmetros positivos tais que $\lim\limits_{k\to\infty} \mu_k = 0$.

Yamashita e Fukushima (2001)

Em [12] foi analisada uma escolha particular para estes parâmetros: $\mu_k \coloneqq \|H_3(x_k)\|_2^2$. Assim como o método de Newton, porém sob hipóteses mais fracas, o método de Levenberg-Marquardt com essa escolha de parâmetros $\mu_k > 0$ é utilizado para determinar um zero de uma função. Assim, em [12] foi garantida convergência quadrática local de uma sequência $\{x_k\}$ gerada por esse método a um ponto x^* satisfazendo (3.1), contanto que o ponto inicial estivesse suficientemente próximo de x^* e determinadas hipóteses fossem satisfeitas.

Fixada uma norma $\|\cdot\|$ em $\mathbb{R}^n$, consideraremos a distância de um ponto $x \in \mathbb{R}^n$ a um conjunto $U \subseteq \mathbb{R}^n$ como sendo

$$\operatorname{dist}(x, U) = \inf\{\|x - u\| \mid u \in U\}\ .$$

Uma das hipóteses utilizadas na prova de convergência local do método de Levenberg-Marquardt com parâmetros $\mu_k := \|H_3(x_k)\|_2^2$, apresentado em [12], é a condição de *error bound* local. Dada uma função continuamente diferenciável $H_3 : \mathbb{R}^n \to \mathbb{R}^m$, um ponto x^* que satisfaça (3.1) e $r_1 > 0$, esta condição consiste em pedir que a norma de $H_3(x)$ forneça uma cota superior para a distância do ponto x ao conjunto de soluções de (3.1) na região $\mathcal{B}(x^*, r_1)$ (ver Definição 2.31).

Uma condição de regularidade da matriz $H_3'(x)$ também é exigida em [12], a qual pode ser garantida em casos nos quais a matriz $H_3'(x)$ seja Lipschitz contínua em uma vizinhança de x^*. Esta condição é satisfeita se existirem $c_2,\ r_1 > 0$ tais que:

$$\|H_3(x') - H_3(x) - H_3'(x')\,(x - x')\| \ \leq c_2\,\|x - x'\|^2 \tag{3.12}$$

para cada par $x,\ x' \in \mathcal{B}(x^*,\ r_1)$.

Considerando a escolha do parâmetro $\mu_k := \|H_3(x_k)\|_2^2$, enunciaremos os dois principais lemas, sob os quais é fundamentada a teoria de convergência local deste método. O primeiro destes dois lemas, [12, Lema 2.1], é muito semelhante ao [8, Lema 2.1]. Uma versão deste lema é enunciada a seguir.

Lema 3.4. *Suponha que a condição de error bound local (ver Definição 2.31) e a condição (3.12) sejam satisfeitas em $\mathcal{B}(x^*, r_1)$, sob a norma euclidiana em $\mathbb{R}^n$. Dada uma sequência $\{x_k\}$ gerada pelo método de Levenberg-Marquardt com parâmetro $\mu_k = \|H_3(x_k)\|_2^2$ e uma iteração k, se $x_k \in \mathcal{B}(x^*, r_1/2)$, então p_k^{LM} solução de (3.11) será tal que :*

$$\begin{aligned} & \|p_k^{LM}\|_2 \ \leq \ c_3\ \mathit{dist}(x_k, X^*) \\ \text{e} \quad & \|H_3'(x_k)\ p_k^{LM} + H_3(x_k)\|_2 \ \leq \ c_4\ \mathit{dist}(x_k, X^*)^2 \end{aligned}$$

para algum $c_3 > 1$ e algum $c_4 > 0$. □

O segundo lema mencionado, [12, Lema 2.2], determina uma relação entre as distâncias entre iterações consecutivas e o conjunto X^*. Uma versão deste lema, essencial para provar a ordem de convergência quadrática local do método, é enunciada a seguir.

Lema 3.5. *Dada uma sequência $\{x_k\}$ gerada pelo método de Levenberg-Marquardt com parâmetros $\mu_k = \|H_3(x_k)\|_2^2$, considere duas iterações consecutivas k e $k+1$. Se $x_k,\ x_{k+1} \in \mathcal{B}(x^*,\ r_1/2)$, então*

$$\mathit{dist}(x_{k+1}, X^*) \leq c_5\ \mathit{dist}(x_k, X^*)^2$$

para algum $c_5 > 0$. □

É apresentado ainda um terceiro lema, o qual garante que, para uma escolha adequada do ponto inicial x_0, todas as iterações seguintes estarão contidas em uma região dentro da qual as hipóteses dos dois lemas anteriores são satisfeitas.

Lema 3.6. *[12, Lema 2.3] Dado $r_1 > 0$, se x_0 for um ponto suficientemente próximo de X^*, então, para toda iteração $k \geq 1$, teremos que*

$$x_k \in \mathcal{B}(x^*, r_1/2) \ . \qquad \square$$

A partir destes três lemas, podemos enunciar uma versão do teorema de convergência local apresentado em [12, Teorema 2.1].

Teorema 3.7. *Seja $H_3 : \mathbb{R}^n \to \mathbb{R}^m$ uma função continuamente diferenciável, $x^* \in \mathbb{R}^n$. Suponha que H_3 satisfaça a condição de error bound local (ver Definição 2.31) em uma vizinhança de x^* e que, nesta mesma região, a condição (3.12) seja satisfeita. Então, para uma sequência $\{x_k\}$ gerada pelo método de Levenberg-Marquardt com parâmetro $\mu_k = \|H_3(x_k)\|_2^2$ com x_0 suficientemente próximo a x^*, a sequência $\{dist(x_k, X^*)\}$ converge para zero com ordem de convergência quadrática e a sequência $\{x_k\}$ converge para um ponto $\bar{x} \in X^* \cap \mathcal{B}(x^*, r_1/2)$, também com ordem de convergência quadrática.* $\square$

Ahookhosh et al. (2019) : Método LM-AR

Em [1], foi apresentada uma escolha para o parâmetro $\mu_k > 0$ que garante convergência local superlinear ao método de Levenberg-Marquardt sob hipóteses diferentes das consideradas em [12]. Este parâmetro foi escolhido da seguinte forma:

$$\mu_k := \epsilon_k \ \|H_3(x_k)\|_2^\eta + \tilde{\epsilon}_k \ \|H_3'(x_k)^T \ H_3(x_k)\|_2^\eta,$$

sendo $\epsilon_k \in [\epsilon_{\min}, \epsilon_{\max}]$, $\tilde{\epsilon}_k \in [0, \tilde{\epsilon}_{\max}]$, para constantes positivas $\epsilon_{\min}, \epsilon_{\max}, \tilde{\epsilon}_{\max}$, e consideraremos $\eta \in \left(\dfrac{2}{q_1} - 2,\ 2\ q_1\ (1 + q_2) - \dfrac{2}{1 + q_2}\right)$, sendo $q_1, q_2 > 0$ constantes definidas pelas duas hipóteses adiante. Este método é chamado LM-AR.

Lembrando que consideramos $H_3 : \mathbb{R}^n \to \mathbb{R}^m$ sendo uma função continuamente diferenciável, dado um ponto $x^* \in X^*$ e $r_2 > 0$, abordaremos a seguir estas hipóteses, entre elas a condição de subregularidade métrica do tipo Hölder da função $H_3(\cdot)$, que pode ser considerada um enfraquecimento da condição de *error bound* local.

Hipótese 3.1 (Subregularidade métrica do tipo Hölder)**.** *Existem $q_1 \in (0, 1]$ e $\tilde{c}_1 > 0$ tais que:*

$$\tilde{c}_1 \ dist(x, X^*) \ \leq \ \|H_3(x)\|^{q_1} \qquad (3.13)$$

para todo $x \in \mathcal{B}(x^*, r_2)$.

Outra hipótese, considerada mais fraca que (3.12), tem relação com a propriedade de Hölder continuidade local (ver Definição 2.34) da matriz jacobiana de $H_3(\cdot)$. Esta propriedade é satisfeita com ordem $q_2 \in (0, 1]$ em $\mathcal{B}(x^*, r_2)$ quando H é uma função contínua e existe $\tilde{c}_2 > 0$ tal que:

$$\|H'(x) - H'(x')\| \ \leq \ \tilde{c}_2 \, \|x - x'\|^{q_2}$$

para todos $x, \ x' \in \mathcal{B}(x^*, r_2)$. Esta hipótese é enunciada abaixo.

Hipótese 3.2. *Existem* $\tilde{c}_2 > 0$ *e* $q_2 \in (0, 1]$ *tais que:*

$$\|H_3(x') - H_3(x) - H_3'(x) \ (x' - x)\| \ \leq \ \tilde{c}_2 \, \|x - x'\|^{1+q_2}$$

para todo $x, \ x' \in \mathcal{B}(x^*, \ r_2)$.

Considera-se, também, que a função $H_3(\cdot)$ seja localmente Lipschitz contínua (ver Definição 2.33) em $\mathcal{B}(x^*, r_2)$.

Hipótese 3.3 (Lipschitz continuidade local)**.** *Existe* $L > 0$ *tal que*

$$\|F(x') - F(x)\| \leq L \, \|x' - x\|$$

para todo $x, x' \in \mathcal{B}(x^*, \ r_2)$.

A teoria de convergência local superlinear deste método fundamenta-se em algumas proposições, das quais mencionaremos duas, consideradas relevantes para este trabalho. A proposição a seguir fornece um limitante superior para o tamanho do passo p_k^{LM} em função da distância do iterando x_k ao conjunto X^*.

Proposição 3.8. *[1, Proposição 2]*
Seja $x_k \notin X^*$ *um iterando gerado pelo método LM-AR sob as hipóteses 3.1, 3.2 e 3.3. Suponha que* $x_k \in \mathcal{B}(x^*, \ r_2/2)$*. Então, o passo* p_k^{LM} *dado por (3.11), com a escolha do parâmetro* $\mu_k > 0$ *do método LM-AR, satisfaz:*

$$\|p_k^{LM}\| \ \leq \ \beta_1 \, dist(x_k, X^*)^{\delta_1} \tag{3.14}$$

sendo $\beta_1 > 0$ *e* $0 < \delta_1 \leq 1$. □

Este limitante é comparável ao Lema 3.4, visto que considera hipóteses mais fracas do que aquele lema e, portanto, não se pode exigir os mesmos resultados. A segunda

proposição aqui apresentada fornece uma comparação entre as distâncias de dois iterados consecutivos ao conjunto X^*.

Proposição 3.9. *[1, Proposição 3]*
Suponha que $q_1 \geq \dfrac{1}{1+q_2}$ *e considere* $r_3 := \min\left\{\dfrac{r_2}{2}, \left(\dfrac{\tilde{c_1}^2}{2^{q_1}\tilde{c_2}^{2q_1}}\right)^{\frac{1}{2q_1(1+q_2)-2}}\right\}$. *Sejam* $x_k \notin X^*$ *e* x_{k+1} *dois iterados consecutivos gerados pelo método LM-AR tais que* $x_k,\ x_{k+1} \in \mathcal{B}(x^*, r_3)$. *Então,*

$$dist(x_{k+1}, X^*) \ \leq \ \beta_2\ dist(x_k, X^*)^{\delta_2} \tag{3.15}$$

sendo $\beta_2 > 0$ *e* $\delta_2 := \dfrac{1}{2-q_1}\min\left\{(1+q_2)\ q_1,\ (1+\eta)\ q_1,\ (1+q_2)\left(q_1 + q_1\ q_2 - \dfrac{\eta}{2}\right)\right\}$. □

Esta proposição também é essencial para se provar convergência local superlinear para o método LM-AR. Este método é interessante por ser um método adaptativo, cujos parâmetros interferem diretamente na teoria de convergência. O teorema de convergência local do LM-AR é apresentado a seguir.

Teorema 3.10. *[1, Teorema 1] Suponha que* $q_1 > \dfrac{1}{1+q_2}$. *Então, para* $r > 0$ *suficientemente pequeno, para toda sequência* $\{x_k\}$ *gerada pelo método LM-AR com* $x_0 \in \mathcal{B}(x^*, r)$ *temos que a sequência* $\{dist(x_k, X^*)\}$ *converge superlinearmente para zero com ordem* $\delta_2 > 1$, *definido pela Proposição 3.9.* □

Observação: Note que os limitantes utilizados para se provar a ordem de convergência local para estas duas escolhas do parâmetro μ_k foram essencialmente os mesmos:

- $\|p_k^{LM}\| \leq a_1\ \text{dist}(x_k, X^*)^{\xi_1}$, com $a_1 > 0$ e $\xi_1 \in (0,\ 1]$, $\forall x_k \in \mathcal{B}(x^*, \tilde{r})$;
- $\text{dist}(x_{k+1}, X^*) \leq a_2\ \text{dist}(x_k, X^*)^{\xi_2}$, com $a_2 > 0$ e $\xi_2 \in (1,\ 2]$, $\forall x_k, x_{k+1} \in \mathcal{B}(x^*, \tilde{r})$;
- $x_{k+1} \in \mathcal{B}(x^*, \tilde{r})$, sempre que $x_k \in \mathcal{B}(x^*, \tilde{r})$ e $\tilde{r} > 0$ suficientemente pequeno.

Uma comparação entre o método de Levenberg-Marquardt e os métodos LP-Newton e LP-(η, θ)

Voltemos, agora, ao raciocínio inicial sobre a estratégia dos métodos de Levenberg-Marquardt em geral, descrita por (3.11). Consideraremos uma função $F : \mathbb{R}^n \to \mathbb{R}^m$

continuamente diferenciável e, dado um iterando x_k, consideraremos agora uma função:

$$\omega_k(p) \coloneqq \|F(x_k) + F'(x_k)\, p\|_2^2 + \mu_k \, \|p\|_2^2 \, . \tag{3.16}$$

Observe que esta é uma função estritamente convexa e, portanto, o problema irrestrito:

$$\underset{p}{\text{minimizar}} \quad \omega_k(p) \tag{3.17}$$

tem solução única dada por:

$$2 \, F'(x_k)^T \, [F(x_k) + F'(x_k) \, p_k] + 2\mu_k \, p_k = 0 \, .$$

Isto significa que o Problema (3.17) fornece mesmo passo p_k^{LM} que a equação (3.11). Agora, consideramos o conjunto de soluções do problema restrito de se determinar um zero de $F(\cdot)$ por:

$$Z = \{x \in \Omega \mid \|F(x)\| = 0\} \, . \tag{3.18}$$

Considere o problema de se determinar $\bar{x} \in Z$, e suponhamos que $\Omega \subseteq \mathbb{R}^n$ seja um conjunto compacto. Então, esta estratégia pode ainda ser modificada para, a cada iteração k, resolver o seguinte subproblema:

$$\underset{x_k + p \, \in \, \Omega}{\text{minimizar}} \quad \omega_k(p) \tag{3.19}$$

sendo $\omega_k(p)$ definida como em (3.16). Denotamos o conjunto dos passos que minimizam o problema restrito (3.19) por $\mathcal{P}_k$ e fazemos $x_{k+1} \coloneqq x_k + p_k^{LMR}$, sendo $p_k^{LMR} \in \mathcal{P}_k$.

Faremos agora uma comparação entre a estratégia (3.19) e o seguinte esquema:

$$\begin{aligned} &\underset{p}{\text{minimizar}} \quad \gamma_k(p) \\ &\text{sujeito a:} \quad x_k + p \in \Omega \end{aligned} \tag{3.20}$$

sendo a função $\gamma_k(\cdot)$ definida como

$$\gamma_k(p) \coloneqq \max \left\{ \|F(x_k) + F'(x_k) \, p\|_2 \, , \; \mu_k^{1/2} \, \|p\|_2 \right\} \, .$$

É possível provar que o valor ótimo de (3.19), denotado por ω_k, será igual a zero se, e somente se, tivermos que $F(x_k) = 0$. Ao mesmo tempo, também é possível provar que o valor ótimo de (3.20), denotado por γ_k, será igual a zero se, e somente se, tivermos que $F(x_k) = 0$. Desta forma, a comparação que se faz entre estas duas estratégias não é em relação às iterações geradas por cada um deles, mas ao fato de que $\gamma_k = 0$ se, e somente se, o valor ótimo de (3.19) for $\omega_k = 0$. Sendo $\omega_k(\cdot)$ e $\gamma_k(\cdot)$ funções contínuas, faz sentido que a convergência destes métodos para uma

solução $x^* \in Z$ ocorra se, e somente se, ambas as sequências não negativas $\{\omega_k(p_k)\}$ e $\{\gamma_k(p_k)\}$ convergirem para zero.

Proposição 3.11. *O valor ótimo de (3.20) será $\gamma_k = 0$ se, e somente se, o valor ótimo de (3.19) for $\omega_k = 0$.*

Demonstração: Temos que

$$\begin{aligned} 2 \min_{x_k+p \,\in\, \Omega} \gamma_k(p) &= \min_{x_k+p \,\in\, \Omega} \left\{ 2 \ \max \left\{ \|F(x_k) + F'(x_k)\ p\|_2,\ \mu_k^{1/2} \|p\|_2 \right\} \right\} \\ &\geq \min_{x_k+p \,\in\, \Omega} \left\{ \|F(x_k) + F'(x_k)\ p\|_2 + \mu_k^{1/2} \|p\|_2 \right\} . \end{aligned}$$

Assim, pelo Corolário 2.29, temos

$$2 \min_{x_k+p \,\in\, \Omega} \gamma_k(p) \geq \min_{x_k+p \,\in\, \Omega} \sqrt{\omega_k(p)} \, .$$

E pelo Corolário 2.30 obtemos

$$\begin{aligned} 2 \min_{x_k+p \,\in\, \Omega} \gamma_k(p) &\geq \min_{x_k+p \,\in\, \Omega} \left\{ \max \left\{ \|F(x_k) + F'(x_k)\ p\|_2,\ \mu_k^{1/2} \|p\|_2 \right\} \right\} \\ &= \min_{x_k+p \,\in\, \Omega} \gamma_k(p). \end{aligned}$$

Portanto, para todo $x_k + p \in \Omega$, temos

$$2 \min_{x_k+p \,\in\, \Omega} \gamma_k(p) \geq \min_{x_k+p \,\in\, \Omega} \sqrt{\omega_k(p)} \geq \min_{x_k+p \,\in\, \Omega} \gamma_k(p). \quad \square$$

Faremos agora mais uma modificação nesta estratégia. Considere o Problema (3.20). É possível mostrar que este problema sempre tem solução, uma vez que $\Omega \subseteq \mathbb{R}^n$ é compacto e $\gamma(\cdot)$ é uma função contínua. Podemos, então, reformular o Problema (3.20) como:

$$\begin{aligned} \underset{\bar{\gamma}\ ,\ p}{\text{minimizar}}\ & \bar{\gamma} \\ \text{sujeito a: } & \|F(x_k) + F'(x_k)\ p\|_2 \leq \bar{\gamma} \\ & {\mu_k}^{1/2}\ \|p\|_2 \leq \bar{\gamma} \\ & x_k + p \in \Omega \\ & \bar{\gamma} \geq 0 \ . \end{aligned} \tag{3.21}$$

Supondo $\|F(x_k)\|_2 \neq 0$, note que (3.21) tem as mesmas soluções p que o problema:

$$\begin{aligned} \underset{\gamma\ ,\ p}{\text{minimizar}}\ & \gamma\ \|F(x_k)\|_2^2 \\ \text{sujeito a: } & \|F(x_k) + F'(x_k)\ p\|_2 \leq \gamma\ \|F(x_k)\|_2^2 \\ & {\mu_k}^{1/2}\ \|p\|_2 \leq \gamma\ \|F(x_k)\|_2^2 \\ & x_k + p \in \Omega \\ & \gamma \geq 0 \end{aligned} \tag{3.22}$$

o qual, por sua vez, tem como soluções os mesmos passos $p \in \mathbb{R}^n$ que o seguinte problema:

$$\begin{aligned} &\underset{\gamma\ ,\ p}{\text{minimizar}}\ \gamma \\ &\text{sujeito a: } \|F(x_k) + F'(x_k)\ p\|_2 \leq \gamma\ \|F(x_k)\|_2^2 \\ &\qquad\qquad \mu_k{}^{1/2}\ \|p\|_2 \leq \gamma\ \|F(x_k)\|_2^2 \\ &\qquad\qquad x_k + p \in \Omega \\ &\qquad\qquad \gamma \geq 0\ . \end{aligned} \tag{3.23}$$

Considerando, agora, a possibilidade de que a matriz jacobiana $F'(x_k)$ possa não estar definida ou ser difícil de calcular, relaxamos esta estratégia e em seu lugar utilizamos uma substituição adequada $G(x_k)$. Escolhendo o parâmetro $\mu_k := \|F(x_k)\|_2^2$, apresentado em [12] e a norma infinito, obtemos a seguinte modificação deste problema:

$$\begin{aligned} &\underset{\gamma\ ,\ p}{\text{minimizar}}\ \gamma \\ &\text{sujeito a: } \|F(x_k) + G(x_k)\ p\|_\infty \leq \gamma\ \|F(x_k)\|_\infty^2 \\ &\qquad\qquad \|p\|_\infty \leq \gamma\ \|F(x_k)\|_\infty \\ &\qquad\qquad x_k + p \in \Omega \\ &\qquad\qquad \gamma \geq 0\ . \end{aligned} \tag{3.24}$$

Este é o subproblema da versão local do método LP-Newton. Neste trabalho, buscamos relaxar ainda mais o método LP-Newton adicionando parâmetros $\eta,\ \theta > 0$ da seguinte forma:

$$\begin{aligned} &\underset{\gamma\ ,\ p}{\text{minimizar}}\ \gamma \\ &\text{sujeito a: } \|F(x_k) + G(x_k)\ p\|_\infty \leq \gamma\ \|F(x_k)\|_\infty^{\eta+\theta} \\ &\qquad\qquad \|p\|_\infty \leq \gamma\ \|F(x_k)\|_\infty^{\theta} \\ &\qquad\qquad x_k + p \in \Omega \\ &\qquad\qquad \gamma \geq 0\ . \end{aligned} \tag{3.25}$$

Note que, pela mesma estratégia utilizada anteriormente, supondo $\|F(x_k)\| \neq 0$, podemos relacionar o método LP-(η, θ) ao método de Levenberg-Marquardt com um parâmetro $\mu_k := \|F(x_k)\|^{2\eta}$. Assim, estes parâmetros são importantes para o estudo da convergência local e também para o estudo da convergência de uma versão global apresentada para este novo método proposto, ao qual chamaremos LP-(η, θ). Estudamos aqui as propriedades de convergência local do método LP-(η, θ), sob a hipótese de subregularidade métrica do tipo Hölder de uma função $F : \mathbb{R}^n \to \mathbb{R}^m$ continuamente diferenciável, cuja matriz jacobiana seja Hölder contínua em uma vizinhança de uma solução $x^* \in Z$.

4 O método LP-Newton

Apresentaremos, nesta e nas seguintes seções deste capítulo, os resultados do método LP-Newton mais relevantes para o desenvolvimento do método LP-(η, θ). Não apresentaremos aqui suas provas, pois resultados mais gerais serão vistos em detalhes no capítulo seguinte.

Seja $F : \mathbb{R}^n \to \mathbb{R}^m$ uma função continuamente diferenciável, considere o problema restrito de determinar uma solução para o sistema de equações:

$$F(x) = 0$$

com $x \in \Omega \subseteq \mathbb{R}^n$, cujo conjunto solução denotaremos por

$$Z := \{x \in \Omega \mid F(x) = 0\}. \tag{4.1}$$

No decorrer deste capítulo, consideraremos $\Omega \subseteq \mathbb{R}^n$ como sendo um conjunto poliédrico, ou seja, que o conjunto Ω possa ser descrito da seguinte forma:

$$\Omega := \{x \in \mathbb{R}^n \mid Ax \leq b\} \ ,$$

sendo $A \in \mathbb{R}^{\tilde{n} \times n}$ e $b \in \mathbb{R}^{\tilde{n}}$ uma matriz e um vetor reais, respectivamente. Também utilizaremos a notação $\|\cdot\|$ para nos referirmos à norma infinito em $\mathbb{R}^n$ ou $\mathbb{R}^m$ e a notação $\mathcal{B}(x, r)$ para denotarmos uma bola fechada em $\mathbb{R}^n$, de centro $x \in \mathbb{R}^n$ e raio $r > 0$. Também consideraremos que o conjunto $Z \subset \mathbb{R}^n$ seja diferente do vazio. Faremos menção aos critérios de estacionaridade, definidos no Capítulo 2, e consideraremos a função $f : \mathbb{R}^n \to \mathbb{R}$ definida por

$$f(x) := \|F(x)\|. \tag{4.2}$$

Localmente, o método LP-Newton consiste em resolver, a cada iteração k, o seguinte problema:

$$\begin{aligned} \underset{p,\gamma}{\text{minimizar}} \ & \gamma \\ \text{sujeito a:} \ & \|F(x_k) + G(x_k)\, p\| \ \leq \ \gamma \, \|F(x_k)\|^2 \\ & \|p\| \ \leq \ \gamma \, \|F(x_k)\| \\ & x_k + p \in \Omega \\ & \gamma \geq 0 \end{aligned} \tag{4.3}$$

onde, para cada $x_k \in \Omega$, consideramos uma substituição adequada $G(x_k)$ para a matriz jacobiana $F'(x_k)$.

Seja γ_k o valor ótimo do Problema (4.3). Localmente, a cada iterando x_k, o método LP-Newton determina um novo ponto $x_{k+1} := x_k + p_k$ tal que (p_k, γ_k) seja solução do Problema (4.3). Resolver este subproblema a cada iteração corresponde a executar o seguinte algoritmo:

Algoritmo 1: Versão local do método LP-Newton

Entrada: $x_0 \in \Omega$; $k := 0$;
Saída: $x_k \in \Omega$;
início
  enquanto $x_k \notin Z$ **faça**
    Calcule (p_k, γ_k) solução do Problema (4.3);
    $x_{k+1} := x_k + p_k$;
    $k := k + 1$;
  fim
fim

Em sua versão local, vimos que o método LP-Newton consiste em resolver, a cada iteração, um problema de programação linear. Este fato, somado à ordem de convergência local garantida pelo método, tornam o LP-Newton uma alternativa interessante ao método de Newton, pois sob abordagem e hipóteses distintas, garante-se a mesma ordem de convergência local. É devido a essa configuração dos subproblemas do método LP-Newton e pelo fato de se encaixar no esquema dos métodos do tipo Newton, o qual pode ser encontrado em [14, Algorithm 3.1], que o método LP-Newton leva este nome. Mas, além desta propriedade, o método LP-Newton também tem a vantagem de contemplar uma classe mais ampla de funções, uma vez que sua teoria de convergência é fundamentada em hipóteses mais fracas do que as do método de Newton. Vejamos a seguir quais são estas hipóteses.

4.1 As hipóteses de convergência local

Considere $x^* \in Z$ e um raio $\epsilon > 0$ arbitrários, mas fixados. Apresentamos, abaixo, as hipóteses de convergência local do método LP-Newton. A Hipótese 4.1, a seguir, é satisfeita, por exemplo, nos casos em que a função $F(x)$ seja Lipschitz contínua em uma vizinhança do ponto x^*.

Hipótese 4.1. *[8, Hipótese 1] Existe $L > 0$ tal que:*

$$\|F(x)\| \leq L\, dist(x, Z)$$

para todo $x \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$.

A Hipótese 4.1 acima e a Hipótese 4.2, a seguir, também chamada condição de *error bound*, são fatores essenciais para a teoria de convergência local do método LP-Newton. Veremos adiante que estas duas hipóteses são utilizadas tanto para se provar a ordem de convergência da sequência $\{\text{dist}(x_k, Z)\}$ para zero quanto a ordem de convergência de $\{f(x_k)\}$, também para zero. Utilizaremos, a partir deste capítulo, a noção de distância apresentada no Capítulo 2, ao considerarmos a norma infinito, a qual denotaremos simplesmente por $\|\cdot\|$, ou seja, para um ponto $x \in \mathbb{R}^n$ e um conjunto $U \subseteq \mathbb{R}^n$, consideraremos

$$\text{dist}(x, U) := \inf_{x' \in U} \|x - x'\| .$$

Assim, podemos enunciar a hipótese a seguir.

Hipótese 4.2. *[8, Hipótese 2] A função* $F(\cdot)$ *satisfaz a condição de error bound local em* $\mathcal{B}(x^*, \varepsilon) \cap \Omega$*, ou seja, existe* $\ell > 0$ *tal que:*

$$\ell \, \textit{dist}(x, Z) \leq \|F(x)\|$$

para todo $x \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$.

Considere uma versão paramétrica do subProblema (4.3), sendo x o parâmetro:

$$\begin{aligned} \underset{p,\gamma}{\text{minimizar}} \;\; & \gamma \\ \text{sujeito a:} \;\; & \|F(x) + G(x)\, p\| \;\leq\; \gamma \, \|F(x)\|^2 \\ & \|p\| \;\leq\; \gamma \, \|F(x)\| \\ & x + p \in \Omega \\ & \gamma \geq 0 \end{aligned} \tag{4.4}$$

cuja solução é denotada por $(p(x), \gamma(x))$. Temos o seguinte resultado, que será visto mais detalhadamente no próximo capítulo:

Proposição 4.1. *[8, Proposição 1 (a)] O Problema (4.4) sempre tem solução.* □

E, como consequência desta proposição, podemos enunciar o seguinte corolário:

Corolário 4.2. *O Problema (4.3) sempre tem solução e, portanto, o Algoritmo 1 está bem definido.* □

A hipótese apresentada a seguir pede que, para todo ponto x em uma vizinhança de $x^* \in Z$, o valor ótimo do Problema (4.4) seja limitado. Esta hipótese fornece um limitante que é utilizado no teorema de convergência do LP-Newton.

Hipótese 4.3. *[8, Hipótese 3] Existe* $\Gamma \geq 1$ *tal que:*

$$\gamma(x) \leq \Gamma$$

para todo $x \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$.

Veremos, no próximo capítulo, que a hipótese de que $\gamma(\cdot)$ seja limitada, ao menos em uma vizinhança de um zero de $f(\cdot)$, também é importante para a teoria de convergência global do novo método apresentado neste trabalho. Já a Hipótese 4.4, abaixo, pede que $F(x) + G(x)\ (x' - x)$ seja uma boa aproximação local para $F(x')$, sendo $x' \in \Omega$ um ponto suficientemente próximo de x.

Hipótese 4.4. *[8, Hipótese 4] Existe* $\hat{\alpha} > 0$ *tal que:*

$$x' \in \mathcal{L}(x, \alpha) := \left\{ x' \in \Omega \mid \|x' - x\| \leq \alpha, \|F(x) + G(x)\ (x' - x)\ \| \leq \alpha^2 \right\}$$

implica que $\|F(x')\| \leq \hat{\alpha}\ \alpha^2$*, para todo* $x \in (\mathcal{B}(x^*, \varepsilon) \cap \Omega) \setminus Z$ *e para todo* $\alpha \in [0, \varepsilon]$.

Assim, ao considerarmos estas hipóteses, o teorema de convergência quadrática local da distância de cada iterando x_k ao conjunto Z pode ser enunciado.

Teorema 4.3. *[8, Teorema 1] Suponha satisfeitas as Hipóteses 4.1 a 4.4. Então, existe* $r > 0$ *tal que toda sequência* $\{x_k\}$ *gerada pelo Algoritmo 1, com* $x_0 \in \Omega \cap \mathcal{B}(x^*, r)$*, converge quadraticamente para algum ponto* $\bar{x} \in Z$. □

A prova deste teorema é fundamentada em dois lemas, apresentados no mesmo trabalho. Estes lemas são enunciados a seguir.

Lema 4.4. *[8, Lema 1] Suponha que a Hipótese 4.3 seja satisfeita e defina o conjunto* $\mathcal{F}(x, \Gamma)$ *por:*

$$\mathcal{F}(x, \Gamma) := \left\{ x' \in \Omega \mid \ \|x' - x\| \leq \Gamma \|F(x)\|, \|F(x) + G(x)\ (x' - x)\| \leq \Gamma \|F(x)\|^2 \right\}.$$

Então, para cada $x \in \mathcal{B}(x^*, \varepsilon)$*, o conjunto* $\mathcal{F}(x, \Gamma)$ *será diferente do vazio. Se, adicionalmente, a Hipótese 4.1 for satisfeita, então teremos que*

$$\|F(x) + G(x)\ (x' - x)\| \ \leq \ \Gamma\ L^2\ dist(x, Z)^2 \tag{4.5}$$

$$e \quad \|x' - x\| \ \leq \ \Gamma\ L\ dist(x, Z) \tag{4.6}$$

para todo $x' \in \mathcal{F}(x, \Gamma)$. □

Observe que todos os elementos do conjunto $\mathcal{F}(x, \Gamma)$ são viáveis para o Problema (4.4) e este conjunto inclui as soluções deste problema. Desta forma, o primeiro resultado é, de certa forma, trivial, enquanto que os outros limitantes são utilizados no Teorema 4.3. Observe que estes limitantes são da forma:

- $\|F(x) + G(x)\ (x' - x)\| \ \leq \ a_1 \ \text{dist}(x, Z)^{\xi_1}$, sendo $a_1 > 0$;
- $\|x' - x\| \ \leq \ a_2 \ \text{dist}(x, Z)^{\xi_2}$, sendo $a_2 > 0$;

sendo que em (4.5) e (4.6) consideramos $\xi_1 = 2$ e $\xi_2 = 1$. Limitantes semelhantes a estes também aparecem na teoria de convergência do método LP-(η, θ), apresentado no próximo capítulo.

O Lema 4.5, enunciado abaixo, fornece informações sobre o comportamento local da sequência $\{\text{dist}(x_k,\ Z)\}$.

Lema 4.5. *[8, Lema 2] Suponha que as Hipóteses 4.1 a 4.4 sejam satisfeitas. Então, existirão $r,\ C > 0$ tais que, dado $x \in \mathcal{B}(x^*, r) \cap \Omega$, a desigualdade*

$$\textit{dist}(x', Z) \ \leq \ C \ \textit{dist}(x, Z)^2 \ \leq \ \frac{1}{2} \ \textit{dist}(x, Z)$$

é verdadeira, para todo $x' \in \mathcal{F}(x, \Gamma)$. □

Note que este lema fornece dois limitantes importantes:

$$\text{dist}(x', Z) \ \leq \ a_3 \ \text{dist}(x, Z)^2 \ , \ \text{sendo } a_3 > 0; \tag{4.7}$$

$$\text{dist}(x', Z) \ \leq \ \frac{1}{2} \ \text{dist}(x, Z) \ . \tag{4.8}$$

Enquanto que o primeiro limitante, (4.7), é utilizado para se obter a ordem de convergência quadrática local do método LP-Newton, o segundo limitante, (4.8), é suficiente para se mostrar que, de fato, a sequência $\{\text{dist}(x_k, Z)\}$ converge para zero. A teoria de convergência local do método LP-(η, θ) também utiliza limitantes semelhantes a estes.

Assim, para $x_0 \in \Omega$ escolhido suficientemente próximo do conjunto Z, a sequência $\{\text{dist}(x_k, Z)\}$ converge para zero com ordem quadrática. A convergência quadrática de $\{\|F(x_k)\|\}$ para zero se dá por meio das Hipóteses 4.1 e 4.2, como se pode ver no corolário a seguir. A mesma técnica será utilizada adiante, para provar a ordem de convergência local de $\{f(x_k)\}$ para zero, pelo método LP-(η, θ).

Corolário 4.6. *Suponha satisfeitas as Hipóteses 4.1 a 4.4. Então, existe $r > 0$ tal que toda sequência $\{x_k\}$ gerada pelo Algoritmo 1, com $x_0 \in \Omega \cap \mathcal{B}(x^*, r)$, é tal que a sequência $\{\|F(x_k)\|\}$ converge quadraticamente para zero.*

Demonstração: Pelo Teorema 4.3, para x_k suficientemente próximo de x^*, sabemos que a sequência $\{\text{dist}(x_k, Z)\}$ converge para zero com ordem de convergência quadrática. Assim, sendo x_0 suficientemente próximo de Z, para todo x_k sabemos que vale a seguinte desigualdade

$$\text{dist}(x_{k+1}, Z) \;\leq\; C\, \text{dist}(x_k, Z)^2.$$

Ora, pela Hipótese 4.1 sabemos que

$$\frac{1}{L}\, \|F(x_k)\| \;\leq\; \text{dist}(x_k, Z)$$

e assim, pelo Teorema do Confronto, a sequência $\{\|F(x_k)\|\}$ converge para zero. Pela Hipótese 4.2, obtemos

$$\text{dist}(x_{k+1}, Z)^2 \;\leq\; \ell^2\, \|F(x_{k+1})\|^2.$$

Portanto, temos que

$$\|F(x_{k+1})\| \;\leq\; L\, \ell^2\, \|F(x_k)\|^2$$

sempre que x_k esteja suficientemente próximo de x^*, o que completa a prova deste corolário. □

4.1.1 Condições suficientes para as hipóteses de convergência

Inicialmente, veremos condições sob as quais as Hipóteses 4.3 e 4.4 podem ser garantidas. Tais condições, descritas a seguir, podem ser encontradas em [8].

A Condição 4.7, a seguir, restringe a escolha da matriz $G(x)$. Esta é uma condição de suavidade que implica que a matriz jacobiana de $F(x)$ esteja definida no interior do conjunto $(\mathcal{B}(x^*, \varepsilon) \cap \Omega) \setminus Z$.

Condição 4.7. *[8, Condição 1] Existem $\varepsilon_1, \kappa_1 > 0$ tais que*

$$\|F(x') - F(x) - G(x)\,(x' - x)\| \leq \kappa_1\, \|x' - x\|^2 \tag{4.9}$$

para todo par x, x' tal que $x \in \mathcal{B}(x^*, \varepsilon_1) \cap \Omega$ e $x \in [\mathcal{B}(x^*, \varepsilon_1) \cap \Omega] \setminus Z$.

Um caso particular no qual a Condição 4.7 é satisfeita é quando a função $F(\cdot)$ é continuamente diferenciável, sua matriz jacobiana $F'(\cdot)$ é Lipschitz contínua em $\mathcal{B}(x^*, \varepsilon_1) \cap \Omega$ e escolhemos a substituição $G(x) \coloneqq F'(x)$, para todo $x \in \mathcal{B}(x^*, \varepsilon_1) \cap \Omega$. A Proposição 4.8 a seguir garante sob a condição de *error bound* que, sendo a Condição 4.7 satisfeita, as Hipóteses 4.3 e 4.4 serão também satisfeitas.

Proposição 4.8. *[8, Proposição 2] Suponha que a Condição 4.7 seja satisfeita. Então, para $\varepsilon > 0$ suficientemente pequeno, temos que:*

(i) Se a Hipótese 4.2 for satisfeita, então a Hipótese 4.3 será também satisfeita.

(ii) A Hipótese 4.4 é satisfeita. □

As hipóteses desta proposição são satisfeitas no exemplo mencionado sobre a Condição 4.7 e neste caso, aliás, pode-se provar que as Hipóteses 4.1 a 4.4 serão satisfeitas. É o que diz o Corolário 4.9, a seguir.

Corolário 4.9. *[8, Corollary 1] Suponha que a matriz jacobiana de $F(x)$ esteja definida e seja Lipschitz contínua em uma vizinhança de x^*. Então, considerando $G(x) = F'(x)$ para todo $x \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$, para $\varepsilon > 0$ suficientemente pequeno, os resultados (i) e (ii) da Proposição 4.8 serão satisfeitos.* □

A Condição 4.10, abaixo, não exige que a matriz jacobiana de $F(\cdot)$ esteja definida.

Condição 4.10. *[8, Condição 2] A função $F(\cdot)$ é Lipschitz contínua em uma vizinhança de x^* e existe $\kappa_2 > 0$ tal que*

$$\sup \{\|F(x) + V\ (x^* - x)\| \mid V \in \mathit{conv}\ \partial_B F(x)\} \leq \kappa_2\ \|x^* - x\|^2 \tag{4.10}$$

para todo $x \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$.

A Proposição 4.11 abaixo fornece condições sob as quais a Hipótese 4.3 é satisfeita, entre elas a condição de *error bound*, a Condição 4.10 e a hipótese de que x^* seja uma solução isolada de (4.1).

Proposição 4.11. *[8, Proposição 4] Suponha satisfeitas a Hipótese 4.2 e a Condição 4.10. Suponha também que x^* seja solução isolada de (4.1) e considere $G(x) \in \mathit{conv}\ \partial_B F(x)$ para todo $x \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$. Então, para $\varepsilon > 0$ suficientemente pequeno, a Hipótese 4.3 será satisfeita.* □

Vemos que esta proposição fornece um conjunto de possíveis escolhas para a matriz $G(x)$ e garante que a Hipótese 4.3 seja válida apenas em uma região suficientemente

próxima de x^*. Para a proposição seguinte, faremos uso de um conjunto relacionado a possíveis substituições da matriz jacobiana de $F(\cdot)$, nos casos em que ela não está definida, o B-subdiferencial (ver Definição 2.44).

A Proposição 4.12 pede que $posto(\partial_B F(x^*)) = n$, o que é uma forma de se generalizar a hipótese de não singularidade de $G(x)$ em casos em que m seja diferente de n ou mesmo em casos nos quais a matriz jacobiana de $F(x)$ não esteja definida.

Proposição 4.12. *[8, Proposição 5] Suponha que a Condição 4.10 seja satisfeita e que $posto(V) = n$, para todo $V \in \partial_B F(x^*)$. Escolhendo $G(x) \in \partial_B F(x)$ para cada $x \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$ teremos que, para $\varepsilon > 0$ suficientemente pequeno, a Hipótese 4.4 será satisfeita.* □

A Proposição 4.12 também determina um conjunto de possibilidades para a escolha da matriz $G(x)$ e garante que a Hipótese 4.4 seja válida apenas em uma região suficientemente próxima de x^*. Temos então o seguinte corolário, que garante as Hipóteses 4.3 e 4.4 em uma região suficientemente próxima de x^*:

Corolário 4.13. *[8, Corollary 2] Suponha que a Condição 4.10 seja satisfeita e que $posto(V) = n$, para cada $V \in \partial_B F(x^*)$. Escolhendo $G(x) \in \partial_B F(x)$ para cada $x \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$ teremos que, para $\varepsilon > 0$ suficientemente pequeno, as Hipóteses 4.3 e 4.4 serão satisfeitas.* □

Também apresentada em [8], a Condição 4.14 pode ser vista como um relaxamento da condição de semi-suavidade de $F(\cdot)$.

Condição 4.14. *[8, Condição 3] Existe $\kappa_3 > 0$ tal que, para cada $x \in \mathcal{B}(x^*, \varepsilon)$, existe $x^\diamond$ tal que*

$$\|x - x^\diamond\| \leq \kappa_3 \, dist(x, Z) \tag{4.11}$$

$$e \quad \|F(x) + G(x)\,(x^\diamond - x)\| \leq \kappa_3 \, dist(x, Z)^2. \tag{4.12}$$

Em certos casos, esta condição pode ser satisfeita tomando-se $x^\diamond = x^*$. A proposição a seguir mostra uma forte relação entre a Condição 4.14 e a Hipótese 4.3.

Proposição 4.15. *[8, Proposição 3] As seguintes afirmações são verdadeiras:*

(i) Se a Hipótese 4.2 e a Condição 4.14 forem satisfeitas, então a Hipótese 4.3 será também satisfeita.

(ii) Se as Hipóteses 4.1 e 4.3 forem satisfeitas, então a Condição 4.14 será satisfeita.

□

Uma família de funções mais específica também foi analisada em [8]: funções de classe C^1 por partes (ver Definição 2.36). Para esta classe de funções, as quatro hipóteses de convergência local deste método puderam ser verificadas. Desta forma, a globalização do algoritmo local apresentado ao início deste capítulo para funções continuamente diferenciáveis foi adaptada, de modo a contemplar também estas funções.

Estas condições e resultados são de interesse para este trabalho, pois no capítulo a seguir apresentamos o método LP-(η, θ), que lida com funções continuamente diferenciáveis $F(\cdot)$, porém consideram-se possíveis e adequadas substituições para a matriz $F'(\cdot)$. A princípio, o fato de que o método LP-Newton lida com funções $F(\cdot)$ de classe C^1 por partes é apresentado como uma característica especial deste método, porém, observa-se que uma das possibilidades de investigações futuras é a de que sejam realizadas modificações no método LP-(η, θ) de modo a lidar também com outras classes de funções, além das continuamente diferenciáveis.

Falaremos, agora, das Hipóteses 4.1 e 4.2. Como mencionado na introdução, a condição de *error bound* pode ser vista como um enfraquecimento de algumas das hipóteses do método de Newton. Em [6] são apresentadas condições sob as quais sabe-se que uma função satisfaz a condição de *error bound*. Tais critérios são mais gerais do que as aplicações encontradas atualmente para o método LP-Newton e, assim sendo, a família de funções que satisfazem a condição de *error bound* é ampla. Além disso, a Hipótese 4.1 é satisfeita quando uma função é Lipschitz contínua em uma vizinhança de um zero de $F(\cdot)$ e temos a seguinte proposição:

Proposição 4.16. *[14, Proposição 4.1] Seja $F : \mathbb{R}^n \to \mathbb{R}^m$ uma função de classe C^1 por partes, então F é localmente Lipschitz contínua. Em particular, a Hipótese 4.1 é satisfeita.* □

Em [14] também são apresentadas situações que garantem a condição de *error bound* local para uma função F de classe C^1 por partes. Entretanto, existem casos relativamente simples em que estas duas hipóteses podem não ser satisfeitas simultaneamente. Ilustramos este fato com dois exemplos a seguir.

Exemplo 4.17. *Seja $F : \mathbb{R} \to \mathbb{R}$ definida por $F(x) := |x|^{1/2}$ e considere o conjunto poliédrico $\Omega = [-1, 1]$, um intervalo fechado em $\mathbb{R}$, conforme a Figura 4.1. Observe que esta função é de classe C^1 por partes. O conjunto dos zeros de $F(\cdot)$ é $Z = \{0\}$ e a distância de um ponto $x \in \mathbb{R}$ a este conjunto poderá ser calculada como dist$(x, Z) = |x|$. Temos então que:*

(i) A Hipótese 4.1 não é satisfeita, pois para todo $L > 0$ existe $r \in (0, 1)$ tal que

$$|F(x)| = |x|^{1/2} > L\ |x| = L\ dist(x, Z)$$

para todo $x \in \mathcal{B}(0, r)$.

(ii) Em contrapartida, a condição de error bound é satisfeita com $\ell = 1$*, pois*

$$dist(x, Z) = |x| \leq |x|^{1/2} = |F(x)|\,,$$

para todo $x \in \Omega$. □

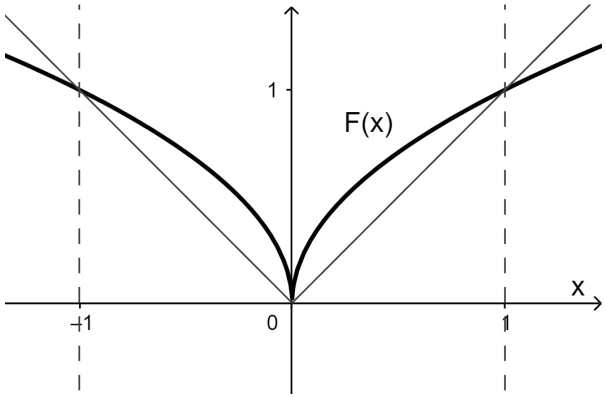


Figura 4.1: Gráfico da função $F(x) \coloneqq |x|^{1/2}$, no intervalo $[-1,\ 1]$. A condição de *error bound* local é satisfeita neste intervalo com $\ell = 1$, porém a Hipótese 4.1 não é satisfeita.

Exemplo 4.18. *Seja* $F : \mathbb{R} \to \mathbb{R}$ *definida por* $F(x) \coloneqq x^2$ *e considere o conjunto poliédrico* $\Omega = [-1, 1]$*, conforme a Figura 4.2. Observe que esta função é continuamente diferenciável. Assim como no exemplo anterior, temos* $Z = \{0\}$ *e* $dist(x, Z) = |x|$*. Temos então que:*

(i) A Hipótese 4.1 é satisfeita, pois

$$|F(x)| = |x|^2 \leq |x| = dist(x, Z)$$

para todo $x \in \Omega$.

(ii) Entretanto, a condição de error bound não é satisfeita, pois para todo $\ell > 0$ *existe* $r \in (0, 1)$ *tal que*

$$\ell\ dist(x, Z) = \ell\ |x| > |x|^2 = |F(x)|\,,$$

para todo $x \in \mathcal{B}(0,\ r)$. □

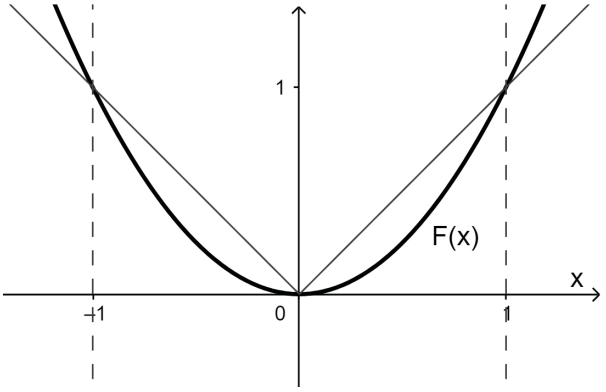


Figura 4.2: Gráfico da função $F(x) := |x|^2$, no intervalo $[-1,\ 1]$. A função $F(\cdot)$ é Lipschitz contínua neste intervalo, portanto a Hipótese 4.1 é satisfeita. Entretanto, neste intervalo $F(\cdot)$ não satisfaz a condição de *error bound* local.

No próximo capítulo, veremos que estas duas funções satisfazem uma condição mais geral do que *error bound* local, a hipótese de subregularidade métrica do tipo Hölder.

4.2 Algoritmo global

Até agora, consideramos o problema de se determinar um ponto $\bar{x} \in Z$. Este é um problema no qual se deseja determinar um zero da função $F(\cdot)$ dentro de um conjunto poliédrico $\Omega \subseteq \mathbb{R}^n$, mas, uma vez que $f(\cdot)$ é uma função não negativa e supomos que Z é diferente do vazio, também pode ser visto como um problema de otimização com restrições em que desejamos determinar um ponto de mínimo global para $f(\cdot)$.

Observando o fato de que nem sempre é fácil determinar um ponto de mínimo global para este problema, escolhe-se um critério de estacionaridade, sobre o qual nos contentamos em determinar um ponto $\bar{x}$ que o satisfaça. Do ponto de vista variacional, devido a este ser um problema com restrições, é considerado satisfatório determinar-se um ponto em que a função $f(\cdot)$ satisfaça o critério de C-estacionaridade, ou seja,

$$0 \in \partial_C f(x) + N_\Omega(x)\ . \tag{4.13}$$

Para isto, consideramos a definição de cone normal apresentada na Definição 2.26. Considerando o caso em que $F(\cdot)$ é uma função continuamente diferenciável, provou-se, sob as hipóteses de convergência global deste método, que o critério de B-estacionaridade, definido no Capítulo 2, é satisfeito para todo ponto C-estacionário de $f(\cdot) = \|F(\cdot)\|$. Este fato é interessante, pois em geral o critério de B-estacionaridade é mais forte que o critério de C-estacionaridade, mas neste caso ambos são equivalentes devido à função $f(\cdot)$ ser Clarke regular (ver Definição 2.41).

Assim, definimos também uma função $\triangle^{LPN}(\cdot) : \Omega \to \mathbb{R}$, a qual relaciona-se ao critério de C-estacionaridade e serve como uma medida de decréscimo de $f(x)$ na direção $p(x)$, sendo $(p(x),\ \gamma(x))$ solução do problema paramétrico (4.4) em um ponto $x \in \Omega$. A importância da função $\triangle^{LPN}(\cdot)$ como critério de parada será discutida no capítulo seguinte, onde é apresentado um critério de parada semelhante.

Seja $(p(x), \gamma(x))$ solução do Problema (4.4), definimos a função $\triangle^{LPN} : \Omega \setminus Z \to \mathbb{R}$ por:

$$\triangle^{LPN}(x) \coloneqq -f(x)\ (1 - \gamma(x) f(x))\ . \tag{4.14}$$

Ainda em relação ao algoritmo global do método LP-Newton, dado um iterando x_k e um valor positivo α, considere a seguinte desigualdade:

$$f(x_k + \alpha_k\ p_k)\ \leq\ f(x_k) + \sigma\ \alpha_k\ \triangle^{LPN}(x_k). \tag{4.15}$$

Não falaremos, neste trabalho, sobre o algoritmo global do método LP-Newton para funções de classe C^1 por partes, pois no capítulo a seguir apresentaremos uma estratégia de globalização parecida com a do Algoritmo 2. Mais informações sobre este outro algoritmo podem ser encontradas em [10].

Uma versão do algoritmo global, desenvolvido em [10, Algorithm 1] para o caso particular de uma função $F(\cdot)$ continuamente diferenciável, pode ser descrita em pseudocódigo da seguinte forma:

Algoritmo 2: Versão global do método LP-Newton

Entrada: $x_0 \in \Omega$; $\sigma\ ,\ \beta \in (0,\ 1)$; $k \coloneqq 0$;
Saída: $x_k \in \Omega$;
início
  enquanto $x_k \notin Z$ **faça**
    Calcule (p_k, γ_k), solução do Problema (4.3);
    Calcule $\triangle^{LPN}(x_k)$, definido em (4.14);
    se $\triangle^{LPN}(x_k) = 0$ **então**
      PARE;
    fim
    $\alpha \coloneqq 1$;
    enquanto *a desigualdade (4.15) não for satisfeita* **faça**
      $\alpha \coloneqq \beta \cdot \alpha$;
    fim
    $x_{k+1} \coloneqq x_k + \alpha\ p_k$;
    $k \coloneqq k + 1$;
  fim
fim

A prova de convergência global do método LP-Newton é obtida a partir de alguns lemas, proposições e teoremas, os quais serão brevemente discutidos a seguir. Primeiramente é apresentado o fato de que $\gamma(\cdot)$, definida como valor ótimo do Problema (4.4), é uma função contínua, para todo ponto $x \in \Omega \setminus Z$.

Lema 4.19. *[10, Lema 3.1] Seja $x \in \Omega \setminus Z$, então a função $\gamma(\cdot)$, definida como valor ótimo do Problema (4.4), é contínua em x, em relação a Ω.* □

Disto, podemos afirmar também que a função $\triangle^{LPN}(\cdot)$, definida por (4.14), é contínua para todo ponto $x \in \Omega \setminus Z$. O próximo lema utiliza a estrutura da derivada direcional de $f(\cdot)$ e do Problema (4.3). Assim, dado um iterando x_k, este lema fornece um limitante superior para a derivada direcional $f'(x_k; p_k)$, sendo (p_k, γ_k) solução do Problema (4.3), e por meio de um teorema de programação linear, uma relação entre a função função $\triangle^{LPN}(\cdot)$ e o critério de C-estacionaridade.

Lema 4.20. *[10, Lema 3.2] Dado um iterando $x_k \in \Omega$, gerado pelo Algoritmo 2, e uma solução (p_k, γ_k) do Problema (4.3), considere as funções $f(\cdot)$ e $\triangle^{LPN}(\cdot)$ definidas por (4.2) e (4.14), respectivamente. Temos que:*

$$f'(x_k;\ p_k)\ \leq\ \triangle^{LPN}(x_k).$$

Além disso, se existir um ponto $\bar{x} \in \Omega$ que não seja C-estacionário, então existirão $\delta,\ C\ >\ 0$ tais que, se x_k for um ponto suficientemente próximo deste ponto $\bar{x}$, teremos que

$$\triangle^{LPN}(x_k)\ \leq\ -\delta\ \ e\ \ \|p_k\|\ \leq\ C.$$ □

Dada um iterando x_k suficientemente próximo do conjunto Z e supondo que $x_k \in \Omega \setminus Z$, teremos $\triangle^{LPN}(x_k) \leq 0$. Desta forma, obtemos pelo Lema 4.20 que, caso a condição

$$\triangle^{LPN}(x) = 0 \tag{4.16}$$

não seja satisfeita, p_k será direção de descida para $f(x_k)$. E, caso esta condição seja satisfeita, pode-se mostrar que o critério de C-estacionaridade para $f(\cdot)$ é satisfeito em x_k, o que será apresentado também no capítulo seguinte. Em particular para funções de classe C^1 por partes, isto é equivalente ao critério de B-estacionaridade, definido no Capítulo 2, devido ao fato de que esta classe de funções é Clarke regular (ver Definição 2.41).

O teorema a seguir é o principal resultado apresentado em [10], e trata-se do teorema de convergência global deste método, relacionado ao Algoritmo 2. Este resultado afirma que, de fato, este algoritmo determina um iterando x_k no qual é satisfeito o critério de B-estacionaridade pela função $f(\cdot)$, verificado pela condição

(4.16) ou, caso isso não ocorra, fornece ao menos algum ponto de acumulação $\bar{x}$ da sequência $\{x_k\}$ gerada por este algoritmo, em que tal critério é satisfeito.

Teorema 4.21. *[10, Teorema 3.1] Seja $F : \mathbb{R}^n \to \mathbb{R}^m$ uma função continuamente diferenciável, $f(x) := \|F(x)\|$, para todo $x \in \Omega$. Então, as seguintes afirmações são válidas:*

(i) O Algoritmo 2 está bem definido.

(ii) O Algoritmo 2 ou termina em um iterando $\bar{x} := x_k \in \Omega$ tal que satisfaz o critério de B-estacionaridade ou gera um ponto de acumulação $\bar{x}$ no qual o critério de B-estacionaridade é satisfeito. Em ambos os casos, o ponto $\bar{x}$ satisfaz a condição (4.16). □

Considerando o fato de que o método global apresentado em [10] foi também adaptado para funções F de classe C^1 por partes, diferentes aplicações são encontradas. De modo geral, este método é aplicável ao problema de se determinar um zero de um sistema de equações não lineares, entretanto por meio de reformulações adequadas, outros problemas podem ser tratados como tal. Uma dessas aplicações são os problemas de Equilíbrio de Nash que, sob condições adequadas, assumem a forma de um sistema de complementaridade não linear [13]. De modo geral, problemas de complementaridade em $\mathbb{R}^n$ restritos a um conjunto poliédrico podem ser reformulados como um sistema de equações não lineares. De modo análogo, existem também reformulações de problemas KKT como sistemas de equação não lineares, e assim o método LP-Newton pode ser considerado aplicável também a esta classe de problemas.

5 Um novo método: LP-(η, θ)

Seja $F : \mathbb{R}^n \to \mathbb{R}^m$ uma função contínua, $\Omega \subseteq \mathbb{R}^n$ um conjunto poliédrico. Denotaremos por $\|\cdot\|$ a norma infinito. Sob hipóteses adequadas, abordaremos o problema de determinar um zero de $F(\cdot)$ em Ω. Desta forma, propomos aqui uma modificação do método LP-Newton, a qual apresenta ordem de convergência local superlinear mesmo em casos nos quais a hipótese de *error bound* local possa não valer, mas em seu lugar se possa garantir a hipótese de subregularidade métrica do tipo Hölder (ver Definição 2.32). Esta hipótese é estritamente mais fraca do que a condição de *error bound* local, sendo que esta última pode ser vista como um caso particular da primeira. Um exemplo desta situação é apresentado a seguir.

Exemplo 5.1. *Considere a função $F : \mathbb{R} \to \mathbb{R}$ definida por $F(x) := x^2$, conforme a Figura 5.1. Sabemos que o único zero desta função é $x^* = 0$, portanto, $Z = \{0\}$ e $dist(x, Z) = |x|$. Além disso, para todo $x \in \mathcal{B}(x^*, 1)$, temos que*

$$dist(x, Z) = |x| = |F(x)|^{1/2}.$$

Portanto a função $F(\cdot)$ satisfaz a hipótese de subregularidade métrica do tipo Hölder em uma vizinhança de x^* com constante $\ell = 1$, para qualquer $q \in (0,\ 1/2]$. Entretanto, vimos pelo Exemplo 4.18 que a condição de error bound local não é satisfeita por $F(\cdot)$ em nenhuma vizinhança de x^*. □

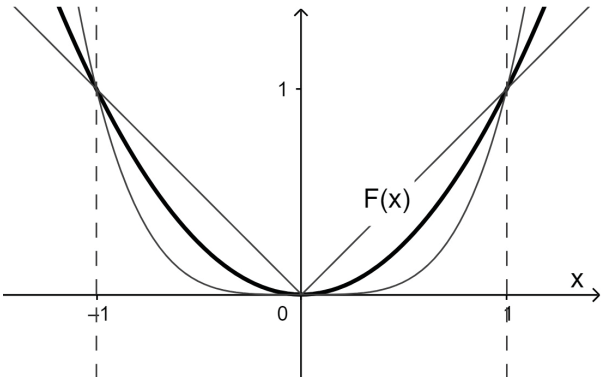


Figura 5.1: Exemplo 5.1. Consideramos a função $F(x) := |x|^2$ e as desigualdades: $|x|^4 \leq |x|^2 \leq |x|$ no intervalo $\Omega := [0,\ 1]$. Estas desigualdades podem ser interpretadas como: $\text{dist}(x, Z)^4 \leq F(x) \leq \text{dist}(x, Z)$ e representam, respectivamente, a condição de subregularidade métrica do tipo Hölder com constantes $c_1 = 1$ e $q_1 = 1/4$, e a Hipótese 4.1 do método LP-Newton, com constante $L = 1$.

Considerando o problema de se determinar um zero da função $F(x) := x^2$, observe que, exatamente no único zero desta função, a matriz jacobiana de $F(\cdot)$ é singular, de modo que, para este problema o método de Newton não está bem definido. Ainda mais, pelo fato de que a função $F(\cdot)$ não satisfaz a condição de *error bound* local em uma vizinhança deste zero, temos que o método LP-Newton também não é aplicável para este problema. Entretanto, o método LP-(η, θ), apresentado neste capítulo, resolve este problema sem a necessidade de reformulações.

Durante este capítulo, consideraremos que $F : \mathbb{R}^n \to \mathbb{R}^m$ seja uma função continuamente diferenciável, logo sempre existirá sua matriz jacobiana $F'(\cdot)$. Mesmo assim, consideraremos ainda possíveis substituições $G(\cdot)$. Estas substituições são de grande importância, pois $F'(\cdot)$ pode ser difícil de calcular e também pelo fato de que, numericamente, em geral, lida-se com aproximações e não com valores exatos. Reiteramos que toda a teoria contida neste capítulo é válida quando substituímos $G(\cdot)$ por $F'(\cdot)$.

5.1 Algoritmo e teoria de convergência local

Considerando o problema de determinar um zero de uma função $F : \mathbb{R}^n \to \mathbb{R}^m$, neste capítulo introduziremos os parâmetros $\eta, \theta > 0$, utilizados na modificação do método LP-Newton, tendo o objetivo de lidar com dois fatores:

- subregularidade métrica do tipo Hölder da função $F(\cdot)$, ao invés da condição de *error bound* local;

- Hölder continuidade da matriz jacobiana de $F(x)$ (ou de sua substituição por uma matriz $G(x)$), ao invés de Lipschitz continuidade local desta matriz.

Observe que o método local aqui apresentado se torna o LP-Newton, quando escolhemos $\eta = \theta = 1$.

A modificação aqui proposta sobre o subproblema local do método LP-Newton é descrita a seguir. Dado um iterando x_k, localmente o método LP-(η, θ) resolve o subproblema:

$$\begin{aligned} \underset{p\,,\,\gamma}{\text{minimizar}} \;\; & \gamma \\ \text{sujeito a:} \;\; & \|F(x_k) + G(x_k)\,p\| \;\leq\; \gamma\,\|F(x_k)\|^{\theta+\eta} \\ & \|p\| \;\leq\; \gamma\,\|F(x_k)\|^{\theta} \\ & x_k + p \in \Omega \\ & \gamma \geq 0 \end{aligned} \tag{5.1}$$

onde η,$\theta > 0$ são parâmetros pré-determinados e $\Omega \subseteq \mathbb{R}^n$ é um conjunto poliédrico (ver Definição 2.24). O algoritmo local resultante desta modificação pode ser definido como:

Algoritmo 3: Versão local do método LP-(η, θ)

Entrada: $x_0 \in \Omega$; $\eta > 0$; $\theta > 0$; $k := 0$;
Saída: $x_k \in \Omega$
início
  enquanto $x_k \notin Z$ **faça**
    Calcule (p_k, γ_k) solução do Problema (5.1);
    $x_{k+1} := x_k + p_k$;
    $k := k + 1$;
  fim
fim

Para a teoria de convergência local deste método, consideramos algumas hipóteses semelhantes às propostas em [1]. Estas foram citadas no Capítulo 2, quando se comentava sobre o método de Levenberg-Marquardt sob a hipótese de subregularidade métrica do tipo Hölder. As hipóteses a seguir são enunciadas considerando algum $x^* \in Z$ e algum $\varepsilon > 0$ fixados.

Hipótese 5.1 (Subregularidade métrica do tipo Hölder)**.** *Existem* $c_1 > 0$, $q_1 \in (0, 1]$ *tais que*

$$c_1 \, dist(x, Z) \leq \|F(x)\|^{q_1}$$

para todo $x \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$.

É interessante notar que, devido às possíveis modificações do método de Levenberg-Marquardt apresentadas no Capítulo 2, a prova de convergência deste método tem relação com a prova de convergência do método de Levenberg-Marquardt com um parâmetro $\mu_k := \|F(x_k)\|^{\eta}$. Por este motivo, o parâmetro $\theta > 0$ não influencia na prova de convergência local aqui apresentada. Entretanto, é possível que outras estratégias de estudo da convergência deste método forneçam limitantes ou estimativas melhores a partir do parâmetro θ. Uma hipótese mais fraca do que a exigência de que a matriz jacobiana $F'(\cdot)$ seja Lipschitz contínua em uma vizinhança de um zero de $F(\cdot)$ é a de que $F'(\cdot)$ seja Hölder contínua. Isto ocorre nos casos em que a hipótese a seguir é satisfeita, considerando-se $G(\cdot) = F'(\cdot)$.

Hipótese 5.2. *Existem* $c_2 > 0$, $q_2 \in (0, 1]$ *tais que a matriz* $G(x)$ *satisfaz*

$$\|F(x') - F(x) - G(x)\,(x' - x)\| \leq c_2 \, \|x' - x\|^{1+q_2}$$

para todo $x, x' \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$.

Uma terceira hipótese a ser considerada, relacionada à hipótese anterior, é a de que a função $F(\cdot)$ seja Lipschitz contínua em uma vizinhança de $x^* \in Z$, ou seja:

Hipótese 5.3. *Existe $c_3 > 0$, tal que*

$$\|F(x) - F(x')\| \leq c_3 \, \|x - x'\|$$

para todo $x, x' \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$.

Nos próximos resultados, faremos uso de algumas funções auxiliares. Considerando um iterando $x_k \in \Omega$ do método LP-(η, θ) tal que $x_k \notin Z$, definimos as funções $\gamma_{k,1}$, $\gamma_{k,2}$, $\gamma_k : \mathbb{R}^n \to \mathbb{R}$ como:

$$\gamma_{k,1}(p) \coloneqq \frac{\|F(x_k) + G(x_k)\, p\|}{\|F(x_k)\|^{\eta+\theta}} \, , \tag{5.2}$$

$$\gamma_{k,2}(p) \coloneqq \frac{\|p\|}{\|F(x_k)\|^{\theta}} \, , \tag{5.3}$$

$$\gamma_k(p) \coloneqq \max\{\gamma_{k,1}(p),\ \gamma_{k,2}(p)\} \ . \tag{5.4}$$

Inicialmente, mostraremos que o Algoritmo 3 está bem definido por meio da proposição a seguir.

Proposição 5.2. *Suponha que as Hipóteses 5.1, 5.2 e 5.3 sejam satisfeitas. Então, dado um iterando $x_k \in \Omega$, o Problema (5.1) tem solução e, sendo (p_k, γ_k) solução do Problema (5.1), teremos que $F(x_k) = 0$ se, e somente se, $\gamma_k = 0$.*

Demonstração: Seja $x_k \in \Omega$ um iterando do método LP-(η, θ). Mostraremos inicialmente que o Problema (5.1) tem solução. No caso em que $x_k \in Z$, esta afirmação é trivial. Supondo que $x_k \notin Z$, consideremos a função $\gamma_k(\cdot)$ definida por (5.4). Lembrando do desenvolvimento apresentado ao final do Capítulo 3, sabemos que o valor ótimo do Problema (5.1) é dado por:

$$\gamma_k \coloneqq \min_{x_k + p \,\in\, \Omega} \gamma_k(p) \ . \tag{5.5}$$

Mostraremos então que é possível determinar um limitante superior finito para γ_k, em função de x_k. Considerando o caso particular $p = 0 \in \mathbb{R}^n$, vemos que $x_k + 0 = x_k \in \Omega$. Além disso, temos que

$$\gamma_k \leq \gamma_k(0) = \|F(x_k)\|^{1-\eta-\theta} \, ,$$

o que mostra que, de fato, este limitante existe. Observe agora que, por (5.1), para qualquer $p \in \mathbb{R}^n$ temos que

$$\gamma_{k,2}(p) \leq \gamma_k(p).$$

Portanto, não há perda de generalidade para (5.5) em restringir o domínio de p de modo que

$$\gamma_{k,2}(p) \leq \|F(x_k)\|^{1-\eta-\theta},$$

ou seja, em considerarmos $p \in \Omega_k := (\Omega - x_k) \cap \mathcal{B}(0, \|F(x_k)\|^{1-\eta})$, que é compacto. Assim, a partir de agora consideraremos a reformulação do Problema (5.5) dada por:

$$\gamma_k = \min_{p \in \Omega_k} \gamma_k(p) \,. \tag{5.6}$$

Observação: Note que, a cada iteração k, o conjunto denotado por Ω_k é uma translação do conjunto $\Omega \cap \mathcal{B}(x_k, \|F(x_k)\|^{1-\eta})$. Este, por sua vez, é compacto, visto que Ω é fechado e a bola $\mathcal{B}(x_k, \|F(x_k)\|^{1-\eta})$ é fechada e limitada.

Sendo x_k fixado, note que as funções $\gamma_{k,1}(\cdot)$ e $\gamma_{k,1}(\cdot)$ são contínuas. Assim, temos também que a função $\gamma_k(\cdot)$ é contínua. Sendo Ω_k compacto, obtemos pelo teorema de Weierstrass que o Problema (5.1) tem solução.

Verificaremos agora a segunda afirmação, de que $x_k \in Z$ se, e somente se, o valor ótimo do Problema (5.1) for $\gamma_k = 0$. Suponha que $\|F(x_k)\| = 0$, então o problema de otimização (5.1) se reduz ao seguinte problema:

$$\begin{aligned} &\underset{\gamma,p}{\text{minimizar}}\ \gamma \\ &\text{sujeito a: } \gamma \geq 0 \end{aligned} \tag{5.7}$$

cuja solução ótima é, claramente, $\gamma = 0$. Reciprocamente, consideremos $x_k \in \Omega$ tal que a solução ótima do Problema (5.1) seja $\gamma_k = 0$ e suponhamos por absurdo que $x_k \notin Z$. Sabendo que este problema sempre tem solução, existirá $\bar{p} \in \Omega_k$ tal que $\gamma_k(\bar{p}) = 0$. Pelas restrições do Problema (5.1), isto significa que $\|\bar{p}\| = 0$ e $\|F(x_k) + G(x_k)\,\bar{p}\| = 0$. Mas, desta forma teremos $\|F(x_k)\| = \|F(x_k) + G(x_k)\,\bar{p}\| = 0$ e, portanto, $x_k \in Z$, o que é uma contradição. Portanto, o valor ótimo do Problema (5.1) é $\gamma_k = 0$ se, e somente se, $x_k \in Z$. □

Dada uma solução (p_k, γ_k) do Subproblema (5.1), a prova de convergência local do método aqui proposto faz uso de dois limitantes principais, correspondentes aos dois lemas a seguir. Estes lemas têm relação com a convergência local do método, portanto não há perda de generalidade em supor que as hipóteses de convergência são válidas em uma região suficientemente próxima do conjunto Z. Para o lema a seguir, por exemplo, é suficiente pedir que estas hipóteses sejam satisfeitas em $\mathcal{B}(x^*, \varepsilon) \cap \Omega$, sendo $x^* \in Z$ e $\varepsilon < 1$.

Lema 5.3. *Dados $x^* \in Z$, $\varepsilon < 1$, seja $x_k \notin Z$ um iterando gerado pelo método LP-(η, θ) sob as Hipóteses 5.1 a 5.3, $x_k \in \mathcal{B}(x^*, \varepsilon/2)$. Seja (p_k, γ_k) uma solução de (5.1).*

Então,

$$\|p_k\| \leq \beta_1 \, dist(x_k, Z)^{\delta_1}$$

com $\beta_1 := \max\left\{c_3 \, c_1^{-\frac{\eta}{q_1}}, \, 1\right\}$ *e* $\delta_1 := \min\left\{1 + q_2 - \frac{\eta}{q_1}, \, 1\right\}$.

Demonstração: Para cada k, sendo Z um conjunto fechado, escolha $\bar{x}_k \in Z$ tal que $\|x_k - \bar{x}_k\| = \text{dist}(x_k, Z)$. Como $x_k \in \mathcal{B}(x^*, \varepsilon/2)$, temos que:

$$\bar{x}_k \in \mathcal{B}(x^*, \varepsilon/2) \cap \Omega \, .$$

Além disso,

$$\|\bar{x}_k - x_k\| = \text{dist}(x_k, Z) \leq \|x_k - x^*\| \leq \varepsilon/2.$$

Agora, para cada p tal que $x_k + p \in \Omega$, defina a função $\gamma_k(p)$ por (5.4). Como (p_k, γ_k) é solução do Subproblema (5.1) e $(\bar{x}_k - x_k) + x_k = \bar{x}_k \in \Omega$, temos que:

$$\gamma_k(p_k) \leq \gamma_k(\bar{x}_k - x_k)$$

Então, pelas restrições do Subproblema (5.1) e pela definição da função $\gamma_k(\cdot)$,

$$\begin{aligned} \|p_k\| \leq \gamma_k \, \|F(x_k)\|^\theta &= \max\left\{\frac{\|F(x_k) + G(x_k)\,(\bar{x}_k - x_k)\|}{\|F(x_k)\|^{\theta+\eta}}, \, \frac{\|\bar{x}_k - x_k\|}{\|F(x_k)\|^\theta}\right\} \cdot \|F(x_k)\|^\theta \\ &= \max\left\{\|F(x_k) + G(x_k)\,(\bar{x}_k - x_k)\| \, \|F(x_k)\|^{-\eta}, \, \|\bar{x}_k - x_k\|\right\} \end{aligned} \tag{5.8}$$

Temos, por (5.8) e pela hipótese de subregularidade métrica do tipo Hölder, que:

$$\begin{aligned} \|p_k\| \; &\leq \; \max\left\{\|F(x_k) + G(x_k)\,(\bar{x}_k - x_k)\| \; (c_1 \, \text{dist}(x_k, Z))^{-\frac{\eta}{q_1}} \, , \; \|\bar{x}_k - x_k\|\right\} \\ &= \; \max\left\{\|F(\bar{x}_k) - F(x_k) - G(x_k)\,(\bar{x}_k - x_k)\| \; (c_1 \, \text{dist}(x_k, Z))^{-\frac{\eta}{q_1}} \, , \; \|\bar{x}_k - x_k\|\right\} \, . \end{aligned}$$

Lembrando que $\bar{x} \in \mathcal{B}(x^*, \varepsilon) \cap \Omega$, pela Hipótese 5.2 temos que

$$\|p_k\| \leq \max\left\{c_2 \, \|\bar{x}_k - x_k\|^{1+q_2} \; (c_1 \, \text{dist}(x_k, Z))^{-\frac{\eta}{q_1}} \, , \; \|\bar{x}_k - x_k\|\right\}.$$

Assim, como $\|\bar{x}_k - x_k\| = \text{dist}(x_k, Z)$, obtemos

$$\|p_k\| \leq \max\left\{c_1^{-\frac{\eta}{q_1}} \; c_2 \, \text{dist}(x_k, Z)^{1+q_2-\frac{\eta}{q_1}}, \; \text{dist}(x_k, Z)\right\} \leq \beta_1 \, \text{dist}(x_k, Z)^{\delta_1}, \tag{5.9}$$

onde $\beta_1 := \max\left\{c_1^{-\frac{\eta}{q_1}} c_2, \; 1\right\}$ e $\delta_1 := \min\left\{1 + q_2 - \frac{\eta}{q_1}, \; 1\right\}$.

□

O lema a seguir fornece um limitante superior finito para $\text{dist}(x_{k+1}, Z)$ em função de $\text{dist}(x_k, Z)$. Assim como no lema anterior, consideramos que as hipóteses deste lema sejam satisfeitas em uma região suficientemente próxima de Z, sendo $x_k, x_{k+1} \in \Omega$ dois pontos nesta região.

Lema 5.4. *Sejam $x_k \in \Omega \setminus Z$ e $x_{k+1} \in \Omega$ duas iterações consecutivas geradas pelo método LP-(η, θ) sob as Hipóteses 5.1 a 5.3, tais que $x_k, x_{k+1} \in \mathcal{B}(x_k, \varepsilon/2)$ e $\varepsilon < 1$. Então,*

$$\mathit{dist}(x_{k+1}, Z) \leq \beta_2 \, \mathit{dist}(x_k, Z)^{\delta_2}$$

com $\beta_2 \coloneqq \max\left\{{c_2}^{q_1},\ L^{\eta\ q_1},\ {c_2}^{q_1}\ {c_1}^{(1+q_2)\ q_1},\ 1\right\}$

e $\delta_2 \coloneqq \min\left\{(1+q_2)\ q_1, (1+\eta)\ q_1, (1+q_2)\left(1+q_2-\frac{\eta}{q_1}\right) q_1\right\}$.

Demonstração: Seja $\bar{x}_k \in Z$ tal que $\|x_k - \bar{x}_k\| = \text{dist}(x_k, Z)$. Defina $\gamma_k(p)$, para $p \in \Omega$, tal como em (5.4). Temos, pelas restrições do Problema (5.1), que:

$$\|F(x_k) + G(x_k)\ p_k\| \leq \gamma_k\ \|F(x_k)\|^{\theta+\eta}.$$

Sabendo que (p_k, γ_k) é solução do Problema (5.1), pela definição da função $\gamma_k(p)$, temos que

$$\begin{aligned}
\|F(x_k) + G(x_k)p_k\| &\leq \gamma_k(p_k)\|F(x_k)\|^{\theta+\eta} \leq \gamma_k(x_k - \bar{x}_k)\|F(x_k)\|^{\theta+\eta} \\
&= \max\{\gamma_{k,1}(\bar{x}_k - x_k),\ \gamma_{k,2}(\bar{x}_k - x_k)\} \cdot \|F(x_k)\|^{\theta+\eta} \\
&= \max\{\|F(x_k) + G(x_k)\ (\bar{x}_k - x_k)\|,\ \|\bar{x}_k - x_k\| \cdot \|F(x_k)\|^{\eta}\} \qquad (5.10) \\
&= \max\{\|F(x_k) + G(x_k)\ (\bar{x}_k - x_k)\|,\ \|\bar{x}_k - x_k\| \cdot \|F(x_k)\|^{\eta}\}\ . \qquad (5.11)
\end{aligned}$$

Observe agora que por (5.10) e pela Hipótese 5.2, temos que

$$\begin{aligned}
\|F(x_k) + G(x_k)\ p_k\| &= \|F(\bar{x}_k) - F(x_k) - G(x_k)\ p_k\| \\
&\leq \max\left\{c_2\ \|\bar{x}_k - x_k\|^{1+q_2},\ \|\bar{x}_k - x_k\|\ \|F(x_k)\|^{\eta}\right\}.
\end{aligned}$$

Dessa forma, pela Hipótese 5.3,

$$\|F(x_k) + G(x_k)\ p_k\| \leq \max\left\{c_2\ \|\bar{x}_k - x_k\|^{1+q_2},\ L^{\eta}\ \|\bar{x}_k - x_k\|^{1+\eta}\right\}.$$

Sabendo que $\|\bar{x}_k - x_k\| = \text{dist}(x_k, Z)$, então,

$$\begin{aligned}
\|F(x_k) + G(x_k)\ p_k\| &\leq \max\left\{c_2\ \text{dist}(x_k, Z)^{1+q_2},\ L^{\eta}\ \text{dist}(x_k, Z)^{1+\eta}\right\} \\
&\leq \bar{\beta}_2\ \text{dist}(x_k, Z)^{\bar{\delta}_2} \qquad (5.12)
\end{aligned}$$

onde $\bar{\beta}_2 \coloneqq \max\{c_2,\ L^{\eta},\ 1\}$ e $\bar{\delta}_2 \coloneqq \min\{1+q_2, 1+\eta\}$. Assim, pela Hipótese 5.1, temos que:

$$[c_1\ \text{dist}(x_{k+1}, Z)]^{\frac{1}{q_1}} \leq \|F(x_{k+1})\| = \|F(x_k + p_k)\|$$

$$\leq \|F(x_k) + G(x_k)\, p_k\| + \|F(x_k + p_k) - F(x_k) - G(x_k)\, p_k\|. \tag{5.13}$$

Então, pela Hipótese 5.2, obtemos

$$[c_1\ \text{dist}(x_{k+1}, Z)]^{\frac{1}{q_1}} \leq \|F(x_k) + G(x_k)\, p_k\| + c_2\, \|p_k\|^{1+q_2}.$$

Esta última relação, por (5.12) e pelo Lema 5.3, respectivamente, nos dá que

$$\begin{aligned}[c_1\ \text{dist}(x_{k+1}, Z)]^{\frac{1}{q_1}} &\leq \bar{\beta}_2\ \text{dist}(x_k, Z)^{\bar{\delta}_2} + c_2\ {\beta_1}^{1+q_2}\ \text{dist}(x_k, Z)^{(1+q_2)\ \delta_1} \\ &\leq \tilde{\beta}_2\ \text{dist}(x_k, Z)^{\tilde{\delta}_2} \end{aligned} \tag{5.14}$$

sendo $\tilde{\beta}_2 := \max\left\{\bar{\beta}_2,\ c_2\ {\beta_1}^{1+q_2}\right\}$ e $\tilde{\delta}_2 := \min\left\{\bar{\delta}_2,\ (1+q_2)\ \delta_1\right\}$.

Portanto, por (5.14), temos que:

$$\text{dist}(x_{k+1}, Z) \leq \tilde{\beta}_2^{q_1}\ \text{dist}(x_k, Z)^{\tilde{\delta}_2\ q_1} = \beta_2\ \text{dist}(x_k, Z)^{\delta_2}$$

sendo $\beta_2 := \max\left\{{c_2}^{q_1},\ {L^{\eta}}^{\ q_1},\ {c_2}^{q_1}\ {\beta_1}^{(1+q_2)\ q_1},\ 1\right\}$
e $\delta_2 := \min\left\{(1+q_2)\ q_1,\ (1+\eta)\ q_1,\ (1+q_2)\ \left(1 + q_2 - \dfrac{\eta}{q_1}\right) q_1\right\}$. □

Com isso, podemos provar convergência local superlinear para este método, por meio do teorema a seguir.

Teorema 5.5. *Seja uma sequência $\{x_k\}$ gerada pelo método LP-(η, θ). Suponha que*

$$\frac{1}{1+q_2} < q_1\ ,\quad \frac{1}{q_1} - 1 < \eta \leq q_1\, q_2$$

e que as Hipóteses 5.1 a 5.3 sejam satisfeitas, para todo x_k, $k = 1, 2, \ldots$. Então existirá $\tilde{r} > 0$ suficientemente pequeno tal que, se o ponto inicial $x_0 \in \mathcal{B}(x^*, \tilde{r})$, teremos que:

(i) a sequência $\{dist(x_k, Z)\}$ converge para zero superlinearmente com ordem de convergência

$$\delta_2 = \min\left\{(1+q_2)\ q_1,\ (1+\eta)\ q_1,\ (1+q_2)\left(1 + q_2 - \frac{\eta}{q_1}\right) q_1\right\}\ ;$$

(ii) a sequência $\{x_k\}$ converge para algum $\bar{x} \in Z$;

(iii) a sequência $\{x_k\}$ apresenta ordem de convergência superlinear δ_2 .

Demonstração:

(i) Sejam δ_1, β_1, δ_2, $\beta_2 > 0$ definidos tais como nos Lemas 5.3 e 5.4. Como $\delta_2 > 1$, temos, para i suficientemente grande, que $\delta_1 {\delta_2}^i > i$. Como $\sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i = 1$, deduzimos que:

$$\sigma := \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{\delta_1 {\delta_2}^i} \leq \sum_{i=1}^{j} \left(\frac{1}{2}\right)^{\delta_1 {\delta_2}^i} + \sum_{i=j+1}^{\infty} \left(\frac{1}{2}\right)^i < \infty$$

Agora, defina

$$\bar{r} := \min \left\{ \frac{1}{2}\, {\beta_2}^{\frac{1}{1-\delta_2}}\, \varepsilon,\ \left(\frac{\varepsilon}{2\,(1+\beta_1+2^{\delta_1}\cdot\beta_1\,\sigma)} \right)^{\frac{1}{\delta_1}} \right\}. \tag{5.15}$$

Como, pelo Lema 5.4, $\beta_2 \geq 1$ e, pelas hipóteses deste teorema, temos que $\delta_2 > 1$, segue que:

$$\frac{1}{2}\, {\beta_2}^{\frac{1}{1-\delta_2}}\, \varepsilon < \frac{1}{2}\, \varepsilon.$$

Além disso, como β_1, $\sigma > 0$, $\varepsilon < 1$ e $\delta_1 \leq 1$, temos que:

$$\left(\frac{\varepsilon}{2(1+\beta_1+2^{\delta_1}\cdot\beta_1\cdot\sigma)} \right)^{\frac{1}{\delta_1}} < \left(\frac{\varepsilon}{2}\right)^{\frac{1}{\delta_1}} < \frac{\varepsilon}{2}.$$

Portanto, sabemos que $\bar{r} \in (0, \varepsilon/2)$. Seja $x^* \in Z$, $x_0 \in \mathcal{B}(x^*, \bar{r})$ e considere uma sequência $\{x_k\}$ gerada pelo método LP-(η, θ). Sabemos pela Proposição 5.2 que esta sequência está bem definida. Dado um iterando qualquer $x_k \in \Omega$, mostraremos por indução que $x_k \in \mathcal{B}(x^*, \varepsilon/2)$. Para a iteração 1, sabemos que

$$\|x_1 - x^*\| = \|x_0 + p_0 - x^*\| \leq \|x_0 - x^*\| + \|p_0\|.$$

Assim, pelo Lema 5.4,

$$\|x_1 - x^*\| = \|x_0 + p_0 - x^*\| \leq \bar{r} + \beta_1\, \text{dist}(x_0, Z)^{\delta_1}.$$

E, como $\bar{r} < 1$,

$$\|x_1 - x^*\| = \|x_0 + p_0 - x^*\| \leq \bar{r}^{\delta_1} + \beta_1\, \|x_0 - x^*\|^{\delta_1} \leq (1+\beta_1)\, \bar{r}^{\delta_1} \leq \varepsilon/2 \tag{5.16}$$

Suponhamos, por hipótese de indução, que $x_i \in \mathcal{B}\,(x^*, \varepsilon/2)$, para cada $i = 1, 2, \cdots, k$ e defina: $\delta_i := \sum_{j=0}^{i-1} {\delta_2}^j$. Pelo Lema 5.4, temos que:

$$\begin{aligned} \text{dist}(x_i, Z) &\leq \beta_2\, \text{dist}(x_{i-1}, Z)^{\delta_2} \leq {\beta_2}^{1+\delta_2}\, \text{dist}(x_{i-2}, Z)^{{\delta_2}^2} \\ &\leq \cdots \leq {\beta_2}^{\delta_i}\, \text{dist}(x_0, Z)^{{\delta_2}^i} \leq {\beta_2}^{\delta_i}\, \|x_0 - x^*\|^{{\delta_2}^i} \\ &= {\beta_2}^{\frac{{\delta_2}^i - 1}{\delta_2 - 1}}\, \|x_0 - x^*\|^{{\delta_2}^i} \end{aligned} \tag{5.17}$$

Sabendo que $x_0 \in \mathcal{B}(x^*, \bar{r})$, então,

$$\text{dist}(x_i, Z) \leq \left(\frac{1}{2\,\bar{r}}\right)^{\delta_2{}^{i}-1} \bar{r}^{\delta_2{}^{i}} = \left(\frac{1}{2}\right)^{\delta_2{}^{i}-1} \bar{r} \tag{5.18}$$

Portanto, temos que

$$\text{dist}(x_i, Z)^{\delta_1} \leq (2\bar{r})^{\delta_1} \cdot \left(\frac{1}{2}\right)^{\delta_1\delta_2{}^{i}}, \tag{5.19}$$

para todo $i = 1, 2, \ldots, k$. Assim, por desigualdades triangulares, obtemos

$$\|x_{k+1} - x^*\| \leq \|x_1 - x^*\| + \sum_{i=1}^{k} \|p_i\|.$$

Desta última relação, por (5.16) e pelo Lema 5.4, respectivamente, temos

$$\|x_{k+1} - x^*\| \leq (1 + \beta_1)\, \bar{r}^{\delta_1} + \beta_1 \sum_{i=1}^{k} \text{dist}(x_i, Z)^{\delta_1}.$$

Assim, por (5.19),

$$\begin{aligned}\|x_{k+1} - x^*\| &\leq (1 + \beta_1)\, \bar{r}^{\delta_1} + \beta_1\, (2\bar{r})^{\delta_1} \sum_{i=1}^{k} \left(\frac{1}{2}\right)^{\delta_1\delta_2{}^{i}} \leq (1 + \beta_1)\, \bar{r}^{\delta_1} + \beta_1 (2\bar{r})^{\delta_1}\sigma \\ &= (1 + \beta_1 + 2^{\delta_1}\beta_1\, \sigma)\, \bar{r}^{\delta_1} \leq \varepsilon/2 \end{aligned} \tag{5.20}$$

Isto completa a indução. Portanto, dado um ponto $x_0 \in \Omega \cap \mathcal{B}(x^*, \bar{r})$, todas as iterações x_k seguintes geradas por este método serão tais que $x_k \in \mathcal{B}(x^*, \varepsilon/2)$.

Assim, sabemos que, para todo iterando x_k, as hipóteses dos Lemas 5.4 e 5.3 são satisfeitas. Para cada iteração k, temos que

$$\text{dist}(x_{k+1}, Z) \leq \beta_2\, \text{dist}(x_k, Z)^{\delta_2}\,. \tag{5.21}$$

Sendo, pelas hipóteses deste teorema, $\delta_2 > 1$, para x_0 suficientemente próximo de Z teremos

$$\frac{1}{2} > \beta_2\, \text{dist}(x_k, Z)^{\delta_2 - 1}\,.$$

Supondo que $x_k \notin Z$ para toda iteração k, por (5.21) teremos que

$$\frac{\text{dist}(x_{k+1}, Z)}{\text{dist}(x_k, Z)} < \frac{1}{2}\,.$$

Logo, a sequência $\{\text{dist}(x_k, Z)\}$ converge para zero. Além disso, por (5.21), obtemos então que a sequência $\{\text{dist}(x_k, Z)\}$ converge superlinearmente com ordem de convergência δ_2.

(ii) Mostraremos agora que a sequência $\{x_k\}$ de fato converge. Pelo Lema 5.3, temos que:

$$\sum_{i=1}^{\infty} \|p_i\| \leq \beta_1 \sum_{i=1}^{\infty} \text{dist}(x_i, Z)^{\delta_1} \leq \beta_1 \, \sigma \, (2\,\bar{r})^{\delta_1} < +\infty$$

Defina a sequência $\{s_k\}$ por $s_k := \sum_{i=1}^{k} \|p_i\|$. Temos, pela desigualdade acima, que esta é uma sequência de Cauchy. De fato, para todo $k, j \in \mathbb{N} \cup \{0\}$, temos que

$$\begin{aligned} \|x_{k+j} - x_k\| &= \|x_{k+j-1} + p_{k+j-1} - x_k\| \leq \|p_{k+j-1}\| + \|x_{k+j-1} - x_k\| \\ &\leq \cdots \leq \sum_{i=k}^{k+j-1} \|p_i\| = s_{k+j-1} - s_{k-1} \, . \end{aligned} \tag{5.22}$$

E assim, obtemos que $\{x_k\}$ é uma sequência de Cauchy. Portanto, $\{x_k\}$ é uma sequência convergente, ou seja, existe $\bar{x} \in \bar{\Omega}$ tal que

$$\lim_{k\to\infty} x_k = \bar{x} \, . \tag{5.23}$$

(iii) Provaremos, agora, a ordem de convergência desta sequência. Sendo $\bar{x}$ determinado por (5.23), como $\eta \leq q_1 \cdot q_2$, como $x_k \in \mathcal{B}\,(x^*, \varepsilon/2)$, para toda iteração k, e $\text{dist}(x_k, Z) \to 0$, segue da continuidade de $\text{dist}(\,\cdot\,, Z)$ que

$$\bar{x} \in Z \cap \mathcal{B}\,(x^*, \varepsilon/2) \, .$$

E, como $\eta \leq q_1 \cdot q_2$, temos $1 + q_2 - \dfrac{\eta}{q_1} > 1$, logo, $\delta_1 = 1$. Assim, substituindo este resultado em (5.22) e fazendo $j \to \infty$, obtemos

$$\|x_k - \bar{x}\| \;=\; \lim_{i\to\infty} \|x_k - x_i\| \;\leq\; \sum_{i=k}^{\infty} \|p_i\| \;\leq\; \beta_1 \sum_{i=k}^{\infty} \text{dist}(x_i, Z) \, .$$

Sabemos que a sequência $\{\text{dist}(x_i, Z)\}$ converge superlinearmente para zero. Então, para k suficientemente grande, teremos que $\text{dist}(x_{k+1}, Z) \leq \frac{1}{2}\,\text{dist}(x_k, Z)$. Desta forma,

$$\begin{aligned} \|x_k - \bar{x}\| \;&\leq\; \beta_1 \sum_{i=k}^{\infty} \frac{1}{2^{i-k}}\, \text{dist}(x_k, Z) \;=\; \beta_1\, \text{dist}(x_k, Z) \sum_{i=k}^{\infty} \frac{1}{2^{i-k}} \\ &= 2\,\beta_1\, \text{dist}(x_k, Z) \;\leq\; 2\,\beta_1\,\beta_2\, \text{dist}(x_{k-1}, Z)^{\delta_2} \\ &\leq\; 2\,\beta_1\,\beta_2\, \|x_{k-1} - \bar{x}\|^{\delta_2} \, . \end{aligned}$$

E assim concluímos a prova de convergência local superlinear da sequência $\{x_k\}$, com ordem δ_2, para algum ponto $\bar{x} \in Z$. □

Deste teorema, temos o seguinte corolário, que fornece condições para a convergência superlinear de $\{\|F(x_k)\|\}$ para zero.

Corolário 5.6. *Seja $\{x_k\}$ uma sequência gerada pelo método LP-(η, θ), tal que $x_0 \in \mathcal{B}(x^*, \bar{r})$, sendo $\bar{r} > 0$ descrito por (5.15), $x^* \in Z$. Suponha também que $\dfrac{1}{1+q_2} < {q_1}^2$ e $\dfrac{1}{{q_1}^2} - 1 < \eta \leq q_1\, q_2$. Então, a sequência $\{\|F(x_k)\|\}$ convergirá para zero superlinearmente com ordem de convergência $q_1\, \delta_2$.*

Demonstração: Como $q_1 \in (0, 1]$ e $\dfrac{1}{1+q_2} < {q_1}^2$, temos que $\dfrac{1}{1+q_2} < q_1$.

E, como $\dfrac{1}{{q_1}^2} - 1 < \eta \leq q_1\, q_2$, temos $\dfrac{1-q_1}{q_1} < \eta \leq q_1\, q_2$.

Assim, sabemos que as hipóteses do Teorema 5.5 são satisfeitas. Por este teorema, obtemos

$$c_3 \text{ dist}(x_{k+1}, Z) \leq c_3\, \beta_2 \text{ dist}(x_k, Z)^{\delta_2} ,$$

para todo x_k. Pela Hipótese 5.3, temos que

$$\|F(x_{k+1})\| \leq c_3 \text{ dist}(x_{k+1}, Z),$$

logo, substituindo o resultado acima na primeira equação, obtemos

$$\|F(x_{k+1})\| \leq c_3\, \beta_2 \text{ dist}(x_k, Z)^{\delta_2}.$$

Como, pela Hipótese 5.1,

$$\text{dist}(x_k, Z)^{\delta_2} \leq {c_1}^{-\delta_2}\, \|F(x_k)\|^{q_1\, \delta_2},$$

segue que

$$\|F(x_{k+1})\| \leq c_3\, \beta_2\, {c_1}^{-\delta_2}\, \|F(x_k)\|^{q_1\, \delta_2},$$

para cada iterando x_k. Como $q_1\, \delta_2 > 1$, segue a tese. □

Para que as hipóteses do Teorema 5.5 sejam satisfeitas, a região viável para as constantes q_1 e q_2 relacionadas às Hipóteses 5.1 e 5.2 é representada pela Figura 5.2, a seguir.

Já a região viável para o parâmetro $\eta > 0$, dependendo de q_1 e q_2 pelo Teorema 5.5 é representada pela Figura 5.3, a seguir. Sendo, por exemplo $q_2 = 0.75$, os pares viáveis $(q_1,\ \eta)$ são representados pela região sombreada abaixo da reta indicada por "$q_2 = 0.75$".

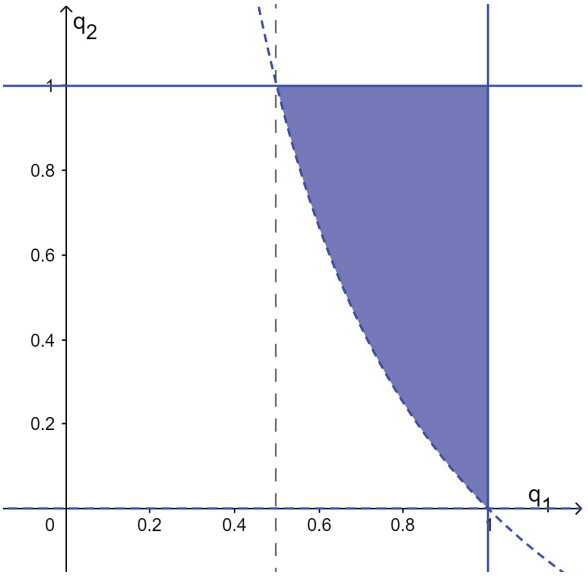


Figura 5.2: Região viável de q_1 e q_2 para o Teorema 5.5. Consideramos a desigualdade $\dfrac{1}{1+q_2} < q_1$.

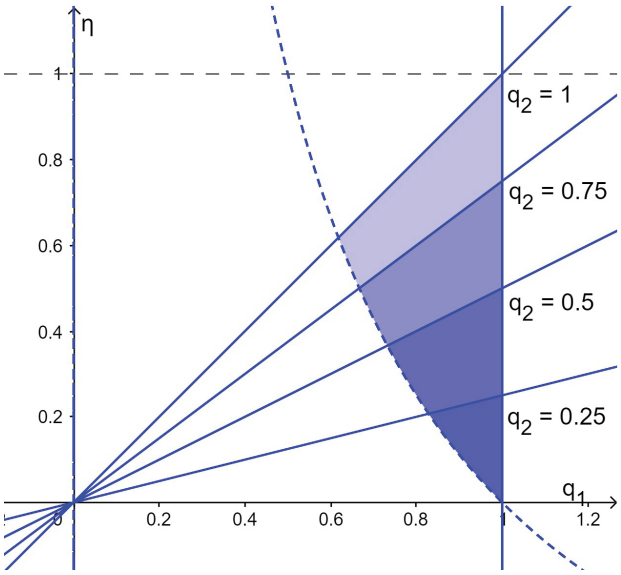


Figura 5.3: Região viável de η para o Teorema 5.5. Consideramos as desigualdades: $\dfrac{1}{q_1} - 1 < \eta \leq q_1\ q_2$. Para cada $q_2 \in (0, 1]$ fixado, a segunda desigualdade determina a região viável para os pares $(q_1,\ \eta)$ pela área sombreada abaixo de cada reta $\eta = q_2\ q_1$.

Para o Corolário 5.6, a região viável para as constantes q_1 e q_2, relacionada às Hipóteses 5.1 e 5.2, é menor do que a região viável para o Teorema 5.5 e é representada pela Figura 5.4. Já a região viável para o parâmetro $\eta > 0$, depende das constantes q_1 e q_2, relacionadas às Hipóteses 5.1 e 5.2 e é representada pela Figura 5.5.

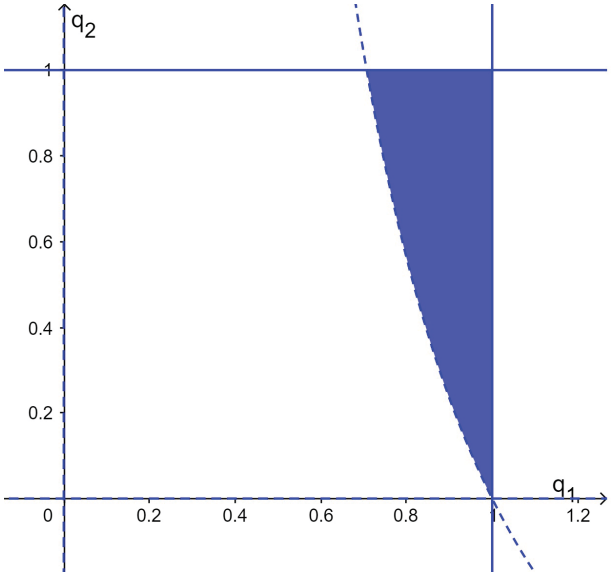


Figura 5.4: Região viável de q_1 e q_2 para o Corolário 5.6. Consideramos a desigualdade $\frac{1}{1+q_2} < {q_1}^2$.

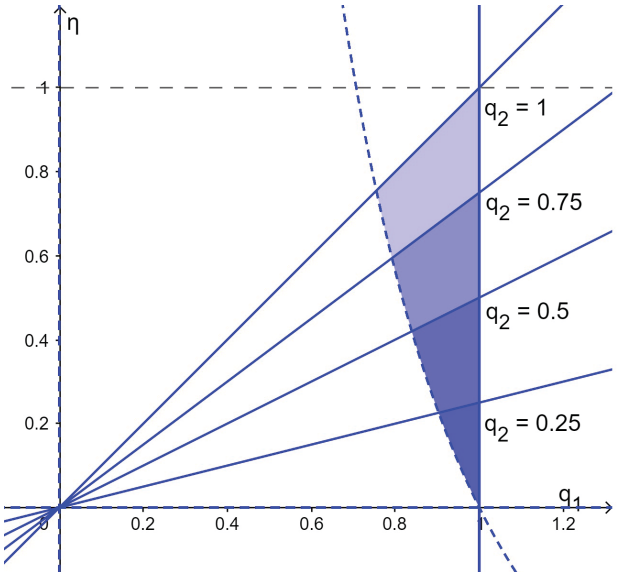


Figura 5.5: Região viável de η para o Corolário 5.6. Consideramos as desigualdades: $\frac{1}{{q_1}^2} - 1 < \eta \leq q_1\, q_2$. Para cada $q_2 \in (0, 1]$ fixado, a segunda desigualdade determina a região viável para os pares $(q_1,\ \eta)$ pela área sombreada abaixo de cada reta $\eta = q_2\, q_1$.

É interessante lembrar que o parâmetro θ não tem relevância para a teoria de convergência local do método LP-(η, θ) de modo que poderíamos simplesmente fixar $\theta = 1$ e estes resultados permaneceriam válidos. Entretanto, o parâmetro θ é fundamental para a teoria de convergência global do método LP-(η, θ), que será apresentada na próxima seção.

5.2 Algoritmo e teoria de convergência global

Na intenção de tornar aplicável a teoria local apresentada na seção anterior, apresentamos nesta seção um algoritmo globalmente convergente do método LP-(η, θ). Assim como anteriormente, consideraremos uma função continuamente diferenciável $F : \mathbb{R}^n \to \mathbb{R}^m$ e, durante esta seção, consideraremos uma função $f : \mathbb{R}^n \to \mathbb{R}$ definida por

$$f(\cdot) \coloneqq \|F(\cdot)\|. \tag{5.24}$$

Desta forma, o problema de se determinar um zero de $F(\cdot)$ é equivalente ao problema de se determinar um minimizador global de $f(\cdot)$ e consideraremos os mesmos conjuntos $\Omega \subseteq \mathbb{R}^n$ poliédrico e $Z \subset \mathbb{R}^n$ definidos anteriormente.

Consideraremos a função $\triangle : \Omega \to \mathbb{R}$, definida por

$$\triangle(x) \coloneqq f(x) - \gamma(x)\, f(x)^{\theta+\eta}\,, \tag{5.25}$$

onde $\gamma(x)$ é o valor ótimo do Problema (5.31) para dado $x \in \Omega$. Assim, podemos apresentar o algoritmo com convergência global do método LP-(η, θ), abaixo.

Algoritmo 4: Versão global do método LP-(η, θ)

Entrada: $x_0 \in \Omega$; $\eta > 0$; $\theta > 0$; $\sigma,\ \beta \in (0,1)$; $k \coloneqq 0$;
Saída: $x_k \in \Omega$;
início
  enquanto $x_k \notin Z$ **faça**
    Calcule (p_k, γ_k), solução do Problema (5.1);
    Calcule $\triangle(x_k)$, definido em (5.25);
    se $\triangle(x_k) = 0$ **então**
      PARE;
    fim
    $\alpha = 1$;
    enquanto $f(x_k + \alpha\, p_k) > f(x_k) - \alpha\, \sigma\, \triangle(x_k)$ **faça**
      $\alpha = \beta \cdot \alpha$;
    fim
    $x_{k+1} \coloneqq x_k + \alpha\, p_k$;
    $k \coloneqq k + 1$;
  fim
fim

Várias estratégias podem ser consideradas para se globalizar o método apresentado na seção anterior. O Algoritmo 4, acima, é uma modificação do Algoritmo 2, visto no

Capítulo 4. A estratégia aqui utilizada busca determinar uma solução do problema de otimização com restrições:

$$\begin{aligned} &\text{minimizar} \quad f(x) \\ &\text{sujeito a:} \quad x \in \Omega \end{aligned} \tag{5.26}$$

Dado $x \in \mathbb{R}^n$, podemos visualizar $F(x)$ como um vetor em $\mathbb{R}^m$,

$$F(x) = [F_1(x) \ \cdots \ F_m(x)]^T.$$

Desta forma, lembrando que $\|\cdot\|$ é considerada aqui como a norma infinito, podemos também visualizar $f(x)$ como

$$f(x) = \max\{|F_1(x)|, \ \ldots \ , \ |F_m(x)|\}.$$

Duas questões naturais então surgem: que condições de otimalidade considerar e que medidas de otimalidade verificáveis são apropriadas? Quando as funções $F_i(\cdot)$, $i = 1, \ldots, m$, são Lipschitz contínuas, a função $f(\cdot)$ também o é. Nestes casos, podemos definir a derivada generalizada de Clarke de $f(x)$ em uma direção $d \in \mathbb{R}^n$ (ver Definição 2.40) como:

$$f^o(x; d) := \limsup_{x' \to x \ , \ t \to 0^+} \frac{f(x' + td) - f(x')}{t}. \tag{5.27}$$

Descreveremos, agora, o cone normal a Ω em um ponto $x \in \Omega$. Este conjunto já foi mencionado anteriormente, e, devido às particularidades do conjunto $\Omega \subseteq \mathbb{R}^n$, é apresentado aqui sob uma formulação mais específica. O cone normal, juntamente com o subdiferencial de Clarke, serão utilizados para definir uma medida de otimalidade para o problema abordado nesta seção. Considerando o fato de que o conjunto $\Omega \subseteq \mathbb{R}^n$ é um conjunto poliédrico, pode ser descrito da seguinte forma:

$$\Omega = \{x \in \mathbb{R}^n \mid A\,x \leq b\} \ , \tag{5.28}$$

em que $A \in \mathbb{R}^{s \times n}$, $b \in \mathbb{R}^s$, com $s \in \mathbb{N}$. Denotamos cada linha da matriz A por A_r e cada elemento do vetor b por b_r, $r = 1, \ldots, s$.

Assim, considerando (5.28), podemos definir o cone normal a Ω em um ponto $x \in \Omega$ como:

$$N_\Omega(x) = \left\{A^T\, y \mid y \in \mathbb{R}^s \text{ é tal que } y \geq 0 \text{ e } (A\,x - b)^T\, y = 0\right\} \ . \tag{5.29}$$

Agora também podemos definir o subdiferencial de Clarke. Sob determinadas condições, este conjunto permite escolher substituições $G(\cdot)$ para a matriz jacobiana de $F(\cdot)$, $F'(\cdot)$. Denotamos o cone tangente ao conjunto Ω em um ponto $x \in \Omega$

por $T_\Omega(x)$ (ver Definição 2.25). Para cada ponto $x \in \Omega$, sabemos que a derivada direcional $f'(x;\ d)$ está definida para todo $d \in T_\Omega$, pois $F(\cdot)$ é continuamente diferenciável. Assim, descrevemos o subdiferencial de $f(\cdot)$ em x por:

$$\partial_C f(x) \coloneqq \left\{ w \in \mathbb{R}^n \mid f^o(x;d) \geq w^T d \ \ \forall\, d \in \mathbb{R}^n \right\} . \tag{5.30}$$

Utilizaremos, no decorrer deste capítulo, a condição também conhecida como Clarke-estacionaridade, ou simplesmente C-estacionaridade, para o conjunto Ω, tal como definido em (5.28). Esta condição de otimalidade já foi mencionada no Capítulo 4, portanto, para este caso particular consideraremos (4.13). Esta é uma condição necessária de otimalidade para o problema de se determinar um ponto de mínimo global do Problema (5.26), ou seja, um ponto $\bar{x} \in Z$. Entretanto, descrita desta forma, pode ser difícil determinar quando esta condição é satisfeita. Por este motivo, é útil definir uma condição facilmente verificável associada ao critério de estacionaridade utilizado. Considere a versão paramétrica do Subproblema (5.1), tendo $x \in \Omega$ como parâmetro:

$$\begin{aligned} \underset{p\ ,\ \gamma}{\text{minimizar}} \ & \gamma \\ \text{sujeito a:} \ & \|F(x) + G(x)\ p\| \ \leq \ \gamma\ \|F(x)\|^{\theta+\eta} \\ & \|p\| \ \leq \ \gamma\ \|F(x)\|^{\theta} \\ & x + p \in \Omega \\ & \gamma \geq 0 \end{aligned} \tag{5.31}$$

Uma boa medida de estacionaridade associada ao critério de C-estacionaridade é baseada na função $\triangle : \Omega \to \mathbb{R}$, definida em (5.25).

A razão pela qual consideramos $\triangle(x)$ como medida de otimalidade vem do fato de que, sob determinadas hipóteses, $\triangle(x) = 0$ se, e somente se, $x \in \Omega$ for um ponto C-estacionário para $f(\cdot)$. Isto será provado mais adiante, no Lema 5.10. Temos, de (5.25), o resultado abaixo.

Proposição 5.7. *Para todo $x \in \Omega$, temos $\triangle(x) \geq 0$.*

Demonstração: Lembremos da Proposição 5.2, onde vimos que, dado $x_k \in \Omega$, temos que $\gamma(x_k) \leq \gamma_k(0) = f(x_k)^{1-\theta-\eta}$. Por raciocínio análogo temos então que, para todo $x \in \Omega$, $\gamma(x) \leq f(x)^{1-\theta-\eta}$. Substituindo esta desigualdade em (5.25), obtemos que $\triangle(x) \geq 0$, para todo $x \in \Omega$. □

Considerando a versão paramétrica do subproblema do método LP-(η, θ) descrita por (5.31), uma questão a ser analisada é como o valor ótimo $\gamma(x)$ do problema paramétrico (5.31) se comporta sob perturbações contínuas do parâmetro x. A proposição a seguir é uma adaptação de [3, Proposition 4.4].

Proposição 5.8. *Considere o problema*

$$\begin{aligned} &\underset{p \in P}{\text{minimizar}} \;\; h(p,x) \\ &\text{sujeito a:} \;\; H(p,x) \in K \end{aligned} \tag{5.32}$$

onde $h : P \times \Omega \to \mathbb{R}$ *é uma função qualquer,* $H : P \times \Omega \to \mathbb{R}^t$ *é uma função contínua,* $K \subseteq \mathbb{R}^t$, $\mathcal{F}(x) \coloneqq \{p \in P \;:\; H(p,x) \in K\}$ *é um conjunto viável e* $S(x) \coloneqq \underset{p \in \mathcal{F}(x)}{\arg\min}\, \gamma$ *é o conjunto de soluções deste problema, quando avaliado em um ponto* $x \in \Omega$. *Seja* $p \in P$. *Suponha que:*

(i) a função objetivo deste problema, denotada por h(p, x), seja contínua em $P \times \Omega$;

(ii) A multifunção[1] $\mathcal{F}(x)$ *seja fechada*[2];

(iii) existam $\alpha > 0$ *e um compacto* $C \subseteq \Omega$ *tais que, para todo* x *em uma vizinhança de* $\bar{x}$ *o conjunto*

$$lev_\alpha(\cdot, x) \coloneqq \{p \in \mathcal{F}(x) \mid h(p,x) \leq \alpha\}$$

seja não vazio e contido em C;

(iv) Para toda vizinhança de $S(\bar{x})$ *exista* $r > 0$ *tal que, para todo* $x \in B(\bar{x}, r)$, *tenhamos*

$$\mathcal{F}(x) \cap B(\bar{x}, r) \neq \emptyset \,.$$

Então, a função $v(x)$, *definida como valor ótimo do Problema (5.32) avaliado em* x, *é contínua em* $\bar{x}$ *em relação a* Ω. □

Desta forma, um caso particular da proposição acima é apresentado a seguir.

Proposição 5.9. *Seja* $\bar{x} \in \Omega \setminus Z$. *Se* $G(\cdot)$ *for uma função contínua em uma vizinhança de* $\bar{x}$, *então* $\gamma(\cdot)$ *é contínua em* $\bar{x}$ *em relação a* Ω, *ou seja, para toda sequência* $\{x_k\}$ *em* Ω *tal que* $x_k \to \bar{x}$, *teremos que* $\gamma(x_k) \to \gamma(\bar{x})$, *sendo* $\gamma(x)$ *o valor ótimo do Problema (5.31) para dado* $x \in \Omega$.

Demonstração: Mostraremos aqui que as hipóteses da Proposição 5.8 são satisfeitas. O último item da Proposição 5.8 é provado de forma análoga à encontrada em [10, Lemma 3.1]. Considere a versão paramétrica do subproblema do método LP-(η, θ) descrita por (5.31), sendo $x \in \Omega$ o parâmetro. Pelo argumento apresentado na Proposição 5.2, consideraremos, sem perda de generalidade, a restrição adicional

[1] Uma multifunção é uma relação que faz corresponder, a cada elemento x de um conjunto X, um subconjunto $Y(x) \subseteq Y$.
[2] Dizemos que uma multifunção $Y(\cdot)$ é fechada quando seu gráfico é fechado em $X \times Y$. (Ver [4, p. 4].)

$p \in \Omega_x \coloneqq (\Omega - x) \cap \mathcal{B}(0, \|F(x)\|^{1-\theta-\eta})$, de modo que o problema considerado nesta proposição será

$$\begin{aligned} &\underset{p \,\in\, \Omega_x \,,\, \gamma}{\text{minimizar}} \;\; \gamma \\ &\text{sujeito a:} \;\; H(p,x) \in K \end{aligned} \tag{5.33}$$

onde definimos

$$H(p,x) \coloneqq \begin{bmatrix} \gamma\, \|F(x)\|^{\theta+\eta} - \|F(x) + G(x)\, p\| \\ \gamma\, \|F(x)\|^{\theta} - \|p\| \\ \gamma \end{bmatrix}_{1}^{1} \,,$$

$K \coloneqq [0,\ +\infty)^3 \subset \mathbb{R}^3$, um conjunto viável $\mathcal{F}(x) \coloneqq \{p \in \Omega_x \ :\ H(p,x) \in K\}$ e um conjunto de soluções $S(x) \coloneqq \underset{p \in \mathcal{F}(x)}{\arg\ \min}\ \gamma$ deste problema, avaliado em um ponto $x \in \Omega$.

Observação: Os índices presentes à direita e abaixo da matriz pela qual é definida $H(p,x)$ indicam o número de linhas e colunas, respectivamente, de cada submatriz. Embora possa ser considerada trivial neste caso particular, esta notação é utilizada neste capítulo com o objetivo de não confundir o leitor quanto às dimensões das matrizes apresentadas.

Seja $\bar{x} \in \Omega$. Temos que:

(i) A função γ é constante em relação a $(p,\ x)$, portanto contínua em $\mathbb{R}^n \times \Omega$;

(ii) A multifunção $\mathcal{F}(\cdot)$ é fechada, pois $H(p,x)$ é contínua e K é um conjunto convexo e fechado;

(iii) Considerando $\alpha \coloneqq \|F\|^{1-\theta-\eta}$ e o fato de que o conjunto Ω_x é compacto, temos que, para todo x suficientemente próximo de $\bar{x}$, o conjunto de nível:

$$\{p \in \mathcal{F}(x) \ :\ \gamma \leq \alpha\}$$

é não vazio, pela Proposição 5.2, e está contido em Ω_x, que é compacto. Agora, considerando que existe $r > 0$ tal que $x \in \mathcal{B}(\bar{x}, r)$, sabemos que $\Omega_x \subset \Omega_{\bar{x}} + \mathcal{B}(0,r)$, que também é um conjunto compacto;

(iv) Resta mostrar que, dado $b_1 > 0$, existirá $b_2 > 0$ tal que, para todo $x \in \mathcal{B}(\bar{x}, b_2) \cap \Omega$ teremos

$$\{S(\bar{x}) + \mathcal{B}(0, b_1)\} \cap \{p \in \Omega_x \ :\ H(p,x) \in K\} \neq \emptyset.$$

Para isto, basta mostrar que, para cada solução $(\bar{p}, \bar{\gamma})$ do Problema (5.31) com $x = \bar{x}$, temos que

$$\text{dist}(\ (\bar{p}, \bar{\gamma}),\ \mathcal{F}(x)\ ) \to 0.$$

quando $x \to \bar{x}$. Para todo $x \in \mathbb{R}^n$ suficientemente próximo de $\bar{x}$, sabemos que $F(x) \neq 0$, portanto podemos definir:

$$p := \bar{x} + \bar{p} - x, \quad \gamma := \max\left\{ \frac{\|F(\bar{x}) + G(\bar{x})\ p\|}{\|F(\bar{x})\|^{\eta+\theta}},\ \frac{\|p\|}{\|F(\bar{x})\|^{\theta}} \right\}.$$

Assim, $(p, \gamma) \in \mathcal{F}(x)$. Além disso, $p \to \bar{p}$ e

$$\gamma \to \bar{\gamma} := \max\left\{ \frac{\|F(\bar{x}) + G(\bar{x})\ \bar{p}\|}{\|F(\bar{x})\|^{\eta+\theta}},\ \frac{\|\bar{p}\|}{\|F(\bar{x})\|^{\theta}} \right\},$$

quando $x \to \bar{x}$. Portanto, $\gamma(\bar{x}) = \bar{\gamma}$, mostrando que a Hipótese (iv) desta proposição é satisfeita.

Assim, mostramos que todas as hipóteses de [3, Proposition 4.4] são satisfeitas. Portanto, segue que $\gamma(\cdot)$ é uma função contínua em $\bar{x}$ em relação a Ω. □

Para continuar a análise da convergência deste algoritmo, é necessário estabelecer mais condições sob as quais a matriz $G(\cdot)$ possa ser considerada uma substituição adequada para a matriz jacobiana de $F(\cdot)$. Uma propriedade relacionada à derivada generalizada de Clarke é a Clarke regularidade (ver Definição 2.41). Dizemos que a função $f(\cdot)$ é Clarke regular em um ponto $x \in \mathbb{R}^n$ se:

(i) Para todo $d \in \mathbb{R}^n$, a derivada direcional $f'(x; d)$ existe;

(ii) Para todo $d \in \mathbb{R}^n$, tivermos $f'(x; d) = f^o(x; d)$.

Assim, consideraremos a seguinte hipótese:

Hipótese 5.4. *Para cada $i = 1, \ldots, m$, a função $F_i(\cdot)$ é Lipschitz contínua e Clarke regular em Ω. Além disso, dado $x \in \mathbb{R}$, para cada $i = 1, \ldots, m$, a i-ésima linha de $G(x)$ é denotada por $G_i(x)$ e é tal que $G_i(x) \in \partial_C F_i(x)$, sendo $G_i(\cdot)$ uma função contínua, para cada $i = 1, \ldots, m$.*

A Hipótese 5.4 claramente é satisfeita quando $F_i(\cdot)$, $i = 1, \ldots, m$, são funções continuamente diferenciáveis e definimos $G(\cdot) := F'(\cdot)$.

Sob a Hipótese 5.4, podemos fazer algumas observações a respeito do subdiferencial de Clarke de $f(\cdot)$, denotado por $\partial_C f(\cdot)$. Notemos, primeiramente, que este conjunto está bem definido, pois, sendo $f(\cdot)$ o máximo de um número finito de funções Lipschitz contínuas, sabemos que $f(\cdot)$ também é Lipschitz contínua.

Definimos os conjuntos de índices:

$$\begin{aligned} I_+(x) &\coloneqq \{i \in \{1,\dots,m\} \mid F_i(x) = f(x)\} \\ I_-(x) &\coloneqq \{i \in \{1,\dots,m\} \mid -F_i(x) = f(x)\}. \end{aligned} \tag{5.34}$$

Como a função $F(\cdot)$ é continuamente diferenciável, as derivadas direcionais de $f(\cdot)$ estão definidas, para todo $x \in \Omega$ e para toda direção $d \in \mathbb{R}^n$. Além disso, fixado $\bar{x} \in \mathbb{R}$, para todo x suficientemente próximo de $\bar{x}$, pela definição de $f(\cdot)$ nota-se que suas derivadas generalizadas de Clarke apresentam a seguinte configuração:

$$f^o(x;d) = \max\left\{ \max_{i\in I_+(\bar{x})} F_i'(x)\, d, \max_{i\in I_-(\bar{x})} -F_i'(x)\, d \right\} \tag{5.35}$$

onde, para cada $i = 1,\dots,m$, $F_i'(\cdot)$ é a i-ésima linha da matriz jacobiana $F'(\cdot)$ (ver [11, p. 2015]). Supondo que a Hipótese 5.4 seja satisfeita, podemos afirmar que as funções $F_i(\cdot)$, $i = 1,\dots,m$, são Clarke regulares em uma vizinhança de $\bar{x}$. Pela regra do máximo (ver Proposição 2.42), $f(\cdot)$ será também Clarke regular em uma vizinhança de $\bar{x}$. Ainda mais, nesta situação, o subdiferencial de Clarke poderá ser expresso da seguinte forma:

$$\partial_C f(\bar{x}) = \operatorname{conv}\left(\{\partial_C F_i(\bar{x}) \ : \ i \in I_+(\bar{x})\} \cup \{-\partial_C F_i(\bar{x}) \ : \ i \in I_-(\bar{x})\}\right). \tag{5.36}$$

Da Clarke regularidade da função $f(\cdot)$, temos que $f^o(x;d) = f'(x;d)$, para cada $x \in \mathbb{R}^n$ e cada direção $d \in \mathbb{R}^n$. Pela Hipótese 5.4, então,

$$f'(x,d) = \max\left\{ \max_{i\ \in I_+(\bar{x})} G_i(x)\, d, \max_{i\ \in I_-(\bar{x})} -G_i(x)\, d \right\}. \tag{5.37}$$

O resultado a seguir relaciona a função $\triangle(\cdot)$ a uma medida de otimalidade, esclarecendo o motivo pelo qual utilizamos a condição $\triangle(x) = 0$ como critério de parada.

Lema 5.10. *Suponha que a Hipótese 5.4 seja satisfeita e seja $x_k \in \Omega$ e que $0 < \eta \leq 1$. Então, toda solução (p_k, γ_k) de (5.1) satisfaz*

$$\max\left\{ \max_{i\ \in I_+(\bar{x})} G_i(x_k)\, p_k, \max_{i\ \in I_-(\bar{x})} -G_i(x_k)\, p_k \right\} \leq -\triangle(x_k). \tag{5.38}$$

Além disso, se para algum $\bar{x} \in \Omega$ tivermos que

$$0 \notin \partial_C f(\bar{x}) + N_\Omega(\bar{x}), \tag{5.39}$$

então existirão $\delta > 0$ e $M > 0$ tais que, para todo $x_k \in \Omega$ suficientemente próximo de $\bar{x}$ e para toda solução (p_k, γ_k) de (5.1), teremos:

$$\triangle(x_k) \geq \delta \ \ e \ \ \|p_k\| \leq M. \tag{5.40}$$

Demonstração: Seja $i \in I_+(x_k)$. Então, temos que

$$\begin{aligned} f(x_k) + G_i(x_k)\, p_k &= F_i(x_k) + G_i(x_k)\, p_k \\ &\leq |F_i(x_k) + G_i(x_k)\, p_k| \\ &\leq \|F(x_k) + G(x_k)\, p_k\| \end{aligned}$$

E assim, pela primeira restrição do Subproblema (5.1), temos que:

$$f(x_k) + G_i(x_k)\, p_k \leq \gamma_k\, \|F(x_k)\|^{\theta+\eta} = \gamma_k\, f(x_k)^{\theta+\eta}.$$

Portanto, segue que

$$G_i(x_k)\, p_k \leq -f(x_k) + \gamma_k\, f(x_k)^{\theta+\eta} = -\triangle(x_k).$$

De modo análogo, podemos afirmar também que, para cada $i \in I_-(x_k)$, vale que

$$-G_i(x_k)\, p_k \leq -\triangle(x_k).$$

Portanto, segue que a desigualdade (5.38) é satisfeita.

Considere agora um ponto $\bar{x} \in \Omega$ tal que a condição (5.39) seja satisfeita. Claramente, $f(\bar{x}) \neq 0$, pois do contrário $\bar{x}$ seria uma solução global do Problema (5.26) e, pela regra de Fermat teríamos que $0 \in \partial_C f(\bar{x}) + N_\Omega(\bar{x})$, uma contradição. Seja $(\bar{p}, \bar{\gamma})$ uma solução de (5.31) com $x = \bar{x}$. Então, pelo mesmo raciocínio utilizado na Proposição 5.2, sabemos que

$$\bar{\gamma} \leq f(\bar{x})^{1-\theta-\eta}\,.$$

Vale a igualdade se, e somente se, $p = 0$ for um passo aceito para o Problema (5.31) com $x = \bar{x}$. Mostraremos que $\bar{\gamma} < f(\bar{x})^{1-\theta-\eta}$. Podemos escrever o Problema (5.31) com $x = \bar{x}$, deixando de lado por enquanto a restrição $\gamma \geq 0$, como o seguinte problema de programação linear:

$$\begin{aligned} &\max \;\; c^T\, z \\ &\text{sujeito a:} \;\; M\, z \leq\; v \end{aligned} \tag{5.41}$$

onde, observando que o conjunto Ω pode ser escrito conforme (5.28), consideramos as matrizes:

$$z = \underset{1}{\begin{bmatrix} p \\ \gamma \end{bmatrix}}_{1}^{n}\,,\; c = \underset{1}{\begin{bmatrix} 0 \\ -1 \end{bmatrix}}_{1}^{n}\,,\; M = \underset{n}{\begin{bmatrix} G(\bar{x}) & -f(\bar{x})^{\theta+\eta}\, 1_m \\ -G(\bar{x}) & -f(\bar{x})^{\theta+\eta}\, 1_m \\ I & -f(\bar{x})^{\theta}\, 1_n \\ -I & -f(\bar{x})^{\theta}\, 1_n \\ A & 0 \end{bmatrix}}_{1}^{\begin{matrix} m \\ m \\ n \\ n \\ s \end{matrix}}\,,\; v = \underset{1}{\begin{bmatrix} -F(\bar{x}) \\ F(\bar{x}) \\ 0 \\ 0 \\ b - A\,\bar{x} \end{bmatrix}}_{1}^{\begin{matrix} m \\ m \\ n \\ n \\ s \end{matrix}}\,.$$

Assim, podemos escrever o dual deste problema de programação linear como:

$$\begin{aligned} &\min \ v^T \, w \\ &\text{sujeito a: } M^T \, w = c \\ &\qquad\quad w \geq 0 \end{aligned} \tag{5.42}$$

Sabemos que $\bar{z} \coloneqq \begin{bmatrix} 0 \\ f(\bar{x})^{1-\theta-\eta} \end{bmatrix}$ é viável para o problema primal (5.41). Assim, observando as condições KKT para o Problema (5.41), sabemos que se esta for uma solução do subproblema paramétrico (5.31), então é ponto KKT para o problema primal (5.41). Isto por sua vez ocorre se, e somente se, forem satisfeitas as condições de folgas complementares para este problema, ou seja, se, e somente se, existir $w \in \mathbb{R}^{3m+n}$, $w \geq 0$, tal que

$$w^T (M \, \bar{z} - v) = 0 \, . \tag{5.43}$$

Escrevendo w como

$$w = \begin{bmatrix} w^{1+} \\ w^{1-} \\ w^{2+} \\ w^{2-} \\ w^3 \end{bmatrix} \begin{matrix} m \\ m \\ n \\ n \\ s \end{matrix} \, ,$$

$$1$$

podemos reescrever a condição (5.43) como:

$$\begin{bmatrix} (w^{1+})^T & (w^{1-})^T & (w^{2+})^T & (w^{2-})^T & (w^3)^T \end{bmatrix} \begin{bmatrix} -f(x) \, 1_m + F(x) \\ -f(x) \, 1_m - F(x) \\ -f(x)^{1-\eta} \, 1_n \\ -f(x)^{1-\eta} \, 1_n \\ b - Ax \end{bmatrix} = 0 \, .$$

Assim, definindo $\mathcal{A}(x) \coloneqq \{r \in \{1, \ldots, s\} \mid A_r \, x = b_r\}$, sendo A_r a r-ésima linha de A, sabemos que:

- $w_i^{1+} = 0$, se $i \notin I_+(x), \quad i = 1, \ldots, m$;
- $w_i^{1-} = 0$, se $i \notin I_-(x), \quad i = 1, \ldots, m$;
- $w^{2+} = 0$;
- $w^{2-} = 0$ e
- $w_r^3 = 0$ se $r \notin \mathcal{A}(x), \quad r = 1, \ldots, s$.

Substituindo estas informações no problema dual (5.42), obtemos que $\bar{z}$ é um ponto KKT para o problema primal (5.41) se, e somente se, existirem valores $w_i^{1+} \geq 0,\ i \in I_+(x)$, $w_i^{1-} \geq 0,\ i \in I_-(x)$ e $w_r^3 \geq 0,\ r \in \mathcal{A}(x)$ tais que:

$$\sum_{i \in I_+(x)} w_i^{1+}\, G_i(x) - \sum_{i \in I_-(x)} w_i^{1-}\, G_i(x) + \sum_{r \in \mathcal{A}(x)} w_r^3\, A_i^T = 0\ ,$$

$$f(x)^{\theta+\eta} \left\{ \sum_{i \in I_+(x)} w_i^{1+} + \sum_{i \in I_-(x)} w_i^{1-} \right\} = 1\ .$$

Por (5.29), temos que $N_\Omega(x) = \text{cone}\,\{A_r \mid r \in \mathcal{A}(x)\}$. E assim, considerando (5.35), vemos pela Hipótese 5.5 que $\partial_C f(x) \cap -N_\Omega(x) \neq \emptyset$, uma contradição. Portanto, temos que

$$\bar{\gamma} < f(\bar{x})^{1-\theta-\eta}$$

e portanto, também temos que $\triangle(\bar{x}) > 0$. Pela Proposição 5.9, sabemos que $\gamma(\cdot)$ é contínua em uma vizinhança de $\bar{x}$ e, portanto, a função $\triangle(\cdot)$ também é contínua em uma vizinhança de $\bar{x}$. A continuidade de $\triangle(\cdot)$ em $\bar{x}$ e o fato de que $\triangle(\bar{x}) > 0$ implicam que existe $\delta > 0$ tal que, para todo $x_k \in \Omega$ suficientemente próximo de $\bar{x}$ teremos que $\triangle(x_k) \geq \delta$.

Além disso, para toda solução (p_k, γ_k) do Problema (5.1), pela segunda restrição deste problema temos que $\|p_k\| \leq \gamma_k\ f(x_k)^\theta = \gamma(x_k)\ f(x_k)^\theta \leq f(x_k)^{1-\eta}$. Sabendo que $\lim_{x_k \to \bar{x}} \gamma(x_k) = \bar{\gamma}$ e que $F(\cdot)$ é contínua, temos então que existe $\tilde{M} > 0$ tal que $\|p_k\| \leq \tilde{M}$, para todo ponto $x_k \in \Omega$ suficientemente próximo de $\bar{x}$. □

Observação: Note que, de fato, dado $x \in \Omega$ a condição $\triangle(x) = 0$ é equivalente ao critério de C-estacionaridade $0 \in \partial_C f(x) + N_\Omega(x)$. O corolário a seguir nos mostra que este critério de parada também impede que o algoritmo gere uma sequência infinita tal que $x_{k+1} = x_k$ a partir de algum iterando $x_{\bar{k}}$, quando a sequência $\{\alpha_k\}$ tem uma cota inferior $\alpha^* > 0$.

Corolário 5.11. *Seja $F : \mathbb{R}^n \to \mathbb{R}$ uma função continuamente diferenciável, $\Omega \subseteq \mathbb{R}^n$ um conjunto poliédrico e $x_k \in \mathbb{R}^n$. Então, temos que $\triangle(x_k) = 0$ se, e somente se, $f(x_k) = 0$ ou $(p_k, \gamma_k) = (0,\ f(x_k)^{1-\theta-\eta})$ for solução do Problema (5.1).*

Demonstração: Primeiramente, mostraremos que, se $f(x_k) = 0$ ou se $(p_k,\ \gamma_k) = (0,\ f(x_k)^{1-\theta-\eta})$ for solução do Problema (5.1), então $\triangle(x_k) = 0$. Note que o caso no qual $f(x_k) = 0$ é trivial, logo provaremos apenas o segundo caso. Lembrando da Proposição 5.2, sabemos que:

$$\gamma_k = \max\left\{ \frac{\|F(x_k) + G(x_k)p_k\|}{\|F(x_k)\|^{\eta+\theta}},\ \frac{\|p_k\|}{\|F(x_k)\|^\theta} \right\}.$$

Logo, considerando $p_k = 0$, obtemos que

$$\gamma_k = \max\left\{\frac{\|F(x_k)\|}{\|F(x_k)\|^{\eta+\theta}},\ 0\right\} = f(x_k)^{1-\eta-\theta}.$$

Substituindo este valor em $\triangle(x_k)$, obtemos

$$\triangle(x_k) = f(x_k) - f(x_k) = 0.$$

Reciprocamente, suponhamos que $\triangle(x_k) = 0$. Então teremos que

$$f(x_k) = \gamma_k\ f(x_k)^{\eta+\theta}.$$

E isto implica que $f(x_k) = 0$ ou $\gamma_k = f(x_k)^{1-\eta-\theta}$. Esta última equação equivale a afirmar que $(0,\ f(x_k)^{1-\theta-\eta})$ é solução do Problema (5.1). □

Considerando que $F(\cdot)$ é uma função Lipschitz contínua, faremos mais uma suposição a respeito da matriz $G(\cdot)$.

Hipótese 5.5. *Para toda direção $d \in \mathbb{R}^n$, $x^* \in Z$, temos que*

$$f'(x^*; d) = \max\left\{\max_{i\ \in I_+(x^*)} G_i(x^*)\ d,\ \max_{i\ \in I_-(x^*)} -G_i(x^*)\ d\right\}. \tag{5.44}$$

Note que, sob a Hipótese 5.5, a hipótese de Clarke regularidade pode ser reformulada como a seguinte igualdade de conjuntos:

$$\partial_C f(x^*) = \text{conv}\ \left(\{G_i(x^*)\ :\ i\ \in I_+(x^*)\} \cap \{-G_i(x^*)\ :\ i\ \in I_-(x^*)\}\right). \tag{5.45}$$

Esta hipótese é claramente satisfeita quando, para todo $i = 1, \ldots, m$, $F_i(\cdot)$ é uma função contínua. O próximo lema mostra o comportamento de um ponto de acumulação de uma sequência $\{x_k\}$ gerada pelo Algoritmo 4.

Observação: Antes de apresentarmos os próximos resultados de convergência global do Algoritmo 4, note que, se o Algoritmo 4 para em um iterando x_k, então $\triangle(x_k) = 0$, pois as únicas condições de parada são $x_k \in Z$ e $\triangle(x_k) = 0$.

Teorema 5.12. *Dado um ponto inicial $x_0 \in \Omega$, suponha que $\{x_k\}$ seja uma sequência infinita gerada pelo Algoritmo 4. Então, para todo ponto de acumulação x^* da sequência $\{x_k\}$ no qual a Hipótese 5.5 seja satisfeita, vale o critério de C-estacionaridade, ou seja, $0 \in \partial_C f(x^*) + N_\Omega(x^*)$.*

Demonstração: Seja $\{x_k\}$ uma sequência infinita gerada pelo Algoritmo 4 com ponto inicial $x_0 \in \Omega$. Pela estrutura do algoritmo, sabemos que

$$f(x_{k+1}) \leq f(x_k) - \alpha_k \, \sigma \triangle (x_k) \, ,$$

com $x_{k+1} = x_k + \alpha_k \, p_k$. Sendo $\alpha_k \, \sigma \triangle (x_k) > 0$, a desigualdade anterior implica que $f(x_{k+1}) < f(x_k)$, para toda iteração k, ou seja, a sequência $\{f(x_k)\}$ é monótona decrescente e limitada, pois $f(x_k) \geq 0$, para todo x_k. Portanto, sabemos que a sequência $\{f(x_k)\}$ é convergente, ou seja, existe $y \geq 0$ tal que $\lim_{k\to\infty} f(x_k) = y$.

Disto, obtemos que $\lim_{k\to\infty} \alpha_k \, \sigma \triangle (x_k) = \lim_{k\to\infty} [f(x_{k+1}) - f(x_k) + \alpha_k \, \sigma \triangle (x_k)] \leq 0$. Sabendo que $\sigma > 0$ é um escalar, $\alpha_k > 0$, $\triangle(x_k) > 0$, para toda iteração k, temos:

$$\lim_{k\to\infty} \alpha_k \triangle (x_k) = 0. \tag{5.46}$$

Seja x^* um ponto de acumulação da sequência $\{x_k\}$ e uma subsequência $\left\{x_{k_j}\right\}$ tal que $x_{k_j} \to x^*$. Prosseguiremos com a análise dos dois possíveis casos.

(i) Caso a sequência $\left\{\alpha_{k_j}\right\}$ possua uma cota inferior $\alpha^* > 0$, teremos que $\lim_{k_j\to\infty} \alpha_{k_j} \geq \alpha^*$. Por (5.46), teremos então que $\lim_{k_j\to\infty} \triangle (x_{k_j}) = 0$. Portanto, pela continuidade da função $\triangle(\cdot)$, obtemos que $\triangle(x^*) = 0$, logo o critério de C-estacionaridade é satisfeito em x^*.

(ii) Caso a situação anterior não ocorra, existirá ao menos uma subsequência convergente $\left\{\alpha_{k_j}\right\}$, para a qual utilizamos os mesmos índices por simplicidade de notação, tal que $\lim_{k_j\to\infty} \alpha_{k_j} = 0$.
Caso $\triangle(x^*) = 0$, sabemos que o critério de C-estacionaridade será satisfeito em x^*. Suponhamos então que $\triangle(x^*) \neq 0$. Por (5.40), temos que a sequência $\left\{\|p_{k_j}\|\right\}$ possui uma cota superior $M > 0$. Portanto existe ao menos uma subsequência convergente $\left\{p_{k_j}\right\}$, para a qual novamente utilizamos os mesmos índices por simplicidade de notação, com $\lim_{k_j\to\infty} p_{k_j} = p^*$. Sendo α_{k_j} um passo aceito pelo Algoritmo 4 em um iterando x_{k_j}, sabemos que $\beta_{k_j} = \alpha_{k_j} \, \beta$ não é um passo aceito pelo algoritmo, e portanto

$$f(x_{k_j} + \beta_{k_j} \, p_{k_j}) > f(x_{k_j}) - \, \sigma \, \beta_{k_j} \triangle (x_{k_j}). \tag{5.47}$$

Isto equivale a afirmar que

$$\frac{f(x_{k_j} + \beta_{k_j} \, p_{k_j}) - f(x_{k_j})}{\beta_{k_j}} > - \, \sigma \triangle (x_{k_j}). \tag{5.48}$$

Tomando o limite em ambos os lados desta desigualdade, obtemos

$$- \, \sigma \triangle (x^*) \; = \; - \, \sigma \lim_{k_j\to\infty} \triangle (x_{k_j}) \quad \leq \; \limsup_{k_j\to\infty} \frac{f(x_{k_j} + \beta_{k_j} \, p_{k_j}) - f(x_{k_j})}{\beta_{k_j}} \, .$$

Observando que $p_{k_j} \to p^*$ e $\beta_{k_j} \to 0$, temos que

$$-\sigma \triangle(x^*) \leq \limsup_{k_j \to \infty} \frac{f(x_{k_j} + \beta_{k_j} p^*) - f(x_{k_j})}{\beta_{k_j}} \leq f^o(x^*, p^*) .$$

E assim, pela Clarke regularidade da função $f(\cdot)$, obtemos

$$-\sigma \triangle(x^*) \leq f'(x^*, p^*).$$

Pelas equações (5.38) e (5.37), então,

$$-\sigma \triangle(x^*) \leq f'(x^*, p^*) \leq -\triangle(x^*). \tag{5.49}$$

Sendo $\triangle(x^*) \neq 0$, isto implica que $\sigma \geq 1$, o que é uma contradição.

Portanto, $\triangle(x^*) = 0$, como queríamos provar. □

Agora, mostraremos que, quando $x_k \in \Omega$ está suficientemente próximo do conjunto Z, o passo $\alpha = 1$ é aceito, ou seja, o método global (Algoritmo 4) resume-se à estratégia local (Algoritmo 3).

Teorema 5.13. *Seja $x^* \in Z$. Suponha que a função $\gamma(\cdot)$, definida como solução ótima do Problema (5.31), seja limitada em Ω em uma vizinhança de x^*. Suponha também que a Hipótese 5.2 seja satisfeita, $\eta + \theta > 1$ e $\theta(1 + q_2) > 1$. Então, para todo iterando $x_k \in \Omega \setminus Z$ suficientemente próximo de x^*, teremos $\triangle(x_k) > 0$ e, para toda solução (p_k, γ_k) do Subproblema (5.1), teremos:*

$$f(x_k + p_k) \leq f(x_k) - \sigma \triangle(x_k) .$$

Demonstração: Primeiramente, mostraremos que $x_k + p_k \in \mathcal{B}(x^*, r) \cap \Omega$, se $x_k \in \Omega$ for suficientemente próximo de $x^* \in Z$. De fato isto ocorre, pois, pelas restrições do Subproblema (5.1) e pela desigualdade triangular, obtemos

$$\|x_k + p_k - x^*\| \leq \|x_k - x^*\| + \|p_k\| \leq \|x_k - x^*\| + \gamma_k \|F(x_k)\|^\theta ,$$

sendo γ_k o valor ótimo do Subproblema (5.1). Da continuidade de $f(\cdot)$ e devido ao fato de $\gamma(\cdot)$ ser uma função limitada, podemos escolher x_k suficientemente próximo de x^* de tal forma que $x_k + p_k \in \mathcal{B}(x^*, r)$.

Agora, pela Hipótese 5.2 e pelas restrições do Subproblema (5.1), obtemos pela desigualdade triangular o seguinte:

$$\begin{aligned} \|F(x_k + p_k)\| &\leq \|F(x_k + p_k) - F(x_k) - G(x_k) p_k\| + \|F(x_k) + G(x_k)\| \\ &\leq c_2 \|p_k\|^{1+q_2} + \gamma_k \|F(x_k)\|^{\eta+\theta} \\ &\leq c_2 \left[\gamma_k \|F(x_k)\|^\theta\right]^{1+q_2} + \gamma_k \|F(x_k)\|^{\eta+\theta} . \end{aligned}$$

Logo,

$$\|F(x_k + p_k)\| \;\leq\; c_2\;{\gamma_k}^{1+q_2}\;\|F(x_k)\|^{\theta\;(1+q_2)} + \gamma_k\;\|F(x_k)\|^{\eta+\theta} \tag{5.50}$$

Observando que $\gamma_k\;\|F(x_k)\|^{\eta+\theta} = \|F(x_k)\| - \triangle(x_k)$, obtemos por (5.50) que

$$\begin{aligned}\|F(x_k + p_k)\| \;\leq\; c_2\;\gamma_k^{1+q_2}\;\|F(x_k)\|^{\theta\;(1+q_2)} + \sigma\;[\|F(x_k)\| - \triangle(x_k)] + \\ + (1-\sigma)\;\;[\|F(x_k)\| - \triangle(x_k)]\;.\end{aligned}$$

Esta desigualdade pode ser reescrita como

$$\begin{aligned}&f(x_k + p_k) - f(x_k) + \sigma\;\;\triangle\,(x_k)\\ &\quad\leq\; -f(x_k)\cdot\Big[(1-\sigma) - \gamma_k\;f(x_k)^{\eta+\theta-1} - c_2\;{\gamma_k}^{1+q_2}\;f(x_k)^{\theta\;(1+q_2)-1}\Big]\;.\end{aligned} \tag{5.51}$$

Agora, mostraremos que esta última expressão não é positiva. Observe que, como $\eta+\theta > 1$, $\theta\;(1+q_2) > 1$ e $\gamma(\cdot)$ é limitada, podemos escolher x_k tão próximo de x^* que $\|F(x_k)\|$ seja arbitrariamente pequeno. Assim, podemos escolher x_k suficientemente próximo de x^* de modo que tenhamos

$$\gamma_k\;f(x_k)^{\eta+\theta-1} + c_2\;{\gamma_k}^{1+q_2}\;f(x_k)^{\theta\;(1+q_2)-1} \;\leq\; (1-\sigma)\;.$$

Isto implica que

$$-f(x_k)\cdot\Big[(1-\sigma) - \gamma_k\;f(x_k)^{\eta+\theta-1} - c_2\;{\gamma_k}^{1+q_2}\;f(x_k)^{\theta\;(1+q_2)-1}\Big] \;\leq\; 0\;.$$

Assim, por (5.51) obtemos que

$$f(x_k + p_k) \;\leq\; f(x_k) - \sigma\;\;\triangle\,(x_k)$$

Finalmente, temos que $\triangle(x_k) > 0$, pois

$$\triangle(x_k) \;=\; f(x_k)\left(1 - \gamma_k\;f(x_k)^{\eta+\theta-1}\right)$$

e podemos escolher $x_k \in \Omega \setminus Z$ suficientemente próximo de $x^* \in Z$ de tal forma que $\gamma_k\;f(x_k)^{\eta+\theta-1}$ seja arbitrariamente pequeno. □

Para este teorema, supusemos que $\gamma(\cdot)$, definida como solução ótima do Problema (5.31), fosse limitada em Ω em uma vizinhança de $x^* \in Z$. O lema abaixo fornece condições sob as quais isto ocorre.

Lema 5.14. *Seja $x^* \in Z$. Suponha que as Hipóteses 5.1 e 5.2 sejam satisfeitas e que $\eta+\theta \leq q_1\;(1+q_2)$, $\theta \leq q_1$. Então, a função $\gamma(\cdot)$ é limitada em Ω, em uma vizinhança de x^*.*

Demonstração: Seja $x \in \mathcal{B}(x^*, r/2)$, sendo $0 < r < \varepsilon$ e r suficientemente pequeno, de modo que $\|F(x)\| < 1$ em $\mathcal{B}(x^*, r)$. Denotemos por $\bar{x}$ um ponto em Z tal que $\|x - \bar{x}\| = \text{dist}(x, Z)$. Como

$$\|\bar{x} - x^*\| \leq \|x - \bar{x}\| + \|x - x^*\| \leq \text{dist}(x, Z) + \frac{r}{2} \leq r,$$

vemos que $\bar{x} \in \mathcal{B}(x^*, r)$. Pela Hipótese 5.2, temos que

$$\begin{aligned} \|F(x) + G(x)\,(\bar{x} - x)\| \;&\leq\; c_2\,\|x - \bar{x}\|^{1+q_2} \;=\; c_2\,\text{dist}(x, Z)^{1+q_2} \\ &\leq\; {c_1}^{-1-q_2}\,c_2\,\|F(x)\|^{(1+q_2)\,q_1} \end{aligned} \tag{5.52}$$

sendo que nesta última desigualdade utilizamos a Hipótese 5.1. Agora, observando que $\eta + \theta \;\leq\; q_1\,(1 + q_2)$ e $\|F(x_k)\| \;\leq\; 1$, temos

$$\|F(x) + G(x)\,(\bar{x} - x)\| \;\leq\; {c_1}^{-1-q_2}\,c_2\,\|F(x)\|^{(1+q_2)\,q_1} \;\leq\; {c_1}^{-1-q_2}\,c_2\,\|F(x)\|^{\eta+\theta}. \tag{5.53}$$

Além disso, sendo $\theta \leq q_1$, obtemos a seguinte desigualdade

$$\|x - \bar{x}\| \;=\; \text{dist}(x, Z) \;\leq\; {c_1}^{-1}\|F(x)\|^{q_1} \;\leq\; {c_1}^{-1}\|F(x)\|^{\theta}. \tag{5.54}$$

Denotando $\bar{p} \coloneqq \bar{x} - x$, $\bar{\gamma} \coloneqq \max\{{c_1}^{-1-q_2}\,c_2,\; {c_1}^{-1}\}$, vemos por (5.53) e (5.54) que:

$$\begin{aligned} \|F(x) + G(x)\,\bar{p}\| \;&\leq\; \bar{\gamma}\,\|F(x)\|^{\eta+\theta} \\ \|\bar{p}\| \;&\leq\; \bar{\gamma}\,\|F(x)\|^{\theta} \end{aligned}$$

Sendo Ω um conjunto convexo, temos então que $(\bar{p}, \bar{\gamma})$ é uma solução viável para o Problema (5.31). Como $x \in \mathcal{B}(x^*, r/2) \cap \Omega$ é arbitrário e $\gamma(x)$ é o valor ótimo de (5.31), vemos que a função $\gamma(\cdot)$ é limitada em $\mathcal{B}(x^*, r/2) \cap \Omega$, logo, $\gamma(\cdot)$ é limitada em uma vizinhança de x^*. □

Considerando o Teorema 5.13 e o Lema 5.14, podemos estimar intervalos para os parâmetros η e θ que satisfazem, simultaneamente, as hipóteses destes dois resultados. Notemos, primeiramente, que, conhecidas as constantes q_1 e q_2, relacionadas às Hipóteses 5.1 e 5.2, tais parâmetros η e θ serão determinados pelas seguintes desigualdades:

$$\begin{cases} \dfrac{1}{1+q_2} \;<\; \theta \;\leq\; q_1 \\ 1 \;<\; \eta + \theta \;\leq\; q_1\,(1+q_2) \end{cases} \tag{5.55}$$

Graficamente, queremos que a interseção da região definida por $\dfrac{1}{1+q_2} \;<\; \theta \;\leq\; q_1$ com a região definida por $1 \;<\; \eta + \theta \;\leq\; q_1\,(1+q_2)$ seja não vazia. Isto é garantido quando os pares (η, θ) satisfazem

$$\begin{cases} \dfrac{1}{1+q_2} \;<\; \theta \;\leq\; q_1 \\ 1 - \dfrac{1}{1+q_2} \;<\; \eta \;\leq\; q_1\, q_2 \end{cases} \tag{5.56}$$

Portanto, também é necessário que

$$q_1 > \max\left\{ \frac{1}{1+q_2}\,,\; \frac{1}{q_2} - \frac{1}{q_2\,(1+q_2)} \right\} \;=\; \max\left\{ \frac{1}{1+q_2}\,,\; \frac{1}{1+q_2} \right\} \;=\; \frac{1}{1+q_2}\,.$$

Esta inequação limita os valores de q_1 e q_2 e nota-se que a região viável para estas constantes, considerando-se as hipóteses do Teorema 5.13 e do Lema 5.14, é igual à região determinada pelas hipóteses do Teorema 5.5 para estas mesmas constantes, conforme a Figura 5.2. As regiões viáveis para os parâmetros η e θ, conforme (5.56), são ilustradas pelas figuras 5.6 e 5.7.

Apresentamos, assim, as teorias de convergência local e global deste novo método. Em relação às aplicações para este método, observa-se que são as mesmas apresentadas em [1], visto que aqui foram consideradas hipóteses semelhantes às encontradas nesse trabalho, além dos casos em que F é continuamente diferenciável e as hipóteses do método LP-Newton são satisfeitas.

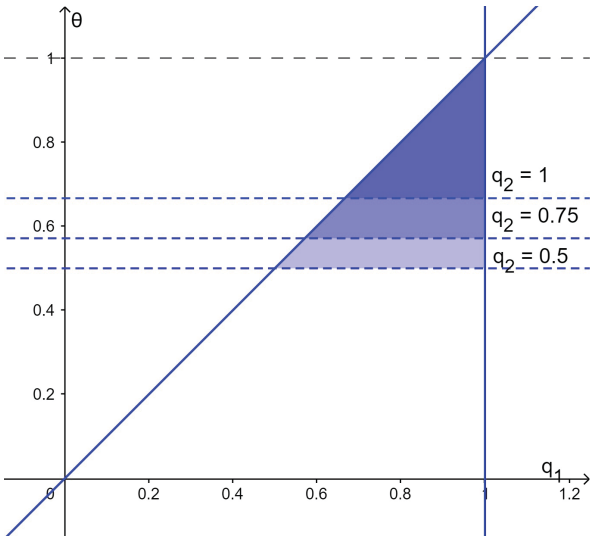


Figura 5.6: Região viável de θ para o Teorema 5.13 e o Lema 5.14. Consideramos as desigualdades: $\dfrac{1}{1+q_2} < \theta \leq q_1$. Para cada $q_2 \in (0,1]$ fixado, a região viável para os pares $(q_1,\ \theta)$ é a área sombreada acima de cada reta $\theta = \dfrac{1}{1+q_2}$.

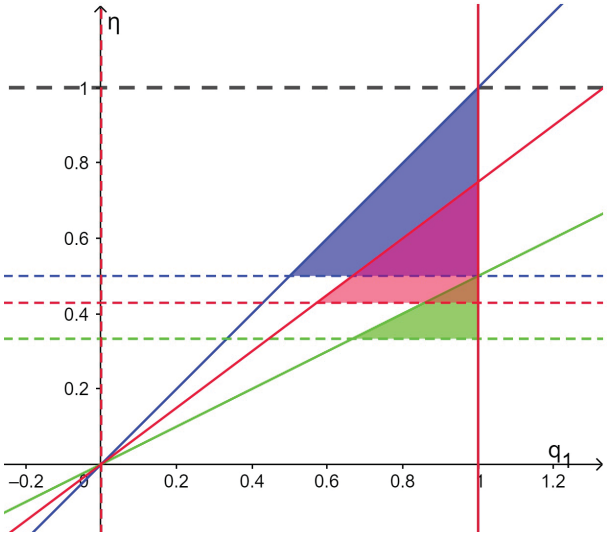


Figura 5.7: Região viável de η para o Teorema 5.13 e o Lema 5.14. Consideramos as desigualdades: $1 - \dfrac{1}{1+q_2} < \eta \leq q_1\, q_2$, as quais são mais restritivas do que $1 < \eta + \theta \leq q_1\,(1+q_2)$ com $\dfrac{1}{1+q_2} < \theta \leq q_1$. Neste gráfico são ilustradas três escolhas para $q_2 \in (0, 1]$: $q_2 = 1$ (azul), $q_2 = 0.75$ (rosa) e $q_2 = 0.5$ (verde). A região viável para os pares $(q_2,\ \eta)$ é o triângulo sombreado com as respectivas cores relacionadas a cada escolha de q_2.

6 Conclusão

Com os avanços da tecnologia, muitos problemas podem apresentar uma melhor modelagem por meio de funções que não sejam suaves ou que não satisfaçam certos critérios de regularidade. Surge, então, a necessidade de se apresentar métodos rápidos capazes de lidar com estas situações. Neste sentido, uma revisão bibliográfica sobre o método LP-Newton foi realizada durante a elaboração deste trabalho, pois este método apresenta convergência quadrática local e lida com critérios de regularidade diferentes dos clássicos, como os do método de Newton ou Levenberg-Marquardt. Aqui, estendemos ainda mais esta linha de pesquisa, de modo a contemplar também funções continuamente diferenciáveis que satisfaçam a condição de subregularidade métrica do tipo Hölder.

Como contribuições deste trabalho, temos:

- o algoritmo local do método LP-(η, θ);
- o algoritmo global do método LP-(η, θ);
- um estudo da convergência local deste método, pelo qual
  - determinam-se intervalos adequados para o parâmetro η;
  - apresentam-se condições suficientes para a convergência local superlinear deste método.
- um estudo da convergência global deste método, pelo qual
  - definem-se intervalos adequados para os parâmetros η e θ;
  - apresentam-se condições suficientes, sob as quais qualquer ponto de acumulação gerado por este método seja C-estacionário.

Para a versão local do método LP-(η, θ), a teoria de convergência aqui apresentada exibe parâmetros q_1 e q_2, relacionados às Hipóteses 5.1 e 5.2, sob os quais esta teoria é válida. Para estes parâmetros, a teoria de convergência local analisada neste

trabalho fornece um intervalo no qual devemos escolher o parâmetro $\eta > 0$ e a ordem de convergência local, que será superlinear.

A globalização do método é realizada, aqui, por meio de uma estratégia de busca linear. Para a versão global deste método, são fornecidas condições sob as quais este algoritmo gera um ponto C-estacionário para o problema de se minimizar a função $f(\cdot) \coloneqq \|F(\cdot)\|$. São também obtidas condições sobre as constantes q_1 e q_2, sob as quais esta teoria de convergência global seja válida.

Como possibilidades de trabalhos futuros, estão a implementação e realização de testes sobre o algoritmo global do método LP-(η, θ), bem como análise da convergência local e global deste método por meio de outros limitantes ou estimativas que, possivelmente, dependam mais do parâmetro $\theta > 0$. Também observa-se a possibilidade de se globalizar o algoritmo local do método LP-(η, θ) a partir de estratégias distintas da aqui utilizada. Embora ainda haja muito a ser desenvolvido, no sentido de abordar situações de minimização que não as clássicas, este novo método é mais um passo nesta direção.

Referências

[1] Ahookhosh, M.; Artacho, F. A.; Fleming, R. M. T. e Voung, P. T. Local convergence of the Levenberg-Marquardt method under Holder metric subregularity. *Adv Comput Math (2019)*, 2019. DOI: 10.1007/s10444-019-09708-7.

[2] Benko, M. e Gfrerer, H. New verifiable stationarity concepts for a class of mathematical programs with disjunctive constraints. *Optimization, Vol. 67, p. 1-23*, 2017. DOI: 10.1080/02331934.2017.1387547.

[3] Bonnans, J. F. e Shapiro, A. *Pertubation Analysis of Optimization Problems*. Springer, Berlin, 2000. DOI: 10.1007/978-1-4612-1394-9.

[4] Borwein, J. M. e Zhu, Q. J. *Techniques of Variational Analysis*. Springer, New York, 2005. DOI: 10.1007/0-387-28271-8.

[5] Clarke, F. H. *Optimization and Nonsmooth Analysis*. Classics in Applied Mathematics, SIAM, Philadelphia, 1990. DOI: 10.1137/1.9781611971309.

[6] Fabian, M. J.; Henrion, R.; Kruger, A. Y. e Outrata, J. V. Error bounds: Necessary and sufficient conditions. *Set-Valued and Variational Analysis, Vol. 18, p. 121-149*, 2010. DOI: 10.1007/s11228-010-0133-0.

[7] S. Facchinei, F. ; Lucidi. Convergence to second-order stationary points in inequality constrained optimization. *Mathematics of Operations Research*, 23:746–766, 1998.

[8] Facchinei, F. ; Fischer, A. e Herrich, M. An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. *Mathematical Programming, Vol. 146, p. 1-36*, 2014. DOI: 10.1007/s10107-013-0676-6.

[9] Facchinei, F. e Pang, J. *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume II*. Springer, New York, 2003.

[10] Fischer, A.; Herrich, M.; Izmailov, A. F. e Solodov, M. V. A globally convergent LP-Newton method. *SIAM Journal on Optimization*, 26:2012–2033, 2016. DOI: 10.1137/15M105241X.

[11] Fischer, A.; Herrich, M.; Izmailov, A. F.; Scheck, W. e Solodov, M. V. A globally convergent LP-Newton method for piecewise smooth constrained equations: escaping nonstationary accumulation points. *Computational Optimization and Applications*, 69:325–349, 2018. DOI: 10.1007/s10589-017-9950-5.

[12] Fukushima, M. e Yamashita, N. *On the rate of convergence of the Levenberg-Marquardt method*. G. Alefeld, X. Chen (eds): Topics in Numerical Analysis, Vol. 15, Springer Vienna, 2001. DOI: 10.1007/978-3-7091-6217-0_18.

[13] Heemels, W. P. M. H.; Schumacher, J. M. e Weiland, S. Applications of complementarity systems. *European Conference Control (ECC)*, 1999. DOI: 10.23919/ECC.1999.7099808.

[14] Herrich, M. Local convergence of Newton-type methods for nonsmooth constrained equations and applications. *Dr. Dissertation, Technischen Universität Dresden*, 2014.

[15] Nocedal, J. e Wright, S. J. *Numerical optimization*. Springer, New York, 2. ed., 2006. DOI: 10.1007/978-0-387-40065-5.

[16] Pták, V. What should be a rate of convergence? *RAIRO Numerical Analysis, Vol. 11 , p. 279-286*, 1977.

[17] Qi, L. e Sun, D. A Survey of Some Nonsmooth Equations and Smoothing Newton Methods. *Springer, Boston*, 1998. DOI: 10.1007/978-1-4613-3285-5_7.

[18] Rockafellar, R. T. e Wets, R. . *Variational Analysis*. Springer, Berlin, Vol. 317, 2009. DOI: 10.1007/978-3-642-02431-3.

[19] Schirotzek, W. . *Nonsmooth analysis*. Springer, Berlin, 2007. DOI: 10.1007/978-3-540-71333-3.

[20] Sun, W. e Yuan, Y. *Optimization Theory and Methods: Nonlinear Programming*. Springer, Boston, 2006. DOI: 10.1007/b106451.