

MATHEUS DE OLIVEIRA MENDONÇA

# Aperfeiçoamentos em métodos de otimização sem derivadas

- i - Otimização unidimensional multimodal
- ii - Método Nelder-Mead simplex restrito

*Orientador:*

Dr. Rodney Rezende Saldanha

*Coorientador:*

Dr. Douglas Alexandre Gomes Vieira

Belo Horizonte

6 de dezembro de 2019

**Universidade Federal de Minas Gerais**

**Escola de Engenharia**

**Programa de Pós-Graduação em Engenharia Elétrica**

**APERFEIÇOAMENTOS EM MÉTODOS DE OTIMIZAÇÃO SEM  
DERIVADAS**

Matheus de Oliveira Mendonça

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do Título de Mestre em Engenharia Elétrica.

Orientador: Prof. Rodney Rezende Saldanha

Belo Horizonte - MG

Dezembro de 2019

M539a	<p>Mendonça, Matheus de Oliveira.  Aperfeiçoamentos em métodos de otimização sem derivadas  [recurso eletrônico] : i - otimização unidimensional multimodal; ii - método  Nelder-Mead simplex restrito / Matheus de Oliveira Mendonça. - 2019.  1 recurso online (58 f. : il., color.) : pdf.</p>
	<p>Orientador: Rodney Rezende Saldanha.  Coorientador: Douglas Alexandre Gomes Vieira.</p>
	<p>Dissertação (mestrado) - Universidade Federal de Minas Gerais,  Escola de Engenharia.</p>
	<p>Apêndices: f. 52-55.</p>
	<p>Bibliografia: f. 56-58.</p>
	<p>Exigências do sistema: Adobe Acrobat Reader.</p>
	<p>1. Engenharia elétrica - Teses. 2. Otimização - Teses. I. Saldanha, Rodney  Rezende. II. Vieira, Douglas Alexandre Gomes. III. Universidade Federal  de Minas Gerais. Escola de Engenharia. IV. Título.</p>
621.3(043)	CDU:

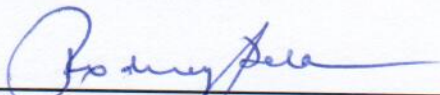
**"Aperfeiçoamentos em Métodos de Otimização sem Derivadas"**

**Matheus de Oliveira Mendonça**

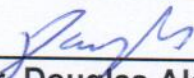
Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Mestre em Engenharia Elétrica.

Aprovada em 06 de dezembro de 2019.

Por:



Prof. Dr. Rodney Rezende Saldanha  
DEE (UFMG) - Orientador



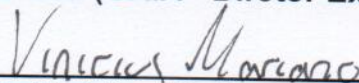
Prof. Dr. Douglas Alexandre Gomes Vieira  
PG-CEFETMG (ENACOM - Diretor Executivo) - Coorientador



Prof. Dr. Carlos Andrey Maia  
DEE (UFMG)



Prof. Dr. Adriano Chaves Lisboa  
PG-CEFETMG (GAIA - Diretor Executivo)



Prof. Dr. Vinícius Mariano Gonçalves  
DEE (UFMG)

# Agradecimentos

Em primeiro lugar, agradeço à minha família pelo amor incondicional, o zelo e a criação que são a base da pessoa que eu sou hoje. Agradeço aos meus pais pelos esforços imensuráveis para me darem uma educação de qualidade que foram fundamentais para que eu conseguisse chegar ao ensino superior.

Agradeço à ENACOM por ser uma empresa que preza pela ciência e por ter criado um terreno propício para que eu conseguisse realizar todas as atividades acadêmicas. Em especial, agradeço ao Gustavo pela confiança e pelos preciosos conselhos que proporcionaram a realização deste mestrado em conjunto com o trabalho.

Agradeço ao Douglas por ter me ajudado a trilhar este caminho da maneira mais natural possível, pelos diversos ensinamentos técnicos e não técnicos, pela sugestão do tema, pela confiança, e por sua “visão” diferenciada que me ajuda a ser um profissional cada vez mais completo.

Agradeço ao Prof. Rodney por essa oportunidade de orientação, por seu entusiasmo único ao transmitir o seu vasto conhecimento, tornando o aprendizado mais natural e me fazendo enxergar os problemas mais difíceis com outros olhos.

Aos meus amigos do “dê um rolê”, agradeço pelo companheirismo e pelas boas risadas, que fazem a minha vida ficar mais leve e mais colorida.

Agradeço à minha namorada Letícia, que me apoia incondicionalmente e me ajuda a ser uma pessoa cada dia melhor, além de ser a revisora mais criteriosa de todos os meus trabalhos.

# Resumo

Este trabalho propõe melhorias para alguns métodos de otimização determinística sem derivadas, em especial, para métodos de otimização unidimensional e para o método Nelder-Mead simplex. Em relação à primeira contribuição, neste trabalho é proposto um *framework* de otimização unidimensional baseado nos conceitos de padrões- $(V, \wedge)$ , que por sua vez são usados para a caracterização de multimodalidade de funções não lineares. Baseado nisso, o *framework* é capaz de mapear múltiplos mínimos ao longo do processo iterativo através de recortes sucessivos do intervalo de busca sempre que um máximo local é encontrado. O *framework* proposto é acoplado com o método da seção áurea, dando origem a um novo método de otimização unidimensional denominado *multimodal golden section*, que herda as propriedades de convergência do método subjacente. Alguns experimentos numéricos conduzidos ilustram a característica multimodal do *framework* desenvolvido. Já em relação à segunda contribuição, este trabalho propõe o uso de um operador lexicográfico como forma de tratamento de restrições de caixa e de desigualdade para o clássico método Nelder-Mead simplex. Além disso, uma nova estratégia de inicialização do simplex inicial é proposta, de modo a evitar uma degeneração prematura. As modificações propostas não alteram a estrutura original do algoritmo. Alguns experimentos são conduzidos e os resultados são comparados com estratégias tradicionais de inicialização do simplex e de tratamento de restrições, demonstrando as principais características dessa contribuição.

Palavras-chave: Otimização sem derivadas. Método da seção áurea. Funções multimodais. Nelder-Mead simplex. Comparação lexicográfica.

# Abstract

This work proposes improvements for some deterministic derivative-free optimization methods, in particular for line search optimization methods and the Nelder-Mead simplex method. Regarding the first contribution, this work proposes a line search optimization framework based on the  $(V, \wedge)$ -patterns, that are used for the multimodality characterization of nonlinear functions. Based on this, the framework can map multiple minima throughout successive interval breakdowns whenever a local maximum is detected. The proposed framework is coupled with the golden section method, originating a novel linear search optimization method called multimodal golden section, which inherits the convergence properties of the underlying method. Numerical experiments depict the multimodal feature of the framework. Regarding the second contribution, this paper proposes the use of a lexicographic operator to deal with box and inequality constraints for the classic Nelder-Mead simplex method. Also, a new initial simplex initialization strategy is proposed to prevent premature degeneration. The proposed modifications do not alter the original structure of the algorithm. Experiments are conducted and the results are compared with traditional simplex initializations and constraint handling strategies, demonstrating the main characteristics of the contribution.

Keywords: Derivative-free optimization. Golden section method. Multimodal functions. Nelder-Mead simplex. Lexicographic comparison.

# Sumário

Agradecimentos	2
Resumo	3
Abstract	4
Sumário	5
Lista de Figuras	7
Lista de Tabelas	9
Simbologia	10
Abreviaturas	12
<b>1 Introdução</b>	<b>13</b>
1.1 Motivação . . . . .	13
1.2 Objetivos . . . . .	14
1.3 Principais contribuições . . . . .	14
1.4 Organização do texto . . . . .	15
<b>2 Revisão teórica</b>	<b>16</b>
2.1 Introdução . . . . .	16
2.2 Nomenclatura . . . . .	16
2.3 Otimização não linear . . . . .	17
2.3.1 Métodos de direção de busca . . . . .	17
2.3.2 Métodos de busca direta . . . . .	19
2.4 Tratamento de restrições . . . . .	20
2.4.1 Métodos de penalidade interna (barreira) . . . . .	21
2.4.2 Métodos de penalidade externa . . . . .	21

2.4.3	Métodos dos multiplicadores de Lagrange . . . . .	22
2.4.4	Tratamento de restrições de caixa . . . . .	23
2.5	Conclusão do capítulo . . . . .	24
<b>3</b>	<b>Otimização unidimensional multimodal</b>	<b>26</b>
3.1	Introdução . . . . .	26
3.2	Caracterização de multimodalidade pelos padrões- $(V, \Lambda)$ . . . . .	26
3.3	Otimização unidimensional multimodal . . . . .	30
3.4	Exemplos . . . . .	33
3.4.1	Exemplo I . . . . .	33
3.4.2	Exemplo II . . . . .	36
3.5	Conclusão do capítulo . . . . .	38
<b>4</b>	<b>Nelder-Mead lexicográfico restrito</b>	<b>39</b>
4.1	Introdução . . . . .	39
4.2	Nelder-Mead lexicográfico restrito . . . . .	39
4.2.1	Comparação lexicográfica . . . . .	39
4.2.2	Inicialização do simplex . . . . .	42
4.3	Experimentos . . . . .	44
4.3.1	Experimento I . . . . .	44
4.3.2	Experimento II . . . . .	47
4.3.3	Experimento III . . . . .	48
4.4	Conclusão do capítulo . . . . .	50
<b>5</b>	<b>Conclusões</b>	<b>51</b>
<b>A</b>	<b>Algoritmos base de otimização</b>	<b>52</b>
A.1	Algoritmo da seção áurea . . . . .	52
A.2	Algoritmo Nelder-Mead simplex . . . . .	54
	<b>Referências Bibliográficas</b>	<b>56</b>

# Lista de Figuras

2.1	Taxonomia simplificada da otimização não linear: famílias de métodos de otimização.	17
2.2	O truncamento de um ponto que ultrapassa os limites da caixa pode mudar a direção de busca da atualização $x_{k+1} = x_k + \alpha_k d_k$ .	23
2.3	Aqui, $x_1$ é o melhor ponto, $x_{n+1}$ é o pior e $x_n$ o segundo pior ponto. O ponto $x_c$ é o centróide relativo a $x_{n+1}$ . Note que $x_r$ reside fora da caixa e o truncamento desse ponto gera um novo simplex com pontos colineares, degenerando o simplex.	24
3.1	Se dois padrões- $\wedge$ podem ser encontrados em uma sequência de pontos, deve existir um $\alpha_j$ tal que $f(\alpha_j) \leq \min\{f(\alpha_2), f(\alpha_i)\}$ , onde $\alpha_2$ e $\alpha_i$ são os pontos centrais dos dois padrões- $\wedge$ , caracterizando um padrão- $\vee$ entre eles.	28
3.2	Se um padrão- $\wedge$ está presente em uma sequência pontos, necessariamente existe um padrão- $\vee$ ou meio-padrão- $\vee$ entre o padrão- $\wedge$ e os limites do espaço de busca (i.e. $\underline{\alpha}$ ou $\bar{\alpha}$ ).	29
(a)	Presença de um meio-padrão- $\vee$ entre um padrão- $\wedge$ e $\bar{\alpha}$ onde $f(\alpha_i) \geq f(\alpha_{i+1}), \forall i \geq 2$ .	29
(b)	Presença de um padrão- $\vee$ entre um padrão- $\wedge$ e $\bar{\alpha}$ quando existe um $\alpha_i$ que satisfaça $f(\alpha_i) \leq f(\alpha_{i+1})$ .	29
3.3	Se um padrão- $\wedge$ está presente em uma sequência pontos, necessariamente existe um padrão- $\vee$ ou meio-padrão- $\vee$ entre o padrão- $\wedge$ e os limites do espaço de busca (i.e. $\underline{\alpha}$ ou $\bar{\alpha}$ ).	30
(a)	Presença de um meio-padrão- $\vee$ entre um padrão- $\wedge$ e $\underline{\alpha}$ quando $f(\alpha_i) \geq f(\alpha_{i-1}), \forall i \leq j$ .	30
(b)	Presença de um padrão- $\vee$ entre um padrão- $\wedge$ e $\underline{\alpha}$ quando existe um $\alpha_i$ que satisfaça $f(\alpha_i) \leq f(\alpha_{i-1})$ .	30
3.4	Ponto de divisão $\alpha_s$ para todos possíveis padrões- $\wedge$ em uma sequência com quatro pontos ordenados $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$ .	31
(a)	Divisão em $\alpha_3$ .	31
(b)	Divisão em $\alpha_2$ .	31
(c)	Divisão em $\alpha_2$ .	31

(d) Divisão em $\alpha_3$ . . . . .	31
3.5 A divisão do intervalo deve garantir a existência de um padrão-V, ou pelo menos um meio-padrão-V e, portanto, um mínimo local. Os pontos pretos representam a sequência $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ da execução atual e os pontos brancos os pontos armazenados de execuções anteriores. . . . .	31
(a) Expansão do intervalo até um padrão-V ser encontrado. . . . .	31
(b) Expansão do intervalo até um meio-padrão-V ser encontrado. . . . .	31
3.6 Instâncias da função de teste para $\omega = 10, 100$ e $\phi = 0$ . Note que independentemente dos valores de $\omega$ e $\phi$ , os menores valores função estão no sentido de $\alpha = 1$ . . . . .	34
3.7 Razão média da quantidade de mínimos locais mapeados ( $m$ ) em relação ao total de mínimos ( $\bar{m}$ ) dentro do espaço de busca. . . . .	35
3.8 Instância do problema para $\omega = 80$ , $\phi = \pi/2$ . Mínimos mapeados em negrito. . . . .	35
3.9 Instâncias de teste para $\eta = 0.5, 2.5$ e $\zeta = 0.25$ . Note que o mínimo global está próximo a $\zeta = 0.25$ e o número de mínimos locais é fixo. . . . .	36
3.10 Mínimos mapeados em negrito para a instância de $f_2(\alpha)$ onde $\zeta = 0.50$ e $\eta = 2.50$ . . . . .	37
3.11 Mínimos mapeados em negrito para a instância de $f_2(\alpha)$ onde $\eta = 0.50$ de $\zeta = 0.0$ . . . . .	37
4.1 Estratégias de inicialização do simplex. . . . .	43
(a) A inicialização proposta em [Luersen et al., 2004] adiciona somente componentes positivas a $x_0$ , direcionando a busca no sentido das componentes mais positivas. . . . .	43
(b) Inicialização proposta. A direção das perturbações são definidas de acordo com a posição relativa em relação aos pontos médios de cada dimensão. . . . .	43
4.2 Curvas de nível de $f_1(x)$ , onde $x^* = [0, 0]^T$ . . . . .	45
4.3 Instâncias do problema e as três regiões possíveis de inicialização de $x_0$ . . . . .	46
(a) Instância 1. . . . .	46
(b) Instância 2. . . . .	46
4.4 Região factível, em cinza, para diferentes valores de $c_3$ . Observe que $c_3$ pode ser visto como uma porcentagem da região de busca que é factível. . . . .	49
(a) Instância do problema com $c_3 = 10$ . . . . .	49
(b) Instância do problema com $c_3 = 1$ . . . . .	49

# Lista de Tabelas

3.1	Valores ótimos de $f_1(\alpha^*)$ obtidos pelos algoritmos EGS e MGS. Os resultados estão dispostos no formato (EGS, MGS), com o menor valor obtido em negrito. . . . .	34
3.2	Razão dos mínimos mapeados ( $m/\bar{m}$ ) do algoritmo MGS, onde 0.125 é o menor valor possível, isto é, o número total de mínimos locais é $\bar{m} = 8$ . . . . .	37
4.1	Valores de $\ x^*\ $ para as duas instâncias do problema, três regiões e as três formas de inicialização do simplex, considerando uma precisão numérica de $10^{-9}$ . . . . .	47
4.2	Resultados do experimento conduzido para uma precisão numérica de $10^{-9}$ . . . . .	48
4.3	Porcentagem das vezes que o método convergiu para o ótimo e para a região factível do problema para os métodos NMR e LA. . . . .	49

# Simbologia

$\mathbb{A}^*$	conjunto com todos os mínimos locais encontrados dentro de um intervalo de busca
$\mathbb{B}$	conjunto viável em relação aos limites inferiores e superiores de um problema de otimização
$d$	uma direção de busca no $\mathbb{R}^n$
$f(\cdot)$	função objetivo $f : \mathbb{R}^n \rightarrow \mathbb{R}$
$F(\cdot)$	função objetivo penalizada $F : \mathbb{R}^n \rightarrow \mathbb{R}$
$g(\cdot)$	restrições de desigualdade $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$
$\mathcal{H}$	matriz Hessiana, pertencente ao $\mathbb{R}^{n \times n}$
$k$	índice da iteração de um algoritmo de otimização
$n$	dimensão de um problema de otimização
$p$	número de restrições de desigualdade em um problema de otimização
$q$	número de comparações lexicográficas
$S$	tamanho do simplex inicial
$\mathbb{S}_n$	conjunto de pontos que formam um simplex no espaço $n$ -dimensional
$u$	escalar aleatório amostrado de uma distribuição uniforme
$\mathcal{U}(0, 1)$	distribuição uniforme no intervalo $[0, 1]$
$x$	variável de otimização pertencente ao $\mathbb{R}^n$
$x^i$	escalar representando o valor da posição $i$ de $x$
$x_e$	ponto de expansão do método Nelder-Mead simplex
$x_{ic}$	ponto de contração interna do método Nelder-Mead simplex

$x_k$	um ponto $k$ pertencente ao $\mathbb{R}^n$
$x_{oc}$	ponto de contração externa do método Nelder-Mead simplex
$x_r$	ponto de reflexão do método Nelder-Mead simplex
$x^*$	solução ótima pertencente ao $\mathbb{R}^n$
$\bar{x}$	limites superiores da variável de otimização $x$
$\underline{x}$	limites inferiores da variável de otimização $x$
$z$	vetor associado a um ponto $x$ pertencente ao $\mathbb{R}^q$
$z_e$	vetor associado ao ponto $x_e$
$z_{ic}$	vetor associado ao ponto $x_{ic}$
$z_{oc}$	vetor associado ao ponto $x_{oc}$
$z_r$	vetor associado ao ponto $x_r$
$\alpha$	escalar que representa o tamanho do passo em uma direção minimizante
$\beta$	parâmetro de expansão do método Nelder-Mead simplex
$\gamma$	parâmetro de contração externa do método Nelder-Mead simplex
$\Gamma(\cdot)$	função do tipo barreira
$\delta$	parâmetro de contração interna do método Nelder-Mead simplex
$\Delta$	largura do intervalo de busca
$\epsilon$	precisão numérica de um algoritmo de otimização
$\lambda$	multiplicador pertencente ao $\mathbb{R}^p$ da função de objetivo penalizada
$\mu$	multiplicador pertencente ao $\mathbb{R}^p$ da função de objetivo penalizada
$\xi$	uma direção coordenada no $\mathbb{R}^n$
$\rho$	parâmetro de reflexão do método Nelder-Mead simplex
$\Phi$	razão áurea, i.e. $(\sqrt{5} - 1)/2$
$\nabla f(\cdot)$	gradiente de $f(\cdot)$ , pertencente ao $\mathbb{R}^n$

# Abreviaturas

EGS	Enhanced golden section
GPS	Generalized pattern search
GSS	Generating set search
LA	Lagrangiano aumentado
MADS	Mesh adaptive search
MGS	Multimodal golden section
MLS	Multimodal line search
NM	Nelder-Mead simplex
NMR	Nelder-Mead simplex restrito

# Capítulo 1

## Introdução

### 1.1 Motivação

A otimização pode ser definida como o processo pelo qual se busca determinar as melhores configurações possíveis para a construção ou funcionamento de sistemas de interesse do ser humano [Takahashi, 2007a]. Por sua natureza prática, ao longo dos anos, pesquisadores, cientistas, engenheiros e diversos outros profissionais foram responsáveis pela elaboração de um arcabouço teórico muito sólido, mas que ao mesmo tempo está em constante evolução, principalmente pelas especificidades de cada problema prático que impõe a necessidade constante de melhorias de métodos já existentes e, muitas vezes, a criação de novos.

Dentre as diversas famílias de algoritmos determinísticos existentes, existe uma classe de algoritmos que não faz uso de informações do gradiente ao longo do processo de otimização. Esses algoritmos são usados em problemas no qual o cálculo do gradiente é custoso computacionalmente de ser obtido, ou não é confiável, ou é até mesmo inexistente, por exemplo [Larson et al., 2019]. Dentro dessa família de algoritmos, enquadram-se o algoritmo da seção áurea [Kiefer, 1953] e o Nelder-Mead simplex [Nelder and Mead, 1965]. O primeiro é usado na otimização unidimensional e o segundo na otimização de funções não lineares. Ao longo dos anos, diversos trabalhos foram publicados com estratégias para aperfeiçoamentos das versões originais desses algoritmos [Larson et al., 2019]. Vieira and Lisboa, 2014, por exemplo, sugerem uma alteração no algoritmo original da seção áurea para lidar com funções multimodais, garantindo que a solução obtida pelo algoritmo não seja pior que o ponto inicial do intervalo de busca. Luenberger and Ye, 2010, por sua vez, sugerem o uso de penalidades lineares como forma de tratamento de restrições no método Nelder-Mead simplex.

Assim, esta dissertação contribui para o aperfeiçoamento contínuo da otimização sem derivadas lidando com dois pontos específicos: multimodalidade da função objetivo na otimização unidimensional e tratamento de restrições no método Nelder-Mead simplex. Em relação ao primeiro ponto, uma estratégia geral para lidar com funções multimodais é desenvolvida, resultando em um

*framework* de otimização capaz de mapear múltiplos mínimos. Já em relação ao segundo ponto, um operador lexicográfico é proposto para tratamento de restrições, mostrando-se mais adequado ao método do Nelder-Mead simplex do que estratégias de penalização.

## 1.2 Objetivos

Os objetivos deste trabalho podem ser resumidos em dois pontos:

- introdução de um *framework* para otimização unidimensional que é capaz de mapear múltiplos mínimos de funções multimodais ao longo do processo iterativo. Esse *framework* possui garantias de convergência derivadas do método ao qual ele é acoplado. Baseado nele, um novo método de busca unidimensional, baseado no método da seção áurea, é proposto;
- apresentação de estratégias para tratamento de restrições e inicialização do algoritmo Nelder-Mead simplex, visando o tratamento de casos degenerados que possam ocorrer em virtude do uso de outras estratégias comuns.

## 1.3 Principais contribuições

Os frutos diretos e indiretos deste trabalho são 4 artigos aceitos em congressos científicos, 1 pôster aceito para apresentação em congresso e 1 artigo em revisão em revista. São eles:

i. Trabalhos completos:

- a. Mendonça, M. et al. (2019). Análise comparativa entre modelos de inteligência computacional para previsão do preço futuro no mercado de energia brasileiro. *XXV Seminário Nacional de Produção e Transmissão de Energia Elétrica* [Mendonça et al., 2019a];
- b. Mendonça, M. et al. (2019). Estratégia para tratamento de restrições no algoritmo Nelder-Mead com aplicações na otimização da programação diária da operação de usinas hidrelétricas. *XXVI Simpósio de Engenharia de Produção* [Mendonça et al., 2019b];
- c. Mendonça, M. et al. (2019). Heurística para unit commitment de turbinas hidráulicas distintas que visa maximizar a eficiência de geração de uma usina hidrelétrica e minimizar o ligamento/desligamento de máquinas considerando indisponibilidades programadas. *XXV Seminário Nacional de Produção e Transmissão de Energia Elétrica* [Mendonça et al., 2019c];
- d. Mendonça, M. et al. (2019). Programação diária da operação considerando incertezas de defluência de usinas a montante operadas por outras empresas. *XXV Seminário Nacional de Produção e Transmissão de Energia Elétrica* [Mendonça et al., 2019d];

## ii. Pôster:

- a. Mendonça, M. et al. (2019). *Constraints Handling and Simplex Initialization Strategy for the Nelder-Mead Simplex Method*. *LI Simpósio Brasileiro de Pesquisa Operacional*.

## iii. Artigo em revista:

- a. *A framework for finding multiple minima in deterministic line search optimization* (em revisão). *Journal of Computational and Applied Mathematics*.

## 1.4 Organização do texto

O presente texto está organizado da seguinte maneira: o [Capítulo 1](#) apresenta o objetivo e a motivação deste trabalho, bem como as principais contribuições científicas realizadas. O [Capítulo 2](#) traz o referencial teórico necessário que norteia esta dissertação, além de apresentar alguns trabalhos relacionados ao tema. O [Capítulo 3](#) e o [Capítulo 4](#) são os mais importantes, neles estão detalhadas as principais contribuições realizadas e os exemplos numéricos de cada contribuição. Por fim, o [Capítulo 5](#) realiza as conclusões desta dissertação e apresenta algumas possibilidades de trabalho futuro.

## Capítulo 2

# Revisão teórica

### 2.1 Introdução

Este capítulo possui a intenção de fornecer o referencial teórico necessário para o completo entendimento das contribuições realizadas. Dentre os conceitos discutidos, encontram-se os relacionados à otimização unidimensional e às abordagens de tratamento de restrições em métodos irrestritos. Aqui também estão inclusos os trabalhos relacionados.

### 2.2 Nomenclatura

Este texto define um problema de otimização não linear da seguinte maneira:

**Definição 1** (otimização restrita não linear).

$$\underset{x}{\text{minimize}} \quad f(x) \tag{2.1}$$

$$\text{sujeito a} \quad g(x) \leq 0 \tag{2.2}$$

$$x \in \mathbb{B}, \tag{2.3}$$

onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  é a função objetivo não linear do problema de otimização,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$  representam as  $p$  restrições de desigualdade,  $x \in \mathbb{R}^n$  representa a variável  $n$ -dimensional do problema de otimização e  $\mathbb{B} = \{x \in \mathbb{R}^n \mid \underline{x} \leq x \leq \bar{x}\}$ , no qual  $\underline{x} \in \mathbb{R}^n$  e  $\bar{x} \in \mathbb{R}^n$  são os limites inferiores e superiores de  $x$ , respectivamente, que também são denominados limites de caixa.

Caso o problema de otimização possua solução única, diz-se que o vetor  $x^*$  é igual ao argumento de  $f(\cdot)$  que faz com que essa função tenha seu mínimo valor, condicionado às restrições (2.2)-(2.3).

## 2.3 Otimização não linear

Dentro do universo da programação matemática, encontra-se um subconjunto de métodos e estratégias numéricas destinadas à otimização de funções não lineares. De maneira simplificada, esses métodos podem ser agrupados em quatro grandes famílias (ver [Figura 2.1](#)) [[Takahashi, 2007b](#)]:

- i. Métodos baseados em direção de busca;
- ii. Métodos baseados em exclusão de semi-espacos;
- iii. Métodos baseados em populações;
- iv. Métodos baseados em busca direta.

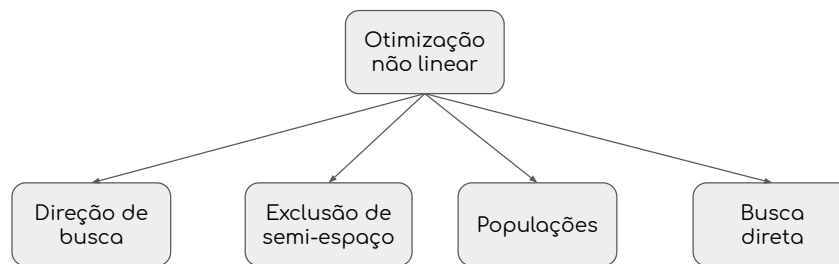


Figura 2.1: Taxonomia simplificada da otimização não linear: famílias de métodos de otimização.

Esta dissertação atém-se à discussão dos métodos baseados em direção de busca e em busca direta.

### 2.3.1 Métodos de direção de busca

Os métodos baseados em direção de busca assemelham-se pela busca local, na iteração  $k$ , por um novo ponto  $x_{k+1}$ , tal que

$$x_{k+1} = x_k + \alpha_k^* d_k,$$

onde  $\alpha_k^* \in [0, \bar{\alpha}_k]$  é um escalar que pondera o tamanho do passo no sentido da direção  $d_k$ , e  $\bar{\alpha}_k$  é o tamanho máximo desse passo, tal que  $x_{k+1} \in \mathbb{B}$ . Esses métodos diferem-se basicamente no cálculo de  $d_k$ . Para o método do gradiente [[Curry, 1944](#); [Courant et al., 1943](#)], por exemplo,

$$d_k = -\nabla f(x_k),$$

já para a família quasi-Newton [Fletcher and Powell, 1963; Broyden, 1965],

$$d_k = -\tilde{\mathcal{H}}_k^{-1} \nabla f(x_k),$$

onde  $\tilde{\mathcal{H}}_k^{-1} \in \mathbb{R}^{n \times n}$  é uma aproximação numérica de  $\mathcal{H}_k^{-1}$ , a inversa da matriz Hessiana avaliada em  $x_k$ . O Algoritmo 1 apresenta a estrutura básica de um método de direção de busca.

---

**Algoritmo 1** Métodos de direção de busca.

---

- 1:  $k \leftarrow 0$
  - 2: **while** critério de parada não for satisfeito **do**
  - 3:    $d_k \leftarrow$  calcula direção de busca
  - 4:    $\alpha_k^* \leftarrow \arg \min_{\alpha} f(x_k + \alpha d_k)$
  - 5:    $x_{k+1} \leftarrow x_k + \alpha_k^* d_k$
  - 6: **end while**
- 

Note que os métodos de direção de busca usam informações locais de primeira e/ou segunda ordem no processo de otimização. Portanto, a diferenciabilidade de  $f(\cdot)$  torna-se relevante para a execução desses métodos. Como discutido em [Takahashi, 2007b], a não diferenciabilidade da função objetivo pode resultar em descontinuidades na função gradiente, produzindo regiões “atratoras” da trajetória do algoritmo. Além disso, a não diferenciabilidade da função objetivo pode gerar vetores de gradiente linearmente dependentes, prejudicando a aproximação iterativa da matriz Hessiana utilizada em métodos quasi-Newton.

A otimização unidimensional [Bazarra and Shetty, 2006] surge como uma etapa natural de métodos baseados em direção de busca (ver linha 4 do Algoritmo 1):

**Definição 2** (otimização unidimensional). *O seguinte problema de otimização:*

$$\begin{aligned} & \underset{\alpha}{\text{minimize}} && f(\alpha) \\ & \text{sujeito a} && \underline{\alpha} \leq \alpha \leq \bar{\alpha}, \end{aligned} \tag{2.4}$$

onde  $\alpha \in \mathbb{R}$ ,  $\underline{\alpha}$  e  $\bar{\alpha}$  são os limites inferiores e superiores de  $\alpha$ ,  $f : \mathbb{R} \rightarrow \mathbb{R}$  é uma função não linear arbitrária, eventualmente multimodal, é conhecido como problema de otimização unidimensional.

Diversos métodos podem ser usados para resolver (2.4), incluindo métodos exatos e inexatos. Uma abordagem ingênua, por exemplo, é o uso de um passo fixo  $\alpha_k^*$ ,  $\forall k$ , que pode resultar em uma convergência muito lenta caso o passo seja pequeno, ou pode divergir, caso o passo seja muito grande, não possuindo, assim, uma convergência garantida [Bertsekas, 2008].

Reduções sucessivas do tamanho do passo [Armijo, 1966; Shi and Shen, 2005] e suas variações, que por sua vez adicionam algumas condições [Goldstein, 1965; Wolfe, 1969; Yuan et. al, 2010], são exemplos de outros métodos aplicados ao problema da otimização unidimensional.

O método da bisseção [Bertsekas, 2008; Luenberger and Ye, 2010; Bazarra and Shetty, 2006] remove metade do intervalo de busca a cada iteração, mas é dependente da diferenciabilidade

da função objetivo. Em contrapartida, o algoritmo de Brent, 1973, e o algoritmo da seção áurea [Kiefer, 1953; Avriel and Wilde, 1966] não dependem de avaliações de derivadas. Este último é baseado em comparação entre dois pontos que reduz o intervalo de busca a cada iteração, de forma que a convergência é atingida. Este método também já foi estendido para problemas multiobjetivo [Vieira et al., 2012].

Recentemente, Vieira and Lisboa, 2014, propuseram algumas modificações em métodos de otimização unidimensional, baseadas no conceito de padrão-V, que garantem que a solução obtida na otimização de funções multimodais não seja pior que a solução inicial. No Apêndice A, o Algoritmo 4 detalha as modificações inseridas no algoritmo da seção áurea. Esta dissertação estende o trabalho desses autores no Capítulo 3 e introduz o conceito de padrão- $\wedge$  (pronunciado como padrão-A), no sentido de criar uma estratégia geral de otimização unidimensional baseada em comparações entre pontos que é capaz de mapear múltiplos mínimos presentes no intervalo de busca de funções multimodais.

### 2.3.2 Métodos de busca direta

Os métodos de busca direta (*direct search/pattern search*, em inglês) pertencem a uma família de algoritmos de otimização que não fazem uso de informações de gradiente, ou de ordem superior, durante o processo de otimização de funções não lineares pertencentes ao  $\mathbb{R}^n$  [MathWorks, 2019b]. Essa família de algoritmos é usada em situações na qual a função objetivo não é continuamente diferenciável, ou até mesmo descontínua. Os métodos dessa família assemelham-se pela avaliação de um conjunto de pontos ao redor do ponto corrente  $x_k$ , com o intuito de se obter um novo ponto  $x_{k+1}$ , de forma que  $f(x_{k+1}) \leq f(x_k)$ .

Hooke and Jeeves, 1961, propuseram um clássico algoritmo de busca baseado em perturbações determinísticas em cada dimensão do ponto  $x_k$ , e a seguinte busca na direção formada pelas perturbações que resultaram nos menores valores de função. Basicamente, seja  $x_k$  o ponto da iteração atual e  $\xi_1, \dots, \xi_n$  as direções coordenadas. Este algoritmo testa os pontos  $x_k \pm \xi_i, \forall i \in [1, n]$ , fazendo um movimento na direção de melhora ou ficando no ponto atual. Após pesquisar todas as coordenadas, o algoritmo termina no ponto  $x_{k+1}$ , no qual é feita uma pesquisa na direção  $x_{k+1} - x_k$ .

Métodos similares também foram propostos, tais como Matyas, 1965, e Luus and Jaakola, 1973, mas com perturbações aleatórias, ao invés de perturbações nas direções coordenadas. Em [MathWorks, 2019b], são implementados os algoritmos *generalized pattern search* (GPS), *generating set search* (GSS) [Kolda et al., 2006] e o *mesh adaptive search* (MADS). O algoritmo GSS utiliza uma direção fixa de vetores para gerar os pontos na vizinhança de  $x_k$  e é similar ao algoritmo GPS, exceto em problemas de otimização que possuam restrições de igualdade. Já o algoritmo MADS seleciona os vetores de perturbação de forma aleatória.

Já o algoritmo proposto por Nelder and Mead, 1965, gera um conjunto  $\mathbb{S}_n = \{x_1, x_2, \dots, x_n, x_{n+1}\}$

com  $n + 1$  pontos para um problema  $n$ -dimensional. Os pontos são então ordenados de modo que

$$f(x_1) \leq f(x_2) \leq \dots \leq f(x_n) \leq f(x_{n+1}).$$

Seja

$$x_c = \frac{1}{n} \sum_{i=1}^n x_i$$

o centróide dos  $n$  melhores pontos (i.e. com o menor valor de função). O algoritmo tenta substituir o pior ponto  $x_{n+1}$  por um novo ponto obtido através de uma das seguintes operações:

- i. **Reflexão:**  $x_r = x_c + \rho(x_c - x_{n+1})$ ,  $\rho = 1.0$ ;
- ii. **Expansão:**  $x_e = x_c + \beta(x_r - x_c)$ ,  $\beta = 2.0$ ;
- iii. **Contração externa:**  $x_{oc} = x_c + \gamma(x_r - x_c)$ ,  $\gamma = 0.5$ ;
- iv. **Contração interna:**  $x_{ic} = x_c - \gamma(x_r - x_c)$ ,  $\gamma = 0.5$ ;
- v. **Encolhimento:**  $x_{i,k+1} = x_{1,k} + \delta(x_{i,k} - x_{1,k})$ ,  $\forall i \in [2, n + 1]$ .

A princípio, qualquer ponto  $(x_r, x_e, x_{oc}, x_{ic})$ , definidos pelas operações i-iv, são candidatos a substituir  $x_{n+1}$ , e os parâmetros comumente utilizados são  $(\rho, \beta, \gamma, \delta) = (1, 2, 0.5, 0.5)$ . A operação v altera todos os pontos, reduzindo o tamanho do simplex no sentido do melhor ponto. As condições para cada operação estão descritas em maiores detalhes no [Algoritmo 5](#), no [Apêndice A](#).

Uma revisão detalhada de métodos sem derivadas pode ser encontrada em [[Rios and Sahinidis, 2013](#)]. É importante ressaltar que o Nelder-Mead simplex é o método base da função “`fminsearch`”, do MATLAB [[MathWorks, 2019a](#)].

Esta dissertação propõe uma estratégia geral para tratamento de restrições no algoritmo Nelder-Mead (NM) baseada na criação de um operador de ordenação lexicográfica, fazendo com que o tratamento de restrições seja naturalmente incorporado pela mesma lógica do método original.

## 2.4 Tratamento de restrições

O tratamento de restrições não lineares genéricas é uma dificuldade antiga que desafia diversos métodos de otimização [[Takahashi, 2007b](#)]. Para as famílias de algoritmos baseadas em direção de busca e busca direta, por exemplo, os métodos são formulados essencialmente para problemas irrestritos. Com isso, diversas estratégias foram desenvolvidas ao longo dos anos para incorporar o tratamento de restrições nesses métodos. Esta seção faz um breve estudo de algumas dessas estratégias.

### 2.4.1 Métodos de penalidade interna (barreira)

A fórmula geral dos métodos de penalidade interna (barreira) é dada pela otimização do seguinte problema irrestrito:

$$\begin{aligned} & \underset{x, \lambda}{\text{minimize}} && F(x, \lambda) \\ & && x \in \mathbb{B}, \end{aligned} \tag{2.5}$$

onde

$$F(x, \lambda) = f(x) + \Gamma(x, \lambda), \tag{2.6}$$

em que  $\Gamma(x, \lambda) : \mathbb{R}^n \rightarrow \mathbb{R}$  é a função de penalidade e  $\lambda \in \mathbb{R}^p$  devem ser tais que [Takahashi, 2007b]:

- $\lim_{g(x) \rightarrow 0^-} \Gamma(x, \lambda) = +\infty$ ;
- $\Gamma(x, \lambda) \approx 0, \forall g(x) < -\epsilon$ ;
- $\lim_{\lambda \rightarrow 0^+} \Gamma(x, \lambda) = 0, \forall g(x) < -\epsilon$ .

Um exemplo de função barreira é:

$$\Gamma(x, \lambda) = -\lambda^T \frac{1}{g(x)}. \tag{2.7}$$

Note que a função (2.7) possui a propriedade de se aproximar do infinito para pontos factíveis próximos da fronteira da região factível definida pelo problema de otimização. Observe ainda que, caso o mínimo do problema esteja na fronteira da região factível, há uma tendência para a sequência de soluções do método, a cada de iteração, de aproximar-se desse mínimo sempre pelo lado de dentro da região factível [Takahashi, 2007b].

Nos métodos de penalidade interna, inicia-se a otimização com um valor relativamente elevado de  $\lambda$ , e faz-se  $\lambda \rightarrow 0^+$  ao longo do processo, diminuindo o seu valor a cada iteração. À medida que  $\lambda \rightarrow 0^+$ , a solução do problema penalizado converge para a solução do problema original [Takahashi, 2007b]. Outras maneiras de se determinar os multiplicadores  $\lambda$  são discutidas em [Luenberger et al., 1984].

### 2.4.2 Métodos de penalidade externa

Os métodos de penalidade externa, ou simplesmente penalidade, também alteram o problema original de otimização para a forma (2.5), sendo agora  $F(\cdot)$  uma função com as propriedades [Takahashi, 2007b]:

- $F(x, \lambda) \gg f(x), \forall g(x) > 0$ ;

- $(g(x_1) - g(x_2) > 0; g(x_2) > 0) \implies F(x_1, \lambda) > F(x_2, \lambda);$
- $F(x, \lambda) \approx f(x), \forall g(x) < \epsilon;$
- $\lim_{\lambda \rightarrow 0^+} F(x, \lambda) = f(x).$

Um exemplo de função penalidade é:

$$\Gamma(x, \lambda) = \lambda^T \max\{0, g(x)\}^2. \quad (2.8)$$

Para esse caso, a função (2.8) apresenta a propriedade de assumir valores elevados para pontos que violem fortemente as restrições, e valores nulos para pontos que não violem restrições. Observe ainda que a função de penalidade nesse caso não é continuamente diferenciável na maioria das vezes. Além disso, uma marcante característica desse método é que a solução é aproximada por pontos fora da região factível [Takahashi, 2007b].

De maneira geral, nos métodos de penalidade externa o vetor  $\lambda$  é inicializado com valores pequenos. Ao longo do processo iterativo, aumenta-se este valor de forma que  $\lambda \rightarrow \infty$  ao final da otimização. Claramente, existem outras maneiras de se determinar os valores de  $\lambda$  [Luenberger et al., 1984; Bertsekas, 2014], tais como:

- usar termos de penalidade distintos para cada restrição, evitando problemas de escala;
- penalizar apenas as restrições mais violadas;
- usar termos de penalidade adaptativos a cada iteração.

Note que, independente da função de penalidade escolhida, seja interna ou externa, a solução converge para a solução ótima quando  $\lambda_k \rightarrow \infty$  (ou  $\lambda_k \rightarrow 0^+$ , para a penalidade interna), entretanto, isso pode acarretar em um mal-condicionamento numérico do problema, visto que para alguns problemas de direção de busca, o cálculo da matriz Hessiana depende de  $\lambda_k$  e a variação de  $\lambda_k$  deve ser gradual, por exemplo. Com isso, a precisão numérica fica bastante limitada.

### 2.4.3 Métodos dos multiplicadores de Lagrange

O método dos multiplicadores de Lagrange, ou Lagrangiano Aumentado [Hestenes, 1969; Powell, 1969; Rockafellar, 1973] também é um método de penalidade tal que

$$\Gamma(x, \mu, \lambda) = \mu^T g(x) + \lambda^T \max\{0, g(x)\}^2, \quad (2.9)$$

onde  $\mu, \lambda \in \mathbb{R}^p$  são os multiplicadores de Lagrange do problema irrestrito. Esse método penalização é mais estável numericamente que os anteriores, visto que não há necessidade de fazer  $\lambda \rightarrow \infty$  ou  $\lambda \rightarrow 0$  e, além disso:

- a solução inicial não precisa ser factível;
- é possível encontrar  $g(x) = 0$  com maior precisão, diferentemente de métodos de penalidade interna e externa;
- os multiplicadores de Lagrange diferentes de zero identificam as restrições ativas no ponto solução.

Em [Pedroso et al., 2011], o método dos multiplicadores de Lagrange é utilizado em conjunto com métodos de otimização sem derivadas, incluindo o algoritmo Nelder-Mead. Outras aplicações recentes também podem ser encontradas [Afonso et al., 2010; Houska et al., 2016; Gramacy et al., 2016].

#### 2.4.4 Tratamento de restrições de caixa

Uma estratégia comum para tratamento de violações de caixa é o truncamento da variável [Luersen et al., 2004]. Isto é,

$$x_k = \max\{\underline{x}, \min\{x_k, \bar{x}\}\}. \quad (2.10)$$

Entretanto, essa operação pode mudar a direção de busca do método, como está ilustrado na Figura 2.2. Implicitamente, o truncamento de  $x_{k+1}$  nos limites da caixa altera a direção de busca de  $d_k$  para  $d'_k$ .

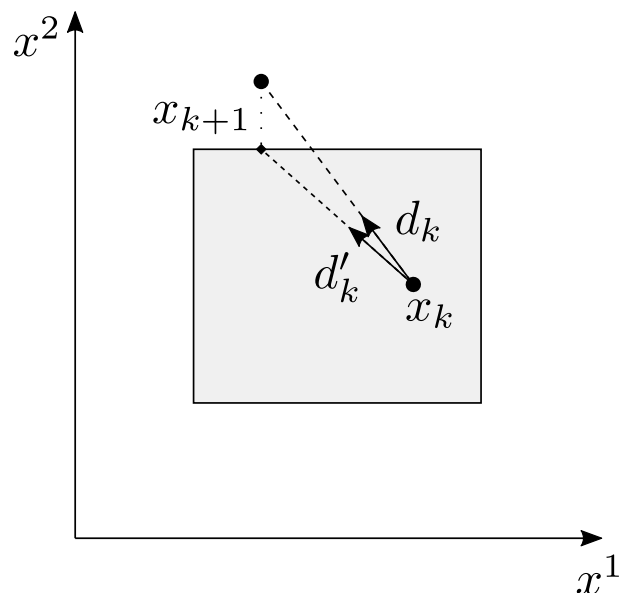


Figura 2.2: O truncamento de um ponto que ultrapassa os limites da caixa pode mudar a direção de busca da atualização  $x_{k+1} = x_k + \alpha_k d_k$ .

Esse comportamento é um dos motivadores que norteiam a solução proposta no Capítulo 4, pois para o método Nelder-Mead simplex esse comportamento indesejado pode resultar em uma

degeneração do simplex e em uma consequente convergência prematura (ver Figura 2.3).

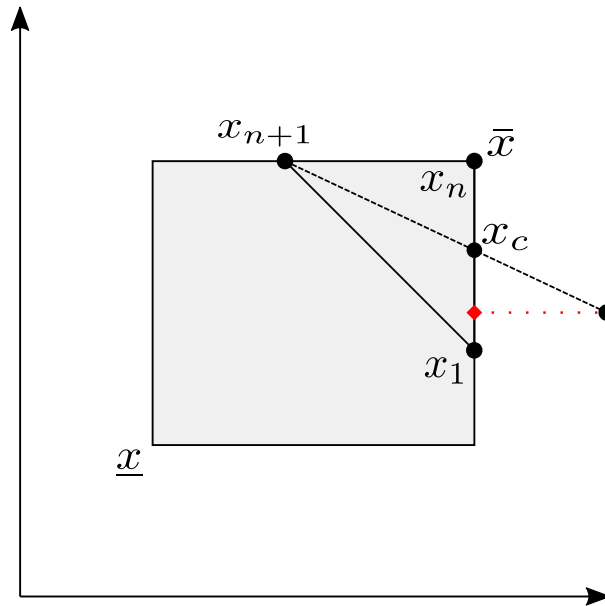


Figura 2.3: Aqui,  $x_1$  é o melhor ponto,  $x_{n+1}$  é o pior e  $x_n$  o segundo pior ponto. O ponto  $x_c$  é o centróide relativo a  $x_{n+1}$ . Note que  $x_r$  reside fora da caixa e o truncamento desse ponto gera um novo simplex com pontos colineares, degenerando o simplex.

Uma outra estratégia é tratar os limites de caixa como restrições lineares de desigualdade [Djikpéssé et al., 2011], acrescentando a complexidade do problema original através da inserção de  $n$  restrições.

Para métodos de direção de busca, também existe a possibilidade de se determinar  $\alpha_k^*$ , o tamanho do passo ótimo da iteração  $k$ , de tal forma que  $x_{k+1} \in \mathbb{B}$ , evitando o truncamento da variável.

## 2.5 Conclusão do capítulo

Esse capítulo resume os principais fundamentos necessários para o entendimento das soluções apresentadas nos próximos capítulos. Além disso, algumas discussões e trabalhos relacionados são apresentados como forma de inserção desta dissertação no universo da otimização não linear.

O primeiro ponto abordado é relacionado com os métodos de direção de busca e métodos de busca direta. Dentro da discussão sobre os métodos de direção de busca, é dada especial atenção para métodos de busca unidirecional, pois este é o tema do Capítulo 3. Já para métodos de busca direta, que não fazem avaliação do gradiente, especial atenção é dada ao método Nelder-Mead simplex (NM), pois este é o foco do Capítulo 4.

O segundo ponto abordado é relativo ao tratamento de restrições utilizados em métodos de

otimização irrestrita. Abordou-se os métodos de penalidades interna, externa e o método dos multiplicadores de Lagrange.

## Capítulo 3

# Otimização unidimensional multimodal

### 3.1 Introdução

Este capítulo apresenta um novo *framework* para otimização unidimensional que é capaz de mapear múltiplos mínimos em funções multimodais. A existência de múltiplos mínimos é provada através dos padrões- $\vee$  [Vieira and Lisboa, 2014] e através dos padrões- $\wedge$ , um conceito aqui introduzido. Toda a fundamentação teórica que suporta o *framework* proposto é apresentada na próxima seção. Em seguida, o algoritmo básico é apresentado e alguns testes são realizados para ilustrar seu comportamento.

### 3.2 Caracterização de multimodalidade pelos padrões-( $\vee, \wedge$ )

Os resultados relacionados ao padrão- $\vee$  são extraídos de [Vieira and Lisboa, 2014] e aqui estendidos para a definição do padrão- $\wedge$ . Enquanto o primeiro padrão é usado para definir a existência de um *mínimo local*, o segundo define a existência de um *máximo local*.

As próximas definições de mínimos e máximos locais são relativas ao problema de otimização unidimensional definido em (2.4).

**Definição 3** (mínimo local). *Para o problema de otimização unidimensional (2.4), um ponto  $\alpha^*$  é um mínimo local se e somente se*

$$\exists \epsilon > 0 : f(\alpha^*) \leq f(\alpha), \forall \alpha \in [\underline{\alpha}, \bar{\alpha}], |\alpha - \alpha^*| < \epsilon. \quad (3.1)$$

Similarmente, um máximo local pode ser definido da seguinte forma:

**Definição 4** (máximo local). *Para o problema de otimização unidimensional (2.4), um ponto  $\alpha^*$  é um máximo local se e somente se*

$$\exists \epsilon > 0 : f(\alpha^*) \geq f(\alpha), \forall \alpha \in [\underline{\alpha}, \bar{\alpha}], |\alpha - \alpha^*| < \epsilon. \quad (3.2)$$

Os próximos resultados são baseados em valores relativos de função dentro de uma sequência ordenada de pontos, extraídos de [Vieira and Lisboa, 2014].

**Definição 5** (padrão- $\vee$ ). *A sequência  $\alpha_1 < \alpha_2 < \alpha_3$  é um padrão- $\vee$  de  $f$  se e somente se  $f(\alpha_2) \leq f(\alpha_1)$  e  $f(\alpha_2) \leq f(\alpha_3)$ .*

**Teorema 1** (padrão- $\vee$ ). *Ao menos um **mínimo local** de  $f(\alpha)$ , sujeito a  $\alpha \in [\underline{\alpha}, \bar{\alpha}]$ , está contido em qualquer **padrão- $\vee$**  quando  $\alpha_1 < \alpha_2 < \alpha_3$ ,  $[\alpha_1, \alpha_3] \subseteq [\underline{\alpha}, \bar{\alpha}]$ .*

*Demonstração.* Ver [Vieira and Lisboa, 2014]. □

**Definição 6** (meio-padrão- $\vee$ ). *A sequência  $\alpha_1 < \alpha_2$  é um meio-padrão- $\vee$  de  $f$  no intervalo  $[\underline{\alpha}, \bar{\alpha}]$  se e somente se  $f(\alpha_1) \leq f(\alpha_2)$  e  $\alpha_1 = \underline{\alpha}$ , ou  $f(\alpha_2) \leq f(\alpha_1)$  e  $\alpha_2 = \bar{\alpha}$ .*

**Teorema 2** (meio-padrão- $\vee$ ). *Todo **meio-padrão- $\vee$**  para  $\alpha_1 < \alpha_2$ ,  $[\alpha_1, \alpha_2] \subset [\underline{\alpha}, \bar{\alpha}]$ , contém pelo menos um **mínimo local** de  $f(\alpha)$  sujeito a  $\alpha \in [\underline{\alpha}, \bar{\alpha}]$  no intervalo  $[\alpha_1, \alpha_2]$  se  $f(\alpha_1) \leq f(\alpha_2)$  e  $\alpha_1 = \underline{\alpha}$ , ou em  $(\alpha_1, \alpha_2]$  se  $f(\alpha_2) \leq f(\alpha_1)$  e  $\alpha_2 = \bar{\alpha}$ .*

*Demonstração.* Veja [Vieira and Lisboa, 2014]. □

A seguir, os resultados obtidos para o padrão- $\vee$ , definidos em [Vieira and Lisboa, 2014], que são usados para definir mínimos locais, são estendidos para o **padrão- $\wedge$** , que está relacionado com a existência de máximos locais.

Enquanto a presença de padrão- $\vee$  garante a existência de pelo menos um mínimo local dentro do intervalo de busca, o padrão- $\wedge$  é usado para dividir o intervalo de tal forma que um padrão- $\vee$ , e conseqüentemente um mínimo local, exista em cada novo intervalo.

**Definição 7** (padrão- $\wedge$ ). *Uma sequência ordenada  $\alpha_1 < \alpha_2 < \alpha_3$  possui um padrão- $\wedge$  da função  $f$  se e somente se  $f(\alpha_2) \geq f(\alpha_1)$  e  $f(\alpha_2) \geq f(\alpha_3)$ .*

**Teorema 3** (padrão- $\wedge$ ). *Todo **padrão- $\wedge$**  para  $\alpha_1 < \alpha_2 < \alpha_3$ ,  $[\alpha_1, \alpha_3] \subseteq [\underline{\alpha}, \bar{\alpha}]$ , contém pelo menos um **máximo local** de  $f(\alpha)$  sujeito a  $\alpha \in [\underline{\alpha}, \bar{\alpha}]$  no intervalo aberto  $(\alpha_1, \alpha_3)$ .*

*Demonstração.* Para um padrão- $\wedge$ , dado um conjunto ordenado  $\alpha_1 < \alpha_2 < \alpha_3$ ,  $f(\alpha_2) \geq f(\alpha_1)$  e  $f(\alpha_2) \geq f(\alpha_3)$ , ou  $\alpha_2 \in (\alpha_1, \alpha_3)$  já é um máximo local dado que existem pontos vizinhos,  $\alpha_1$  e  $\alpha_3$ , em ambos os lados  $\alpha_2$  cujos valores de função não são maiores, ou existe um outro ponto no intervalo  $(\alpha_1, \alpha_3)$  com o valor de função maior do que  $f(\alpha_2)$ .

Para o último caso, basta escolher um ponto dentro de  $(\alpha_1, \alpha_3)$  com o maior valor de função e este será, por definição, um máximo local. □

**Definição 8** (meio-padrão- $\wedge$ ). *Uma sequência ordenada  $\alpha_1 < \alpha_2$  possui um meio-padrão- $\wedge$  da função  $f$  no intervalo  $[\underline{\alpha}, \bar{\alpha}]$  se e somente se  $f(\alpha_1) \geq f(\alpha_2)$  e  $\alpha_1 = \underline{\alpha}$ , ou  $f(\alpha_2) \geq f(\alpha_1)$  e  $\alpha_2 = \bar{\alpha}$ .*

**Teorema 4** (meio-padrão- $\wedge$ ). *Todo meio-padrão- $\wedge$  para  $\alpha_1 < \alpha_2$ ,  $[\alpha_1, \alpha_2] \subset [\underline{\alpha}, \bar{\alpha}]$ , contém pelo menos um máximo local de  $f(\alpha)$  sujeito a  $\alpha \in [\underline{\alpha}, \bar{\alpha}]$  no intervalo  $[\alpha_1, \alpha_2]$  se  $f(\alpha_2) \leq f(\alpha_1)$  e  $\alpha_1 = \underline{\alpha}$ , ou em  $(\alpha_1, \alpha_2]$  se  $f(\alpha_1) \leq f(\alpha_2)$  e  $\alpha_2 = \bar{\alpha}$ .*

*Demonstração.* A prova vem da definição de meio-padrão- $\wedge$  e de máximo local. Como  $\alpha_1 < \alpha_2$ , considere o tipo de meio-padrão- $\wedge$  no qual  $f(\alpha_2) \leq f(\alpha_1)$  e  $\alpha_1 = \underline{\alpha}$ , ou  $\alpha_1 \in (\alpha_1, \alpha_2)$  já é um máximo local, visto que existe um vizinho  $\alpha_2$  cujo valor de função não é maior, ou deve existir um ponto em  $[\alpha_1, \alpha_2]$  com o maior valor de função. Para o último caso, qualquer ponto em  $(\alpha_1, \alpha_2)$  com o maior valor de função é, por definição, um máximo local.

Para o tipo de meio-padrão- $\wedge$  no qual  $f(\alpha_1) \leq f(\alpha_2)$  e  $\alpha_2 = \bar{\alpha}$ , a demonstração vem por simetria.  $\square$

A principal contribuição deste capítulo é a elaboração de um algoritmo de otimização unidimensional capaz de mapear múltiplos mínimos. Os seguintes teoremas servem de suporte para a elaboração do mesmo.

**Teorema 5.** *(existência de um padrão- $\vee$  entre padrões- $\wedge$ ) Considere a sequência  $\alpha_1 < \alpha_2 < \dots < \alpha_w$ , onde  $w \in \mathbb{N}$ ,  $w \geq 5$ , é o número de pontos da sequência. Se dois padrões- $\wedge$  podem ser encontrados na sequência, um padrão- $\vee$  necessariamente tem que existir entre os padrões- $\wedge$  encontrados.*

*Demonstração.* Considere, sem perda de generalidade, que o primeiro padrão- $\wedge$  foi encontrado em  $\alpha_1 < \alpha_2 < \alpha_3$ , e o segundo padrão- $\wedge$  em  $\alpha_{i-1} < \alpha_i < \alpha_{i+1}$ , tal que  $i \in \{4, \dots, w-1\}$ .

Qualquer ponto  $\alpha_j$ ,  $j \in \{3, \dots, w-2\}$ , tal que  $f(\alpha_j) \leq f(\alpha_2)$  e  $f(\alpha_j) \leq f(\alpha_i)$ , caracteriza uma sequência  $\alpha_2 < \alpha_j < \alpha_i$  com um padrão- $\vee$  (ver Figura 3.1).

Pela definição de padrão- $\wedge$ ,  $f(\alpha_j) \leq \min\{f(\alpha_2), f(\alpha_i)\}$  é sempre verdadeira para  $j = 3$  ou  $j = i-1$ , garantindo a existência de um padrão- $\vee$  para algum  $3 \leq j \leq i-1$ .  $\square$

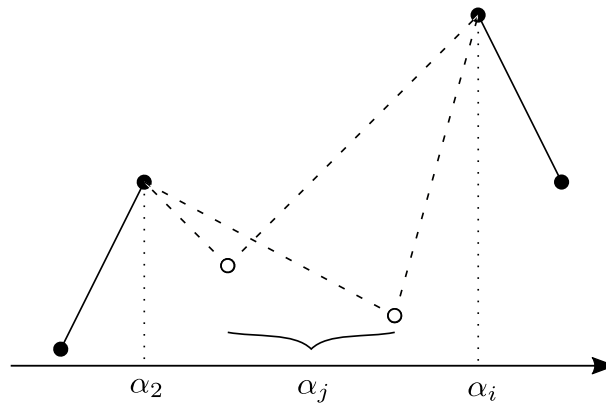


Figura 3.1: Se dois padrões- $\wedge$  podem ser encontrados em uma sequência de pontos, deve existir um  $\alpha_j$  tal que  $f(\alpha_j) \leq \min\{f(\alpha_2), f(\alpha_i)\}$ , onde  $\alpha_2$  e  $\alpha_i$  são os pontos centrais dos dois padrões- $\wedge$ , caracterizando um padrão- $\vee$  entre eles.

**Teorema 6.** (existência de um meio-padrão- $\vee$  ou um padrão- $\vee$  entre um padrão- $\wedge$  e  $\bar{\alpha}$ ) Considere a sequência  $\alpha_1 < \alpha_2 < \dots < \alpha_w$ , onde  $w \in \mathbb{N}$ ,  $w \geq 3$  é o número de pontos e  $\alpha_w = \bar{\alpha}$ . Se um padrão- $\wedge$  pode ser encontrado nessa sequência, um padrão- $\vee$  ou um meio-padrão- $\vee$  devem existir na sequência.

*Demonstração.* Considere, sem perda de generalidade, que o padrão- $\wedge$  está em  $\alpha_1 < \alpha_2 < \alpha_3$ . Qualquer sequência terminando em  $\bar{\alpha}$  define um meio-padrão- $\vee$  ou um meio-padrão- $\wedge$ . Caso ele termine com um meio-padrão- $\wedge$ ,  $\alpha_{w-1} < \alpha_w$  implica que  $f(\alpha_w) \geq f(\alpha_{w-1})$ . O primeiro ponto  $i$ ,  $\alpha_i$ ,  $i \in \{3, \dots, w-1\}$ , tal que  $f(\alpha_i) \leq f(\alpha_{i+1})$ , forma um padrão- $\vee$  em  $\alpha_{i-1} < \alpha_i < \alpha_{i+1}$ , ou um meio-padrão- $\wedge$  no caso em que  $i = w-1$ . Ver Figura 3.2 para maiores detalhes.  $\square$

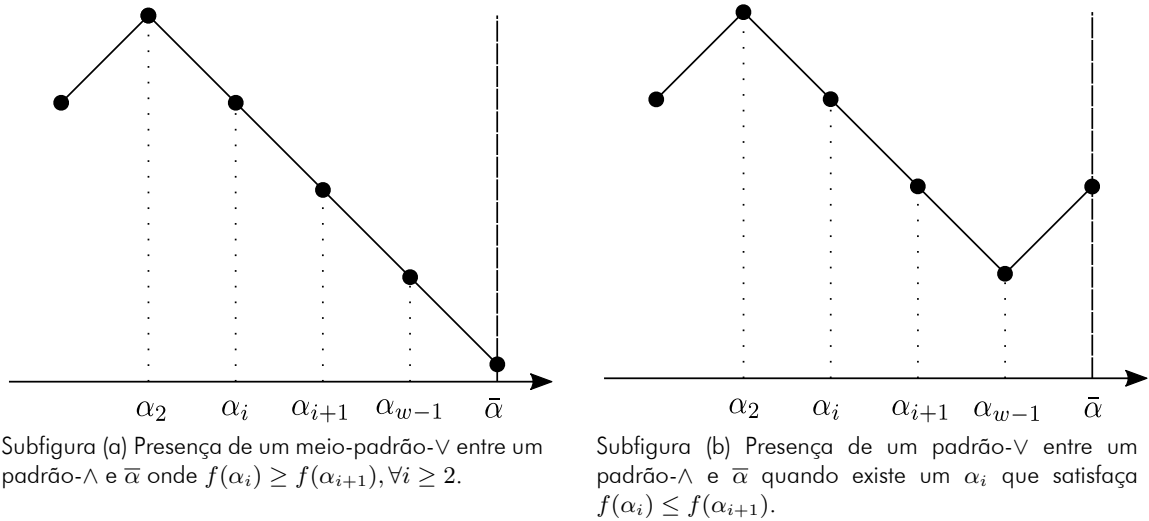


Figura 3.2: Se um padrão- $\wedge$  está presente em uma sequência pontos, necessariamente existe um padrão- $\vee$  ou meio-padrão- $\vee$  entre o padrão- $\wedge$  e os limites do espaço de busca (i.e.  $\underline{\alpha}$  ou  $\bar{\alpha}$ ).

**Teorema 7.** (existência de um meio-padrão- $\vee$  ou um padrão- $\vee$  entre um padrão- $\wedge$  e  $\underline{\alpha}$ ) Considere a sequência  $\alpha_1 < \alpha_2 < \dots < \alpha_w$ , onde  $w \in \mathbb{N}$ ,  $w \geq 3$ , é o número de pontos e  $\alpha_1 = \underline{\alpha}$ . Se um padrão- $\wedge$  pode ser encontrado nessa sequência, um padrão- $\vee$  ou um meio-padrão- $\vee$  devem existir nessa sequência.

*Demonstração.* A prova é similar à do Teorema 6. Detalhes gráficos podem ser encontrados na Figura 3.3.  $\square$

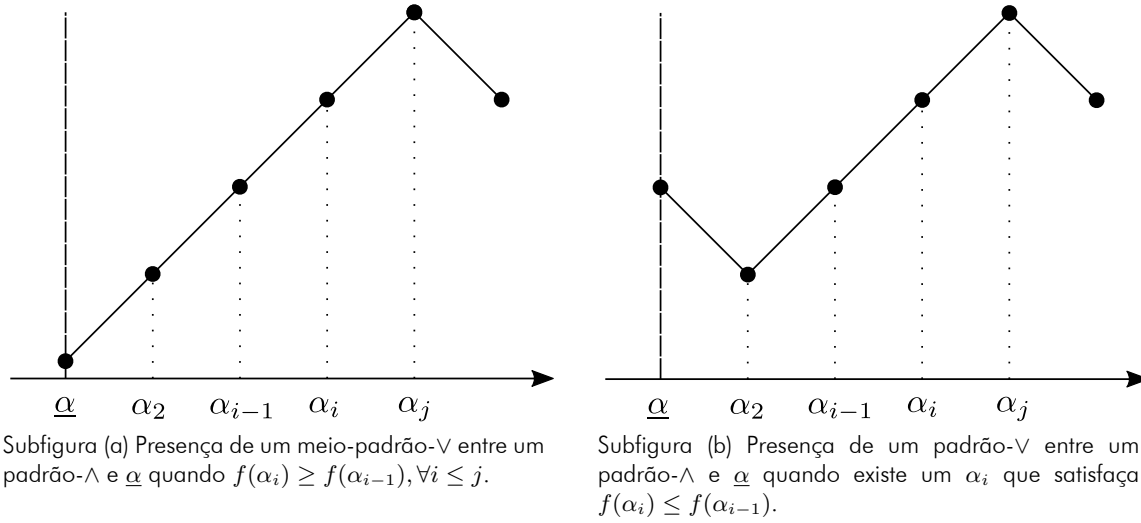


Figura 3.3: Se um padrão- $\wedge$  está presente em uma sequência pontos, necessariamente existe um padrão-V ou meio-padrão-V entre o padrão- $\wedge$  e os limites do espaço de busca (i.e.  $\underline{\alpha}$  ou  $\bar{\alpha}$ ).

### 3.3 Otimização unidimensional multimodal

Esta seção apresenta um *framework* para otimização unidimensional baseado em comparação entre pontos, que pode ser usado para mapear múltiplos mínimos em funções multimodais. A estratégia do *framework* é baseada no fato de que quando um padrão- $\wedge$  é encontrado, a região de busca pode ser dividida em novos subintervalos que contenham um padrão-V ou um meio-padrão-V. Então, um método de busca unidimensional, como o método da seção áurea, pode ser empregado em cada novo subintervalo, preservando as propriedades e convergência do método original.

O *framework* proposto é apresentado no [Algoritmo 2](#). Note que as entradas do *framework* proposto são as mesmas de um método de otimização unidimensional tradicional, enquanto que a saída é um conjunto  $\mathbb{A}^*$  com todos os mínimos locais encontrados dentro do intervalo de busca. Um importante aspecto do algoritmo é o armazenamento dos pontos  $(\alpha_1, \dots, \alpha_m)$  de cada execução (linha 2), que definem a sequência ordenada do método subjacente. Se um padrão- $\wedge$  é encontrado em  $\alpha_s$  (linha 3), o intervalo  $[\alpha_1, \alpha_m]$  é dividido em dois novos subintervalos (ver [Figura 3.4](#)), isto é,  $[\alpha_l, \alpha_s]$  e  $[\alpha_s, \alpha_u]$ . Os pontos  $\alpha_l$  e  $\alpha_u$  são definidos de tal forma que um padrão-V ou um meio-padrão-V deva existir dentro de cada subintervalo, conforme [Teorema 5](#), [Teorema 6](#) e [Teorema 7](#).

A [Figura 3.5](#) ilustra um exemplo de como  $\alpha_u$  deve ser definido: dentre os pontos armazenados,  $\alpha_u$  é o menor ponto maior que  $\alpha_m$  onde um padrão-V pode ser encontrado no intervalo  $[\alpha_s, \alpha_u]$  (ver [Subfigura 3.5a](#)). Se um padrão-V não pode ser encontrado,  $\alpha_u = \bar{\alpha}$  (ver [Subfigura 3.5b](#)), caracterizando um meio-padrão-V. De qualquer modo, um mínimo local existe em cada subintervalo.

A determinação de  $\alpha_l$  vem por simetria:  $\alpha_l$  é o maior ponto menor que  $\alpha_1$  onde um padrão-V pode ser definido no intervalo  $[\alpha_l, \alpha_s]$ . Se um padrão-V não pode ser encontrado,  $\alpha_l = \underline{\alpha}$ , caracterizando um meio-padrão-V. As linhas 4 a 13 do algoritmo são responsáveis pela determinação de cada novo subintervalo.

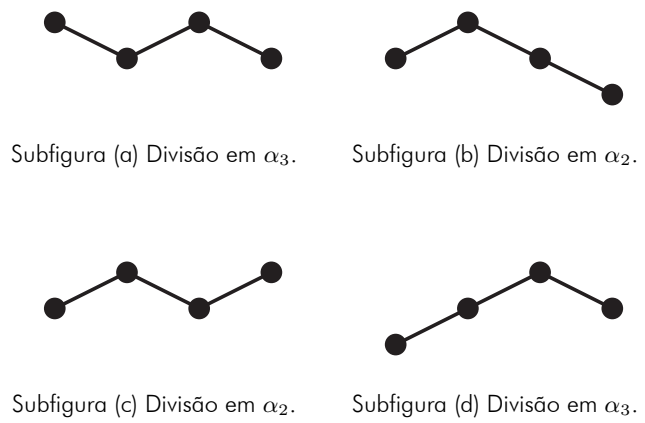


Figura 3.4: Ponto de divisão  $\alpha_s$  para todos possíveis padrões- $\wedge$  em uma seqüência com quatro pontos ordenados  $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$ .

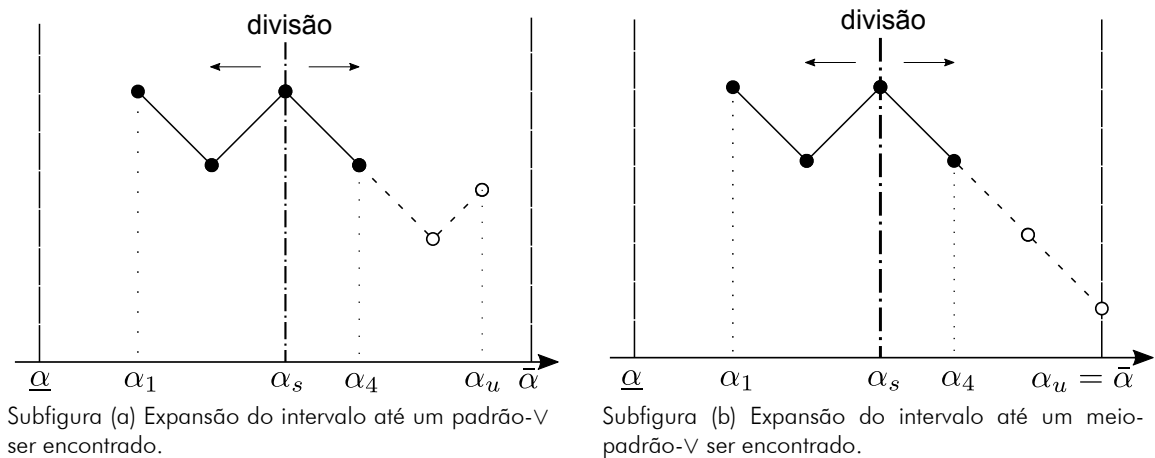


Figura 3.5: A divisão do intervalo deve garantir a existência de um padrão-V, ou pelo menos um meio-padrão-V e, portanto, um mínimo local. Os pontos pretos representam a seqüência  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$  da execução atual e os pontos brancos os pontos armazenados de execuções anteriores.

---

**Algoritmo 2** *Framework* proposto: multimodal line search (MLS).

---

**Entrada**

$[\underline{\alpha}, \bar{\alpha}]$  intervalo de busca  
 $f : \mathbb{R} \rightarrow \mathbb{R}$  função  
 LINESEARCH método de otimização unidimensional  
 $\epsilon$  precisão numérica

**Saída**

$\mathbb{A}^*$  conjunto com todos os mínimos locais  $\alpha_i^*$  encontrados no intervalo  $[\underline{\alpha}, \bar{\alpha}]$

```

1: function MULTIMODALLINESEARCH( $f, \underline{\alpha}, \bar{\alpha}, \epsilon$ )
2:    $\kappa \leftarrow \kappa \cup \{\alpha_1, \dots, \alpha_m\}, m > 2$  ▷ armazena os pontos
3:   if  $f(\alpha_s) \geq f(\alpha_{s-1})$  and  $f(\alpha_s) \geq f(\alpha_{s+1}), s = 2, \dots, m - 1$  then
4:     if  $\exists j \mid \kappa_j < \alpha_s$  and  $f(\kappa_j) \leq f(\kappa_{j-1})$  and  $f(\kappa_j) \leq f(\kappa_{j+1})$  then
5:        $\alpha_l \leftarrow \max\{\kappa_j\}, \forall j$ 
6:     else ▷ nenhum padrão-V encontrado
7:        $\alpha_l \leftarrow \underline{\alpha}$  ▷ meio-padrão-V
8:     end if
9:     if  $\exists j \mid \kappa_j > \alpha_s$  and  $f(\kappa_j) \leq f(\kappa_{j-1})$  and  $f(\kappa_j) \leq f(\kappa_{j+1})$  then
10:       $\alpha_u \leftarrow \min\{\kappa_j\}, \forall j$ 
11:     else ▷ nenhum padrão-V encontrado
12:       $\alpha_u \leftarrow \bar{\alpha}$  ▷ meio-padrão-V
13:     end if
14:      $\mathbb{A}^* \leftarrow \mathbb{A}^* \cup \text{MULTIMODALLINESEARCH}(f, \alpha_l, \alpha_s, \epsilon)$ 
15:      $\mathbb{A}^* \leftarrow \mathbb{A}^* \cup \text{MULTIMODALLINESEARCH}(f, \alpha_s, \alpha_u, \epsilon)$ 
16:   else ▷ nenhum padrão-Λ encontrado
17:      $\mathbb{A}^* \leftarrow \text{LINESEARCH}(f, \underline{\alpha}, \bar{\alpha}, \epsilon)$ 
18:   end if
19:   return  $\mathbb{A}^*$ 
20: end function

```

---

Uma vez definidos os novos subintervalos, o algoritmo proposto é chamado recursivamente (linhas 14 e 15). Se em algum momento um padrão- $\wedge$  não é encontrado em uma sequência, o método de otimização subjacente é executado (linha 17).

A convergência do algoritmo proposto é garantida pela precisão numérica  $\epsilon$  do método de busca unidimensional. Isto é, se um padrão- $\wedge$  é encontrado em todo novo subintervalo durante a execução, o algoritmo é finalizado quando a largura do intervalo atingir  $\epsilon(\alpha_m - \alpha_1)$ .

**Teorema 8** (existência de um mínimo local em todo novo subintervalo do algoritmo de otimização unidimensional multimodal). *Existe um mínimo local em cada novo subintervalo  $[\alpha_l, \alpha_s]$  e  $[\alpha_s, \alpha_u]$  do framework multimodal, denominado *multimodal line search algorithm (MLS)*.*

*Demonstração.* Como  $\alpha_s$  é um ponto que define um padrão- $\wedge$ , qualquer ponto  $\alpha_u$ ,  $\alpha_u > \alpha_s$ , tal que  $f(\alpha_u) \geq f(\alpha_{u-1})$  e  $f(\alpha_u) \geq f(\alpha_{u+1})$  caracteriza um intervalo  $[\alpha_s, \alpha_u]$  com um padrão- $\vee$  (Teorema 5), e portanto um mínimo local. Similarmente, qualquer ponto  $\alpha_l$ ,  $\alpha_l < \alpha_s$ , tal que  $f(\alpha_l) \geq f(\alpha_{l-1})$  e  $f(\alpha_l) \geq f(\alpha_{l+1})$  caracteriza um intervalo  $[\alpha_l, \alpha_s]$  com um mínimo local. Além do mais, se um padrão- $\vee$  não pode ser encontrado nos intervalos  $[\underline{\alpha}, \alpha_s]$  e  $[\alpha_s, \bar{\alpha}]$ , a existência de meio-padrão- $\vee$  é garantida pelo resultado do Teorema 6 e do Teorema 7, respectivamente, que por sua vez também garantem a existência de um mínimo local em cada novo subintervalo.  $\square$

Note que a característica multimodal do *framework* é dependente do primeiro padrão encontrado no espaço de busca original. Ou seja, dado que um padrão- $\wedge$  não é encontrado na primeira iteração, o algoritmo se comporta como um método de otimização unidimensional ao qual ele foi acoplado. Este comportamento é explorado em mais detalhes na próxima seção.

## 3.4 Exemplos

Apesar do *framework* proposto poder ser incorporado a qualquer método de otimização unidimensional baseado em comparação de pontos, neste trabalho ele é acoplado ao algoritmo *enhanced golden section* (EGS) [Vieira and Lisboa, 2014], resultando em novo algoritmo denominado *multimodal golden section* (MGS).

### 3.4.1 Exemplo I

O primeiro exemplo visa testar a influência do número de mínimos locais dentro do espaço de busca e suas posições relativas. Nesse sentido, esse experimento é conduzido para o seguinte problema unidimensional:

$$\begin{aligned} & \underset{\alpha}{\text{minimize}} && f_1(\alpha) = -0.5\alpha + \cos(\omega\alpha + \phi) \\ & \text{sujeito a} && 0 \leq \alpha \leq 1. \end{aligned} \tag{3.3}$$

O parâmetro  $\omega$  controla a componente de frequência de  $f_1(\alpha)$ . Ou seja, controla o número de mínimos locais dentro de  $\alpha \in [0, 1]$ : quando  $\omega$  aumenta, o número de mínimos locais também aumenta (ver Figura 3.6). O parâmetro  $\phi$  controla a fase da função, que é responsável pela posição relativa dos padrões dentro do espaço de busca. Testes são conduzidos para  $\omega$  variando de 10 a 100, com um passo de 10, e  $\phi = \{0, \pi/6, \pi/4, \pi/2\}$ .

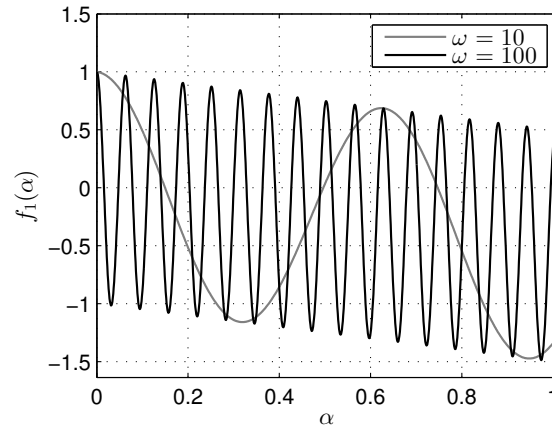


Figura 3.6: Instâncias da função de teste para  $\omega = 10, 100$  e  $\phi = 0$ . Note que independentemente dos valores de  $\omega$  e  $\phi$ , os menores valores função estão no sentido de  $\alpha = 1$ .

Inicialmente, são comparadas as melhores solução encontradas pelo MGS com a solução encontrada pelo método subjacente EGS. Em seguida, a característica multimodal do *framework* proposto é ilustrada em maiores detalhes.

A Tabela 3.1 apresenta a solução ótima do algoritmo EGS e melhor solução encontrada pelo algoritmo MGS. Note que a solução encontrada pelo método proposto nunca é pior que a solução do algoritmo subjacente e para 80% das instâncias desse problema a solução do algoritmo MGS é melhor que a solução do EGS.

Tabela 3.1: Valores ótimos de  $f_1(\alpha^*)$  obtidos pelos algoritmos EGS e MGS. Os resultados estão dispostos no formato (EGS, MGS), com o menor valor obtido em negrito.

$\omega$	$\phi = 0$	$\phi = \pi/6$	$\phi = \pi/4$	$\phi = \pi/2$
10	(-1.1583, <b>-1.4725</b> )	(-1.1321, <b>-1.4463</b> )	(-1.1191, <b>-1.4332</b> )	( <b>-1.3939</b> , <b>-1.3939</b> )
20	(-1.0789, <b>-1.3930</b> )	(-1.2228, <b>-1.3799</b> )	(-1.2163, <b>-1.3734</b> )	(-1.1967, <b>-1.4129</b> )
30	(-1.2619, <b>-1.4714</b> )	(-1.0438, <b>-1.3579</b> )	( <b>-1.3536</b> , <b>-1.3536</b> )	( <b>-1.2358</b> , <b>-1.2358</b> )
40	(-1.1179, <b>-1.4320</b> )	(-1.1899, <b>-1.4255</b> )	(-1.1866, <b>-1.4222</b> )	(-1.1768, <b>-1.4910</b> )
50	( <b>-1.4713</b> , <b>-1.4713</b> )	(-1.2776, <b>-1.3404</b> )	(-1.1493, <b>-1.4634</b> )	(-1.2043, <b>-1.4556</b> )
60	(-1.2357, <b>-1.4975</b> )	(-1.0742, <b>-1.4931</b> )	(-1.0720, <b>-1.4909</b> )	(-1.1178, <b>-1.3796</b> )
70	(-1.1122, <b>-1.4264</b> )	(-1.2431, <b>-1.4226</b> )	(-1.2413, <b>-1.4208</b> )	(-1.2356, <b>-1.4600</b> )
80	( <b>-1.3731</b> , <b>-1.3731</b> )	( <b>-1.4091</b> , <b>-1.4091</b> )	( <b>-1.4074</b> , <b>-1.4074</b> )	(-1.1276, <b>-1.4811</b> )
90	(-1.1920, <b>-1.4713</b> )	(-1.1891, <b>-1.4334</b> )	(-1.1178, <b>-1.4320</b> )	(-1.1484, <b>-1.4276</b> )
100	( <b>-1.4555</b> , <b>-1.4555</b> )	(-1.2644, <b>-1.4529</b> )	(-1.2631, <b>-1.4516</b> )	(-1.1649, <b>-1.4791</b> )

A Figura 3.7 apresenta a razão média, para todos valores de  $\phi$ , entre o número de mínimos mapeados ( $m$ ) e o número total de mínimos locais presentes ( $\bar{m}$ ) no espaço de busca para cada valor de  $\omega$ . Como o algoritmo EGS sempre encontra um mínimo,  $m = 1$ , sua razão decresce à

medida que  $\omega$  aumenta ( $\bar{m} \propto \omega$ ). A curva do MGS possui um comportamento similar na média, entretanto a razão obtida para este algoritmo é sempre maior que a razão do EGS,  $m > 1$ , realçando a capacidade do *framework* proposto de mapear múltiplos mínimos.

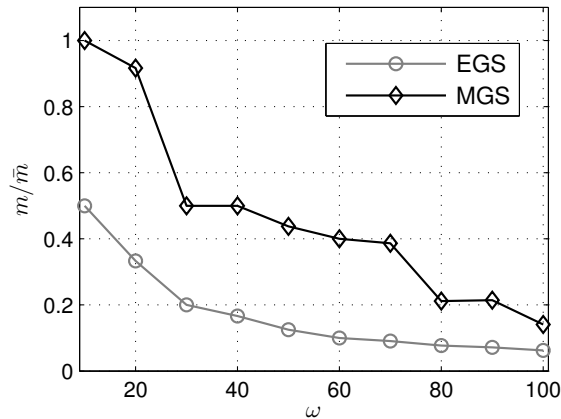


Figura 3.7: Razão média da quantidade de mínimos locais mapeados ( $m$ ) em relação ao total de mínimos ( $\bar{m}$ ) dentro do espaço de busca.

É importante ressaltar que o *framework* proposto não garante que todos os mínimos presentes no espaço de busca original sejam mapeados (ver Figura 3.8). Entretanto, uma vez que um padrão- $\Lambda$  é encontrado, a divisão consecutiva do intervalo mapeia mais de um mínimo, aumentando a chance de se encontrar um mínimo melhor que a solução do método unidimensional subjacente.

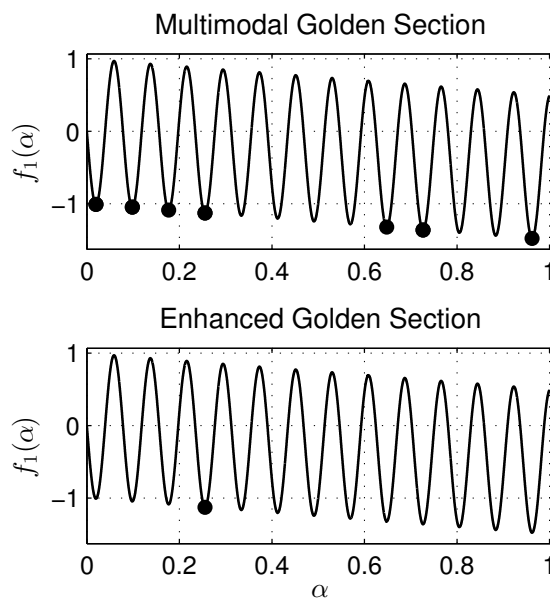


Figura 3.8: Instância do problema para  $\omega = 80$ ,  $\phi = \pi/2$ . Mínimos mapeados em negrito.

### 3.4.2 Exemplo II

O segundo exemplo testa a influência da posição relativa do mínimo global dentro do espaço de busca e a influência da magnitude da multimodalidade, para um fixo número de mínimos locais. Nesse sentido, o seguinte problema de otimização é formulado:

$$\begin{aligned} & \text{minimize} && f_2(\alpha) = |\alpha - \zeta| + \eta \sin(50\alpha) \\ & \text{sujeito a} && 0 \leq \alpha \leq 1. \end{aligned} \tag{3.4}$$

O parâmetro  $\zeta \in [0, 1]$  controla a posição relativa do mínimo global de  $f_2(\alpha)$  entre  $\alpha \in [0, 1]$ , enquanto  $\eta$  controla a magnitude do termo multimodal em relação ao termo  $|\alpha - \zeta|$  (ver [Figura 3.9](#) para detalhes). Testes são conduzidos para  $\zeta = \{0.00, 0.25, 0.50, 0.75, 1.00\}$  e  $\eta = \{0.5, 1.0, 1.5, 2.5\}$ .

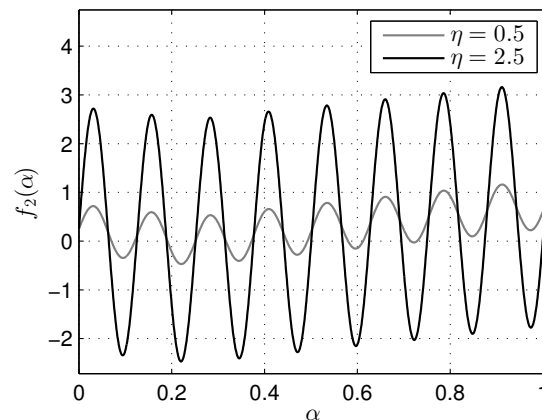


Figura 3.9: Instâncias de teste para  $\eta = 0.5, 2.5$  e  $\zeta = 0.25$ . Note que o mínimo global está próximo a  $\zeta = 0.25$  e o número de mínimos locais é fixo.

A [Tabela 3.2](#) mostra a razão do número de mínimos mapeados pelo algoritmo MGS para instâncias do problema (3.4). Note que o algoritmo tende a achar mais mínimos à medida que  $\zeta$  diminui e  $\eta$  aumenta. Basicamente, para esse problema-exemplo um menor valor de  $\zeta$  favorece a presença de um padrão- $\Lambda$  na primeira recorrência do algoritmo. O comportamento relacionado ao parâmetro  $\eta$  é relacionado ao fato que este controla o nível da componente multimodal de  $f_2(\alpha)$ , isto é, quanto maior for  $\eta$ , mais acentuada é a multimodalidade da função de teste, até o ponto em que a influência de  $\eta$  é a maior que a influência de  $\zeta$ , fato que pode ser notado nas instâncias em que  $\eta = 2.5$ . A [Figura 3.10](#) e a [Figura 3.11](#) ilustram esses comportamentos.

Tabela 3.2: Razão dos mínimos mapeados ( $m/\bar{m}$ ) do algoritmo MGS, onde 0.125 é o menor valor possível, isto é, o número total de mínimos locais é  $\bar{m} = 8$ .

$\eta$	$\zeta = 0.0$	$\zeta = 0.25$	$\zeta = 0.50$	$\zeta = 0.75$	$\zeta = 1.00$
0.5	0.750	0.750	0.125	0.125	0.125
1.0	0.750	0.750	0.125	0.125	0.125
1.5	0.750	0.750	0.125	0.125	0.125
2.5	0.875	0.875	1.000	1.000	1.000

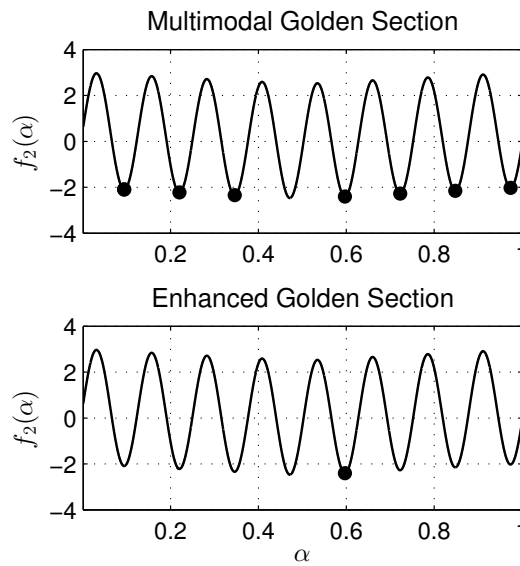


Figura 3.10: Mínimos mapeados em negrito para a instância de  $f_2(\alpha)$  onde  $\zeta = 0.50$  e  $\eta = 2.50$ .

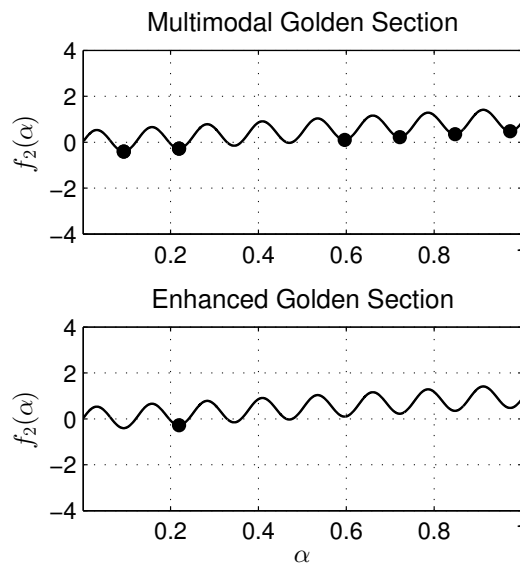


Figura 3.11: Mínimos mapeados em negrito para a instância de  $f_2(\alpha)$  onde  $\eta = 0.50$  de  $\zeta = 0.0$ .

### 3.5 Conclusão do capítulo

Este capítulo apresentou um novo *framework* para otimização unidimensional, MLS - *multimodal line search*, que é capaz de mapear múltiplos mínimos em funções multimodais. A existência de múltiplos mínimos é provada através dos padrões- $(\vee, \wedge)$ . A existência de um padrão- $\wedge$ , em problemas de minimização, é suficiente para a identificação e a subdivisão consecutiva em novos intervalos, cada um com o seu próprio mínimo local.

O *framework* proposto não garante que todos os mínimos presentes no intervalo de busca sejam mapeados. Entretanto, uma vez que um padrão- $\wedge$  é identificado, mais de um mínimo será mapeado e as chances de se mapear um ponto melhor que o ponto obtido pelo método subjacente aumentam, como mostram os experimentos realizados.

Essa estratégia pode ser acoplada a qualquer método de busca unidirecional baseado em comparação de pontos. O acoplamento com o algoritmo da seção áurea resultou em novo algoritmo denominado *multimodal golden section* (MGS).

## Capítulo 4

# Nelder-Mead lexicográfico restrito

### 4.1 Introdução

Este capítulo apresenta as modificações propostas para o algoritmo Nelder-Mead simplex, no sentido de adaptá-lo para a otimização restrita e tratar algumas questões problemáticas do método que podem resultar em uma rápida degeneração do simplex.

### 4.2 Nelder-Mead lexicográfico restrito

A seguir, os principais pontos de contribuição são detalhados.

#### 4.2.1 Comparação lexicográfica

Como discutido no [Capítulo 2](#), existem diversas técnicas para tratamento de restrições que podem ser incorporadas ao método Nelder-Mead simplex (NM). Este trabalho propõe uma [comparação lexicográfica](#), que é um meio de comparação entre pontos no qual há uma relação de ordem preestabelecida. Formalmente:

**Definição 9** (Dominância lexicográfica). *Dados dois pontos  $z_1, z_2 \in \mathbb{R}^q$ , dizemos que  $z_1$  domina lexicograficamente  $z_2$  se existe um índice  $l \leq q$  tal que  $z_1^i = z_2^i$  para  $i = 1, \dots, l-1$ , e  $z_1^l < z_2^l$ , onde  $z^i$  é a posição  $i$  do vetor  $z$ . Esta relação pode ser escrita como  $z_1 \leq_{\text{lex}} z_2$  [[Ehrgott, 2005](#)].*

Do ponto de vista de otimização restrita, a ordenação lexicográfica é uma alternativa aos métodos de penalidades. Observe que a ordenação preserva a ordem de grandeza da função objetivo e das violações, ao passo que métodos baseados em penalidades produzem uma função equivalente que deixa de ter essa informação.

De posse do operador lexicográfico, é criado um vetor  $z$ , associado a cada ponto  $x$ , de tal

maneira que

$$z = \begin{bmatrix} \min\{\lceil \max\{0, 1_n^T \max\{\underline{x} - x, 0\} + 1_n^T \max\{x - \bar{x}, 0\}\} \rceil, 1\} \\ 1_p^T \max\{0, g(x)\} \\ f(x) \end{bmatrix},$$

onde  $1_{n,p}$  é o vetor unitário  $n, p$ -dimensional, sendo  $n$  a dimensão da variável e  $p$  o número de restrições de desigualdade. Observe que:

- i. um ponto  $x \in \mathbb{B}$  possui  $z^1 = 0$ , dominando um ponto fora dos limites de caixa, cujo  $z^1 = 1$ ;
- ii. caso ambos pontos pertençam aos limites de caixa  $\mathbb{B}$ , o segundo critério,  $z^2$ , é o nível de violação das restrições do problema: um ponto factível domina um ponto infactível e, entre dois pontos infactíveis, o que tiver menor somatório de violações domina o outro;
- iii. casos ambos pontos sejam factíveis, o terceiro critério,  $z^3$ , é o valor da função objetivo.

Uma vez definidos esses vetores  $z$ , os pontos do simplex são ordenados de tal maneira que

$$z_1 \leq_{\text{lex}} z_2 \leq_{\text{lex}} \dots \leq_{\text{lex}} z_n \leq_{\text{lex}} z_{n+1}.$$

Desse modo, as operações tradicionais do método agora podem ser definidas de acordo com as condições:

- i. **Reflexão:**  $z_r \leq_{\text{lex}} z_{n+1}$ ;
- ii. **Expansão:**  $z_e \leq_{\text{lex}} z_r \leq_{\text{lex}} z_1$ ;
- iii. **Contração externa:**  $z_{oc} \leq_{\text{lex}} z_r \leq_{\text{lex}} z_{n+1}$  ou  $z_{oc} \leq_{\text{lex}} z_{ic}$ ;
- iv. **Contração interna:**  $z_{ic} \leq_{\text{lex}} z_{n+1}$ ;
- v. **Encolhimento:** caso contrário.

O método NM modificado é detalhado no [Algoritmo 3](#). Note que uma vez definida a comparação lexicográfica, a estrutura original do algoritmo é preservada e o tratamento de restrições é incorporado de forma natural.

**Algoritmo 3** Nelder-Mead simplex lexicográfico restrito.**Entrada**

$f$  função objetivo  
 $g$  restrições do problema  
 $\mathbb{B}$  restrições de caixa  
 $\mathbb{S}_n$  conjunto de pontos  $\{x_1, \dots, x_{n+1}\} \in \mathbb{B}$   
 $\rho$  parâmetro de reflexão  
 $\beta$  parâmetro de expansão  
 $\gamma$  parâmetro de contração externa  
 $\delta$  parâmetro de contração interna

- 1: **while** critério de parada não for satisfeito **do**
- 2:    $x_r \leftarrow x_c + \rho(x_c - x_{n+1})$
- 3:   **if**  $z_r \leq_{\text{lex}} z_1$  **then**
- 4:      $x_e \leftarrow x_c + \beta(x_r - x_c)$
- 5:     **if**  $z_e \leq_{\text{lex}} z_1$  **then**
- 6:        $x_{n+1} \leftarrow x_e$  ▷ Expansão
- 7:     **else**
- 8:        $x_{n+1} \leftarrow x_r$  ▷ Reflexão
- 9:     **end if**
- 10:  **else if**  $z_r \leq_{\text{lex}} z_n$  **then**
- 11:     $x_{n+1} \leftarrow x_r$  ▷ Reflexão
- 12:  **else if**  $z_r \leq_{\text{lex}} z_{n+1}$  **then**
- 13:     $x_{oc} \leftarrow x_c + \gamma(x_r - x_c)$
- 14:    **if**  $z_{oc} \leq_{\text{lex}} z_r$  **then**
- 15:       $x_{n+1} \leftarrow x_{oc}$  ▷ Contração externa
- 16:    **else**
- 17:       $x_{n+1} \leftarrow x_r$  ▷ Reflexão
- 18:    **end if**
- 19:  **else**
- 20:     $x_{ic} \leftarrow x_c - \gamma(x_r - x_c)$
- 21:    **if**  $z_{ic} \leq_{\text{lex}} z_{n+1}$  **then**
- 22:       $x_{n+1} \leftarrow x_{ic}$  ▷ Contração interna
- 23:    **else**
- 24:       $x_i \leftarrow x_1 + \delta(x_i - x_1), \forall i \in [2, n + 1]$  ▷ Encolhimento
- 25:    **end if**
- 26:  **end if**
- 27: **end while**

Observe que a comparação lexicográfica é uma alternativa natural para o tratamento de restrições do método NM devido à natureza comparativa entre pontos do mesmo. Além do mais,

o uso de uma estratégia de penalização da função objetivo para comparação entre pontos pode falhar em alguns casos. Como exemplo, considere dois pontos,  $x_1$  e  $x_2$ , de tal forma que a função penalizada

$$F(x_i) = f(x_i) + \Gamma(x_i, \lambda_i),$$

associada a cada ponto  $i$ , seja

$$F(x_1) = f(x_1) + \Gamma(x_1, \lambda_1) = -10 + 10 = 0$$

e

$$F(x_2) = f(x_2) + \Gamma(x_2, \lambda_2) = 0 + 0 = 0.$$

Ou seja, para esse caso, a comparação através do valor de  $F(x)$  diz que um ponto inactivo,  $x_1$ , é tão bom quanto um ponto activo  $x_2$ . De maneira geral, qualquer comparação baseada em  $F(x) = 0$  pode ser dúbia:

$$F(x_3) = f(x_3) + \Gamma(x_3, \lambda_3) = -10^{-3} + 10^{-3} = 0$$

também é um ponto melhor que  $x_1$ , por exemplo.

## 4.2.2 Inicialização do simplex

Em [Luersen et al., 2004], é proposta uma inicialização do simplex baseada na adição de vetores de perturbação ao ponto inicial  $x_0$  de acordo com

$$x_i = x_0 + \nu \xi_i + \sum_{\substack{k=1 \\ k \neq i}}^n \iota \xi_k, \quad i = 1, 2, \dots, n, \quad (4.1)$$

onde

$$\nu = \frac{S}{n\sqrt{2}}(\sqrt{n+1} + n - 1), \quad (4.2)$$

$$\iota = \frac{S}{n\sqrt{2}}(\sqrt{n+1} - 1), \quad (4.3)$$

onde  $S$  é um escalar positivo que controla o tamanho do simplex e  $\xi_i$  é a  $i$ -ésima coluna da matriz identidade  $n$ -dimensional. Um valor comum é  $S = 10$  [Luersen et al., 2004], considerando que as variáveis estejam escaladas entre 0 e 100.

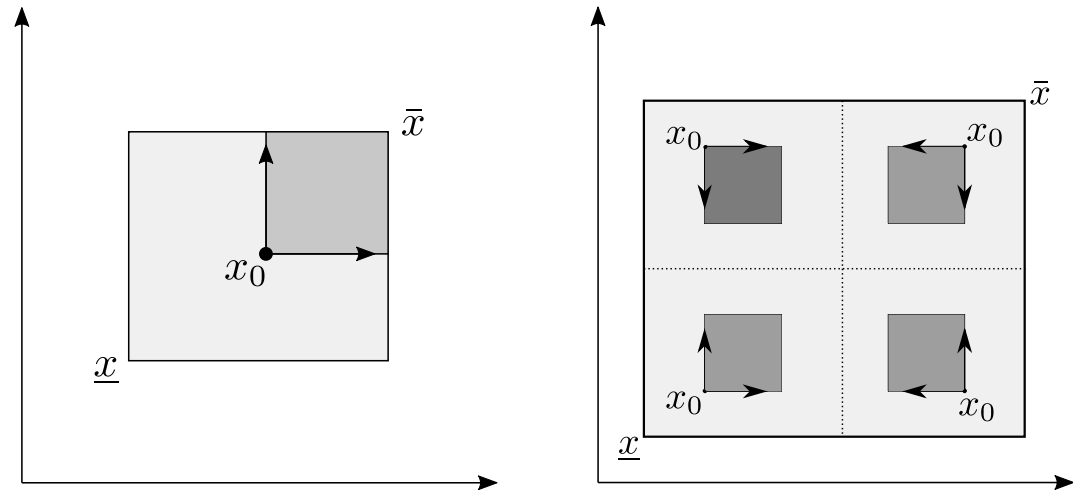
Note que essas perturbações são sempre positivas (ver [Subfigura 4.1a](#)), fazendo com o que os pontos gerados no início possam exceder os limites superiores do espaço de busca, acelerando o processo de degeneração do simplex. Como exemplo, para  $x_0 = \bar{x}$ , todos os pontos gerados por essa estratégia seriam iguais a  $\bar{x}$ , considerando que haja um truncamento dos pontos gerados nos limites das variáveis.

A inicialização proposta aumenta o espaço de busca do simplex inicial, consistindo na alteração do sentido da perturbação adicionada a cada novo ponto gerado de acordo com a posição relativa

de  $x_0$  em relação à média dos limites de caixa de cada dimensão. Isto é, para um dado valor de  $S$ , o  $i$ -ésimo ponto do simplex inicial também é calculado por (4.1), mas agora

$$\xi_i = \begin{cases} \frac{0.5(\bar{x}^i + \underline{x}^i) - x_0^i}{\|0.5(\bar{x}^i + \underline{x}^i) - x_0^i\|} e_i, & \text{se } \|0.5(\bar{x}^i + \underline{x}^i) - x_0^i\| \neq 0, \\ e_i, & \text{caso contrário,} \end{cases} \quad (4.4)$$

onde  $e_i$  é a  $i$ -ésima coluna da matriz identidade  $n$ -dimensional. Note que  $\xi_i$  apresenta dois valores possíveis:  $e_i$  ou  $-e_i$ . Observe ainda que essa estratégia altera o sinal de  $\xi_i$  se a coordenada  $x_0^i$  estiver no metade de cima do espaço delimitado por  $\underline{x}^i$  e  $\bar{x}^i$ , aumentando o espaço de busca do método em casos que haja necessidade de truncamento. A Subfigura 4.1b ilustra a inicialização proposta.



Subfigura (a) A inicialização proposta em [Luersen et al., 2004] adiciona somente componentes positivas a  $x_0$ , direcionando a busca no sentido das componentes mais positivas.

Subfigura (b) Inicialização proposta. A direção das perturbações são definidas de acordo com a posição relativa em relação aos pontos médios de cada dimensão.

Figura 4.1: Estratégias de inicialização do simplex.

Violações de restrições de caixa são tratadas indiretamente pela comparação lexicográfica proposta, sendo esta o primeiro critério ( $z^1$ ) do vetor associado a cada ponto  $x$  do simplex. Desse modo, pontos fora do espaço de busca são preteridos em relação à pontos dentro do espaço de busca, fazendo com que o método modificado proposto absorva o tratamento de restrições de caixa sem a necessidade de uma ação específica (i.e. truncamento), além de garantir que todos os novos pontos gerados pelo método estejam dentro do espaço viável.

Entretanto, é importante ressaltar que o truncamento das variáveis apresentado em (2.10) é aplicado caso o ponto inicial  $x_0$  esteja fora dos limites da caixa.

### 4.3 Experimentos

Esta seção descreve três experimentos usados para exemplificar as características do método NM modificado. O primeiro é baseado na otimização de problemas irrestritos e é usado para ilustrar o comportamento do método em relação à inicialização do simplex proposta. O segundo apresenta o comportamento do método para um problema sintético restrito em diferentes dimensões. Já o terceiro experimento testa o efeito da proporção entre o tamanho da região factível e o tamanho do espaço total de busca na convergência do método.

#### 4.3.1 Experimento I

Para este experimento, é considerado o seguinte problema de otimização:

$$\underset{x}{\text{minimize}} \quad f_1(x) = \max\{f_a(x), f_b(x)\} \quad (4.5)$$

$$\text{sujeito a } \underline{x} \leq x \leq \bar{x}, \quad (4.6)$$

onde

$$f_a(x) = (x - c_1)^T Q (x - c_1) \quad (4.7)$$

$$f_b(x) = (x - c_2)^T Q (x - c_2) \quad (4.8)$$

e

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad c_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{e} \quad c_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

O mínimo de  $f_1(x)$  ocorre no ponto  $x^* = [0, 0]^T$  (ver [Figura 4.2](#)). Note que  $f_1(x)$  apresenta uma descontinuidade “atratora” em  $x^1 = 0$ , dividindo o espaço em duas regiões disjuntas (i.e.,  $x^1 < 0$  e  $x^1 > 0$ ), e que faz com que métodos baseados em avaliações do gradiente fiquem estacionados nessa região [[Takahashi, 2007b](#)].

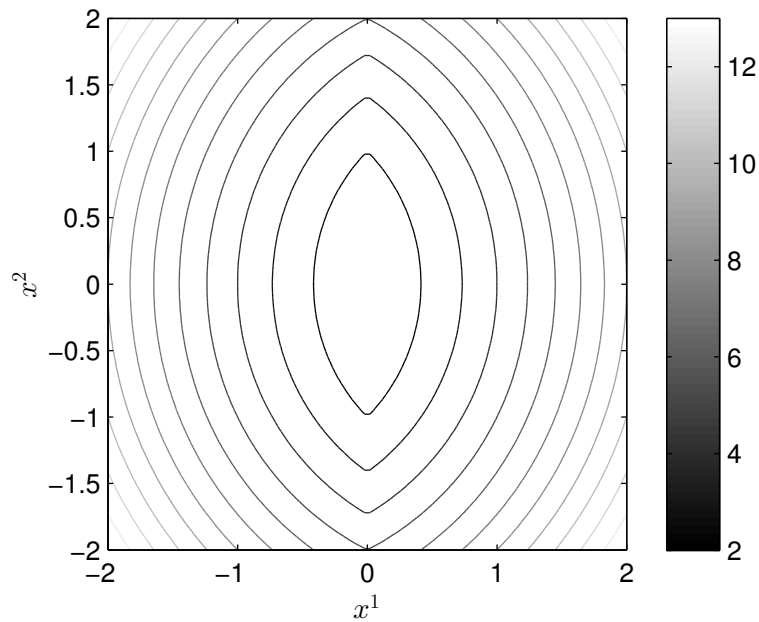


Figura 4.2: Curvas de nível de  $f_1(x)$ , onde  $x^* = [0, 0]^T$ .

Para exemplificar o comportamento da inicialização do simplex proposta, duas instâncias do problema de otimização são criadas:

- Instância 1:  $\underline{x} = [-2, -2]^T$  e  $\bar{x} = [0, 2]^T$ ;
- Instância 2:  $\underline{x} = [0, -2]^T$  e  $\bar{x} = [2, 2]^T$ .

Além disso, são definidas três regiões possíveis para definição aleatória do ponto inicial  $x_0$ . Seja  $u \sim \mathcal{U}(0, 1)$ , onde  $u$  é um escalar aleatório amostrado da distribuição uniforme  $\mathcal{U}(0, 1)$ :

- Região 1 (R1):  $x_0 = [\underline{x}^1, \underline{x}^2 + u(\bar{x}^2 - \underline{x}^2)]^T$ ;
- Região 2 (R2):  $x_0 = [0.5(\underline{x}^1 + \bar{x}^1), \underline{x}^2 + u(\bar{x}^2 - \underline{x}^2)]^T$ ;
- Região 3 (R3):  $x_0 = [\bar{x}^1, \underline{x}^2 + u(\bar{x}^2 - \underline{x}^2)]^T$ .

A [Figura 4.3](#) ilustra as duas instâncias do problema, bem como as três regiões possíveis de inicialização de  $x_0$ . Observe que cada instância do problema contempla umas das regiões disjuntas do problema original e que o ótimo do problema se encontra nos limites da caixa do problema.

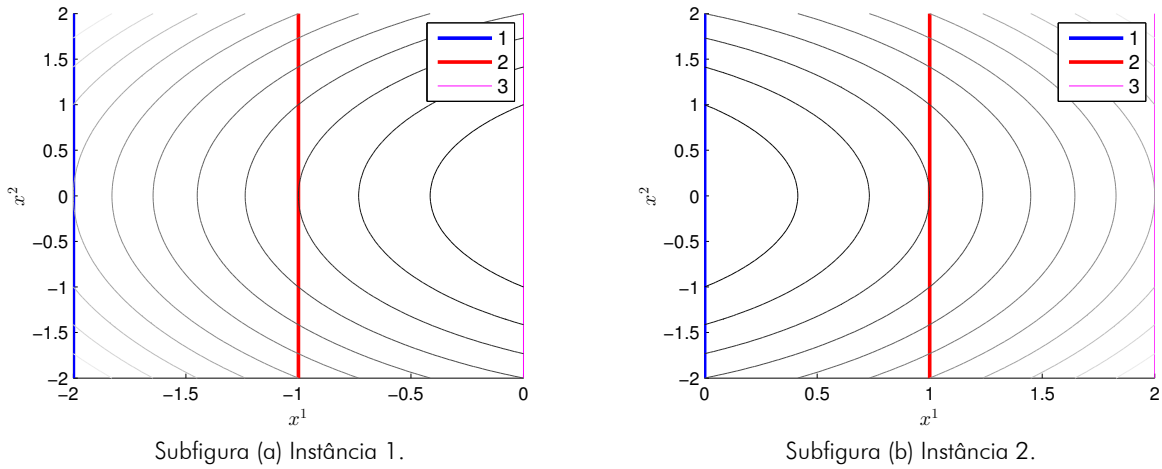


Figura 4.3: Instâncias do problema e as três regiões possíveis de inicialização de  $x_0$ .

Este experimento consiste em 30 inicializações aleatórias de  $x_0$  em cada uma das três regiões, para cada instância do problema, utilizando o método proposto, considerando três formas de determinação do simplex inicial:

- Inicialização 1 (I1): inicialização proposta neste trabalho;
- Inicialização 2 (I2): inicialização utilizada na função “`fminsearch`”, que produz os  $n$  pontos do simplex através de perturbações de +5% em cada componente de  $x_0$  [MathWorks, 2019a];
- Inicialização 3 (I3): inicialização utilizada em [Luersen et al., 2004], na qual este trabalho se baseia, mas que adiciona somente perturbações positivas.

Para todas as inicializações, caso um ponto gerado inicialmente viole as restrições de caixa, esse ponto é truncado conforme (2.10).

A Tabela 4.1 apresenta os resultados das execuções do método em relação ao valor de  $\|x^*\|$ , que idealmente é nulo. Para a instância 1, na qual o ótimo do problema encontra-se na região R3, os resultados obtidos nas regiões R1 e R2 para as três formas de inicialização são equivalentes, isto é, não é observado nenhum ganho significativo de uma inicialização em relação à outra. Já para a região R3, as inicializações I2 e I3 tiveram uma convergência melhor para o ótimo do que a inicialização I1. Como as inicializações I2 e I3 sempre inserem perturbações positivas, o simplex inicial dessas inicializações já começa degenerado, fazendo com que a busca somente aconteça na região R3. Diferentemente da inicialização proposta R1, que também converge para o ótimo, mas com um simplex inicial não-degenerado e que é capaz de explorar as regiões internas da caixa antes de convergir. Já para a instância 2, que possui o ótimo na região R1, é possível observar a maior característica de exploração da região de busca que a inicialização proposta possibilita. Mesmo com  $x_0$  inicializado em R3, o método converge para o ótimo na inicialização I1. Já para as

demais inicializações em R3, a degeneração do simplex inicial faz com que a busca ocorra somente nessa região, impossibilitando a convergência do método para o ótimo do problema.

Além disso, pode-se observar que para a instância 2, mesmo na região R2, a inicialização I3 teve um pior caso muito elevado, muito próximo ao caso em que  $x_0$  localiza-se em R3. Esse valor elevado ocorre quando  $x_0$  encontra-se próximo ao limite superior da caixa, ocasionando um efeito similar ao que ocorre quando o simplex é inicializado na região R3. Nota-se ainda que o pior caso da inicialização proposta, referente à instância 2 e região R1, é aproximadamente 7 vezes menor que o pior caso das demais inicializações do simplex.

Tabela 4.1: Valores de  $\|x^*\|$  para as duas instâncias do problema, três regiões e as três formas de inicialização do simplex, considerando uma precisão numérica de  $10^{-9}$ .

Inicialização	$\ x^*\ $	Instância do problema de otimização					
		Instância 1			Instância 2		
		R1	R2	R3	R1	R2	R3
I1	média	0.0008	0.0003	0.0036	0.028	0.0545	0.0001
	melhor	0.	0.	0.	0.	0.	0.
	pior	0.0194	0.0095	0.0503	0.2742	0.0997	0.0034
	desvio padrão	0.0036	0.0017	0.0103	0.0728	0.1982	0.0006
I2	média	0.0022	0.0003	0.	0.0002	0.0026	2.
	melhor	0.	0.	0.	0.	0.	2.
	pior	0.0391	0.0057	0.	0.0045	0.0696	2.
	desvio padrão	0.0074	0.0011	0.	0.0008	0.0125	0.
I3	média	0.0036	0.0001	0.	0.001	0.0669	2.
	melhor	0.	0.	0.	0.	0.	2.
	pior	0.0848	0.0011	0.	0.0184	1.9858	2.
	desvio padrão	0.0153	0.0002	0.	0.0036	0.3563	0.

### 4.3.2 Experimento II

Nesta seção, um experimento é criado para demonstrar o funcionamento do método proposto, com a inicialização proposta, em instâncias de um problema restrito de diferentes dimensões. Para esse caso, o problema de otimização é dado por

$$\underset{x}{\text{minimize}} \quad 1_n^T \left| \frac{x}{10} \right| \quad (4.9)$$

$$\text{sujeito a} \quad 1_n^T x^3 - 4 \leq 0 \quad (4.10)$$

$$(x - 1_n)^T (x - 1_n) - \frac{1}{\pi} \leq 0 \quad (4.11)$$

$$-10 \leq x \leq 10, \quad (4.12)$$

onde  $n$  é a dimensão do problema, de 2 a 7 (para outros valores de  $n$ , este problema é infactível) e  $1_n$  é o vetor  $n$ -dimensional unitário. Note que (4.9) não é continuamente diferenciável e (4.10) possui pontos de inflexão, tornando o problema difícil de lidar para métodos baseados em avaliações do gradiente.

Para cada instância do problema, 30 inicializações aleatórias são realizadas ao longo de todo espaço de busca. Para cada dimensão do problema, a Tabela 4.2 apresenta o valor ótimo, a melhor solução obtida, a média das soluções, o pior valor e o desvio padrão do valor de função obtido na otimização. O ótimo é a forma aproximada da solução exata dada por

$$f(x^*) = \frac{n}{10} \left( 1 - \frac{1}{\sqrt{n\pi}} \right).$$

Tabela 4.2: Resultados do experimento conduzido para uma precisão numérica de  $10^{-9}$ .

$n$	ótimo	melhor	média	pior	desvio padrão
2	0.1202	0.1202	0.1202	0.1319	0.0021
3	0.2022	0.2022	0.2024	0.2032	0.0002
4	0.2871	0.2871	0.2874	0.3151	0.0049
5	0.3738	0.3738	0.3750	0.4026	0.0067
6	0.4618	0.4618	0.4630	0.4996	0.0090
7	0.5507	0.5508	0.5521	0.5693	0.0056

As colunas de média e desvio padrão mostram a capacidade de convergência do método proposto, assim como foi demonstrado em maiores detalhes no exemplo anterior. Além disso, os valores obtidos no pior caso demonstram que não houve degeneração do simplex inicial, dadas as propriedades matemáticas da função objetivo e que o valor está próximo do ótimo.

### 4.3.3 Experimento III

Este experimento é desenhado para testar a influência da proporção entre o tamanho da região factível e o tamanho do espaço total de busca para o método modificado proposto, denominado Nelder-Mead restrito (NMR). Além disso, os resultados obtidos neste experimento são contrastados com os resultados obtidos ao se aplicar o método dos multiplicados de Lagrange (LA) em conjunto com o método Nelder-Mead tradicional. Para esse fim, o seguinte problema é proposto:

$$\underset{x}{\text{minimize}} \quad 1_n^T |x| \tag{4.13}$$

$$\text{sujeito a} \quad (x - [9, 0]^T)^T (x - [9, 0]^T) \leq c_3 \tag{4.14}$$

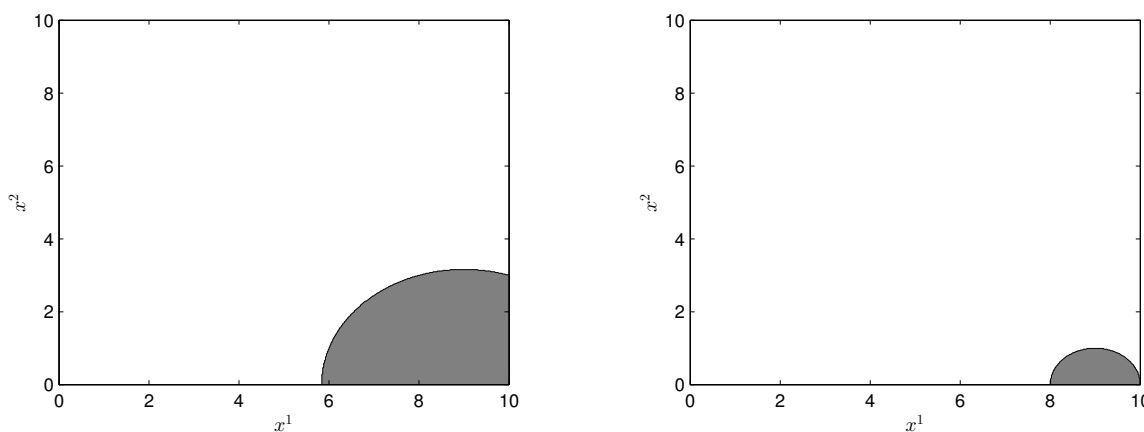
$$0 \leq x \leq 10, \tag{4.15}$$

onde  $x \in \mathbb{R}^2$  e  $c_3 \geq 0$  é um escalar que controla o tamanho do espaço factível (ver Figura 4.4). Note que (4.14) é a equação de uma circunferência de raio  $\sqrt{c_3}$  centrada em  $[9, 0]^T$ . Assim, o

ótimo desse problema ocorre em

$$x^* = [9 - \sqrt{c_3}, 0]^T, \forall 0 \leq c_3 \leq 81, \quad (4.16)$$

que é o ponto mais próximo à origem, que por sua vez é a solução do problema irrestrito.



Subfigura (a) Instância do problema com  $c_3 = 10$ .

Subfigura (b) Instância do problema com  $c_3 = 1$ .

Figura 4.4: Região factível, em cinza, para diferentes valores de  $c_3$ . Observe que  $c_3$  pode ser visto como uma porcentagem da região de busca que é factível.

Assim, o experimento foi conduzido considerando  $c_3 = \{10, 1, 10^{-2}, 10^{-3}, 10^{-5}, 0\}$ , e 30 inicializações aleatórias de  $x_0$  para cada valor. Considerando uma precisão numérica de  $10^{-9}$ , registrou-se na Tabela 4.3 a porcentagem das vezes que o método convergiu para o ótimo do problema e a porcentagem das vezes que o método convergiu para a região factível.

Tabela 4.3: Porcentagem das vezes que o método convergiu para o ótimo e para a região factível do problema para os métodos NMR e LA.

$c_3$	NMR		LA	
	% ótimo	% factível	% ótimo	% factível
$10^1$	90	97	93	100
$10^0$	87	97	93	97
$10^{-2}$	87	97	0	0
$10^{-3}$	100	100	0	0
$10^{-5}$	97	97	3	3
0	93	93	0	0

Os resultados mostram que a convergência do NMR é independente do tamanho do espaço factível para o problema analisado, demonstrando uma boa capacidade de exploração do método restrito proposto, inclusive no limite em que a região factível corresponde a um ponto (i.e.  $c_3 = 0$ ). Além disso, para essas instâncias do problema, o método convergiu para a região factível em pelo

menos 93% das vezes, sendo um indicativo de robustez do algoritmo. A porcentagem das vezes que o método convergiu para o ótimo também é elevada, mas não observou-se nenhum padrão em relação ao valor de  $c_3$ , demonstrando que a estratégia proposta é invariante ao valor de  $c_3$ .

Em contrapartida, o acoplamento da estratégia de penalização usada no método LA com o método Nelder-Mead clássico mostrou ter capacidade de convergência limitada à medida que a região factível diminui. Para valores de  $c_3$  menores que  $10^{-2}$ , por exemplo, essa estratégia de penalização praticamente não convergiu em nenhum dos casos.

#### 4.4 Conclusão do capítulo

Este capítulo apresentou modificações que tornam o Nelder-Mead simplex um método restrito. As modificações inseridas são basicamente duas: inicialização do simplex inicial e ordenação lexicográfica para comparação entre pontos. A primeira modificação faz com o que o simplex inicial não se degenere em decorrência do ponto inicial estar localizado próximo aos limites superiores da caixa do problema. Já a segunda modificação associa um novo vetor  $z \in \mathbb{R}^q$  a cada ponto  $x \in \mathbb{R}^n$ , possibilitando a escrita do método em função de comparações lexicográficas dos  $n + 1$  pontos  $z$ . Através dessas comparações, o tratamento de violações de restrições de caixa e de restrições de desigualdade são naturalmente incorporadas ao método.

Os experimentos conduzidos mostram que a inicialização proposta, de fato, torna o método invariante ao ponto inicial  $x_0$ , evitando a degeneração do simplex inicial, diferentemente de estratégias encontradas na literatura, que possuem raízes na natureza irrestrita da formulação original. Além disso, os experimentos também mostram a boa convergência do método para diferentes dimensões e para diferentes tamanhos de região factível, demonstrando uma boa característica exploratória do método proposto.

## Capítulo 5

# Conclusões

Este trabalho apresentou algumas propostas de melhorias para algoritmos de otimização sem derivadas de busca unidimensional e busca direta  $n$ -dimensional. Na primeira proposta, é introduzido o conceito de padrão- $\wedge$  que, usado em conjunto com o conceito de padrão- $\vee$ , previamente definido, possibilita a criação de um *framework* (*multimodal line search* - MLS) de otimização unidimensional capaz de mapear múltiplos mínimos locais de funções multimodais ao ser acoplado a um método unidimensional sem derivadas. Baseado nesse *framework*, um novo algoritmo, denominado *multimodal golden section* (MGS), é proposto usando o algoritmo da seção áurea. O *framework* proposto herda todas as propriedades de convergência do algoritmo subjacente. Os exemplos realizadas demonstram a natureza multimodal do algoritmo proposto, que também é determinístico e pode ser acoplado a métodos de direção de busca tradicionais.

A segunda proposta faz do Nelder-Mead simplex um método restrito através da implementação de uma comparação lexicográfica que considerada três critérios, em ordem: (i) factibilidade em relação aos limites de caixa, (ii) factibilidade em relação às restrições de desigualdade e (iii) valor de função objetivo. Através dessa comparação, a natureza das funções é preservada, evitando problemas numéricos que ocorrem em estratégias de penalização. Além disso, este trabalho mostra que uma estratégia de penalização pode ser dúbia ao se comparar pontos, e esse ponto pode ser explorado em mais detalhes não só no algoritmo Nelder-Mead simplex, mas em qualquer algoritmo de otimização baseado em comparação de pontos, como é o caso de algoritmos genéticos. Além disso, uma nova forma de inicialização do simplex também é proposta, evitando a degeneração do simplex ao se inicializar o ponto inicial próximo aos limites superiores do espaço de busca. Os experimentos conduzidos demonstram as características de convergência do método proposto, tornando-o robusto à inicialização e ao tratamento de restrições de desigualdade. Como possibilidade de trabalho futuro nessa parte, há a necessidade de estender o método para o tratamento de restrições de igualdade e problemas de otimização multiobjetivo.

## Apêndice A

# Algoritmos base de otimização

### A.1 Algoritmo da seção áurea

O algoritmo da seção áurea é baseado em uma sequência ordenada de pontos  $[\alpha_1, \alpha_2, \alpha_3, \alpha_4]$  e na comparação sucessiva entre os pontos  $\alpha_2$  e  $\alpha_3$ , de forma que uma nova sequência seja produzida na iteração  $k + 1$  através da eliminação ou do intervalo  $(\alpha_3, \alpha_4]$ , ou do intervalo  $[\alpha_1, \alpha_2)$ . O [Algoritmo 4](#) apresenta a estrutura básica, note que ele já possui a modificação proposta por [Vieira and Lisboa, 2014](#), que engloba uma comparação extra (i.e.,  $\min\{f(\alpha_2), f(\alpha_3)\} \geq f(\underline{\alpha})$ ) na linha 11 do algoritmo. Esta verificação adicional garante que  $f(\alpha^*) \leq f(\underline{\alpha})$ .

**Teorema 9** (Taxa de convergência do algoritmo da seção áurea). *O algoritmo da seção áurea converge em no máximo  $(\log \epsilon / \log \Phi)$  iterações, sendo  $\epsilon$  a precisão numérica adotada e  $\Phi = (\sqrt{5}-1)/2$  a razão áurea.*

*Demonstração.* A cada iteração  $k$ , a redução do intervalo é de  $\Delta_k/\Delta_{k-1} = \Phi$ , sendo  $\Delta$  a largura do intervalo. Portanto, após  $k$  iterações a redução agregado do intervalo é de  $\Delta_k/\Delta_0 = \Phi^k$ . Como o único critério de parada é  $\epsilon < \Delta_k/\Delta_0 = \Phi^k$ , o número máximo de iterações é  $\log \epsilon / \log \Phi$ . [[Vieira and Lisboa, 2014](#)] □

---

**Algoritmo 4** Algoritmo da seção áurea.

---

**Entrada**

$f$  função objetivo  
 $[\underline{\alpha}, \bar{\alpha}]$  intervalo de busca

**Saída**

$\alpha^*$  passo ótimo

```

1: function ENHANCEDGOLDENSECTION( $f, \underline{\alpha}, \bar{\alpha}$ )
2:    $\Phi \leftarrow (\sqrt{5} - 1)/2$  ▷ razão áurea:  $\Phi = (\sqrt{5} - 1)/2$ 
3:    $k \leftarrow 0$  ▷ iteração
4:    $\Delta_0 \leftarrow \bar{\alpha} - \underline{\alpha}$ 
5:    $\alpha_1 \leftarrow \underline{\alpha}$ 
6:    $\alpha_2 \leftarrow \bar{\alpha} - \Phi\Delta_0$ 
7:    $\alpha_3 \leftarrow \underline{\alpha} + \Phi\Delta_0$ 
8:    $\alpha_4 \leftarrow \bar{\alpha}$ 
9:   while  $\Delta_k > \epsilon\Delta_0$  do ▷  $\epsilon \in (0, 1]$ 
10:     $k \leftarrow k + 1$ 
11:    if  $\min\{f(\alpha_2), f(\alpha_3)\} \geq f(\underline{\alpha})$  or  $f(\alpha_2) \leq f(\alpha_3)$  then
12:       $\Delta_k \leftarrow \alpha_3 - \alpha_1$ 
13:       $\alpha_4 \leftarrow \alpha_3$ 
14:       $\alpha_3 \leftarrow \alpha_2$ 
15:       $\alpha_2 \leftarrow \alpha_4 - \Phi\Delta_k$ 
16:    else ▷ elimina  $[\alpha_1, \alpha_2]$ 
17:       $\Delta_k \leftarrow \alpha_4 - \alpha_2$ 
18:       $\alpha_1 \leftarrow \alpha_2$ 
19:       $\alpha_2 \leftarrow \alpha_3$ 
20:       $\alpha_3 \leftarrow \alpha_1 + \Phi\Delta_k$ 
21:    end if
22:  end while
23:   $\alpha^* \leftarrow (\alpha_2 + \alpha_3)/2$ 
24: end function

```

---

## A.2 Algoritmo Nelder-Mead simplex

O algoritmo Nelder-Mead (NM) [Nelder and Mead, 1965] foi inicialmente proposto como um método irrestrito e possui convergência assegurada para funções estritamente convexas [Lagarias et al., 1998]. O método gera, a princípio, um conjunto  $\mathbb{S}_n = \{x_1, x_2, \dots, x_n, x_{n+1}\}$  com  $n + 1$  pontos para um problema  $n$ -dimensional. Os pontos são então ordenados de modo que

$$f(x_1) \leq f(x_2) \leq \dots \leq f(x_n) \leq f(x_{n+1}).$$

Seja

$$x_c = \frac{1}{n} \sum_{i=1}^n x_i$$

o centróide dos  $n$  melhores pontos (i.e. como menor valor de função). O algoritmo tenta substituir o pior ponto  $x_{n+1}$  como um novo ponto obtido através de uma das seguintes operações:

- i. **Reflexão:**  $x_r = x_c + \rho(x_c - x_{n+1})$ ;
- ii. **Expansão:**  $x_e = x_c + \beta(x_r - x_c)$ ;
- iii. **Contração externa:**  $x_{oc} = x_c + \gamma(x_r - x_c)$ ;
- iv. **Contração interna:**  $x_{ic} = x_c - \gamma(x_r - x_c)$ ;
- v. **Encolhimento:**  $x_i = x_1 + \delta(x_i - x_1), \forall i \in [2, n + 1]$ .

A princípio, qualquer ponto  $(x_r, x_e, x_{oc}, x_{ic})$ , definidos pelas operações i-iv, são candidatos a substituir  $x_{n+1}$ , e os parâmetros comumente utilizados são  $(\rho, \beta, \gamma, \delta) = (1, 2, 0.5, 0.5)$ . A operação v altera todos os pontos, reduzindo o tamanho do simplex no sentido do melhor ponto. As condições para cada operação estão descritas em maiores detalhes no [Algoritmo 5](#).

---

**Algoritmo 5** Método Nelder-Mead simplex.

---

**Entrada**

$f$  função objetivo  
 $\mathbb{S}_n$  conjunto de pontos  $\{x_1, \dots, x_{n+1}\} \in \mathbb{B}$   
 $\rho$  parâmetro de reflexão  
 $\beta$  parâmetro de expansão  
 $\gamma$  parâmetro de contração externa  
 $\delta$  parâmetro de contração interna

- 1: **while** critério de parada não for satisfeito **do**
- 2:    $x_r \leftarrow x_c + \rho(x_c - x_{n+1})$
- 3:   **if**  $f(x_r) < f(x_1)$  **then**
- 4:      $x_e = x_c + \beta(x_r - x_c)$
- 5:     **if**  $f(x_e) < f(x_r)$  **then**
- 6:        $x_{n+1} = x_e$  ▷ Expansão
- 7:     **else**
- 8:        $x_{n+1} = x_r$  ▷ Reflexão
- 9:     **end if**
- 10:  **else if**  $f(x_r) < f(x_n)$  **then**
- 11:     $x_{n+1} = x_r$  ▷ Reflexão
- 12:  **else if**  $f(x_r) < f(x_{n+1})$  **then**
- 13:     $x_{oc} = x_c + \gamma(x_c - x_{n+1})$
- 14:    **if**  $f(x_{oc}) \leq f(x_{n+1})$  **then**
- 15:       $x_{n+1} = x_{oc}$  ▷ Contração externa
- 16:    **end if**
- 17:  **else if**  $f(x_r) \geq f(x_{n+1})$  **then**
- 18:     $x_{ic} = x_c - \gamma(x_c - x_{n+1})$
- 19:    **if**  $f(x_{ic}) \leq f(x_{n+1})$  **then**
- 20:       $x_{n+1} = x_{ic}$  ▷ Contração interna
- 21:    **end if**
- 22:  **else**
- 23:     $x_i = x_1 - \delta(x_i - x_1), \forall i \in [2, n + 1]$  ▷ Encolhimento
- 24:  **end if**
- 25: **end while**

---

# Referências Bibliográficas

- Afonso, M. V., Bioucas-Dias, J. M., and Figueiredo, M. A. (2010). An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681--695.
- Armijo, L. (1966). Minimization of functions having Lipschitz conditions for partial derivatives. *Pacific Journal of Mathematics*, 16:1--3.
- Avriel, M. and Wilde, D. J. (1966). Optimality proof for the symmetric fibonacci search technique. *Fibonacci Quarterly*, 4:265--269.
- Bazaraa, M. S. and Shetty, C. M. (2006). *Nonlinear Programming - Theory and Algorithms*. Wiley Interscience.
- Bertsekas, D. P. (2008). *Nonlinear Programming*. Athena Scientific.
- Bertsekas, D. P. (2014). *Constrained optimization and Lagrange multiplier methods*. Academic press.
- Brent, R. P. (1973). *Algorithms for minimization without derivatives*. P. Hall.
- Broyden, C. G. (1965). A class of methods for solving nonlinear simultaneous equations. *Mathematics of computation*, 19(92):577--593.
- Courant, R. et al. (1943). *Variational methods for the solution of problems of equilibrium and vibrations*. Verlag nicht ermittelbar.
- Curry, H. B. (1944). The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2(3):258--261.
- Djikpéssé, H., Couët, B., and Wilkinson, D. (2011). A practical sequential lexicographic approach for derivative-free black-box constrained optimization. *Engineering Optimization*, 43(7):721--739.
- Ehrgott, M. (2005). *Multicriteria optimization*, volume 491. Springer Science & Business Media.
- Fletcher, R. and Powell, M. J. (1963). A rapidly convergent descent method for minimization. *The computer journal*, 6(2):163--168.
- Goldstein, A. (1965). On steepest descent. *SIAM J. Control*, 3:147--151.

- Gramacy, R. B., Gray, G. A., Le Digabel, S., Lee, H. K., Ranjan, P., Wells, G., and Wild, S. M. (2016). Modeling an augmented lagrangian for blackbox constrained optimization. *Technometrics*, 58(1):1--11.
- Hestenes, M. R. (1969). Multiplier and gradient methods. *Journal of optimization theory and applications*, 4(5):303--320.
- Hooke, R. and Jeeves, T. A. (1961). Direct search solution of numerical and statistical problems. *Journal of the ACM (JACM)*, 8(2):212--229.
- Houska, B., Frasch, J., and Diehl, M. (2016). An augmented lagrangian based algorithm for distributed nonconvex optimization. *SIAM Journal on Optimization*, 26(2):1101--1127.
- Kiefer, J. (1953). Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, 4(3):502--506.
- Kolda, T. G., Lewis, R. M., Torczon, V., et al. (2006). A generating set direct search augmented lagrangian algorithm for optimization with a combination of general and linear constraints. *Sandia National Laboratories*, 6.
- Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the nelder--mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112--147.
- Larson, J., Menickelly, M., and Wild, S. M. (2019). Derivative-free optimization methods. *Acta Numerica*, 28:287--404.
- Luenberger, D. and Ye, Y. (2010). *Linear and Nonlinear Programming*. Springer.
- Luenberger, D. G., Ye, Y., et al. (1984). *Linear and nonlinear programming*, volume 2. Springer.
- Luersen, M. A., Riche, R. L., Guyon, F., Le Riche, R., and Guyon, F. (2004). A constrained, globalized, and bounded Nelder-Mead method for engineering optimization. *Structural and Multidisciplinary Optimization*, 27(1-2):43--54.
- Luus, R. and Jaakola, T. (1973). Optimization by direct search and systematic reduction of the size of search region. *AICHE Journal*, 19(4):760--766.
- MathWorks (2019a). Find minimum of unconstrained multivariable function using derivative-free method. <https://www.mathworks.com/help/matlab/ref/fminsearch.html>. Acessado em: 2019-09-30.
- MathWorks (2019b). What is direct search? <https://www.mathworks.com/help/gads/what-is-direct-search.html>. Acessado em: 2019-08-30.
- Matyas, J. (1965). Random optimization. *Automation and Remote control*, 26(2):246--253.
- Mendonça, M. et al. (2019a). Análise comparativa entre modelos de inteligência computacional para previsão do preço futuro no mercado de energia brasileiro. *Seminário Nacional de Produção e Transmissão de Energia Elétrica*.

- Mendonça, M. et al. (2019b). Estratégia para tratamento de restrições no algoritmo nelder-mead com aplicações na otimização da programação diária da operação de usinas hidrelétricas. *Simpósio de Engenharia de Produção*.
- Mendonça, M. et al. (2019c). Heurística para unit commitment de turbinas hidráulicas distintas que visa maximizar a eficiência de geração de uma usina hidrelétrica e minimizar o ligamento/-desligamento de máquinas considerando indisponibilidades programadas. *Seminário Nacional de Produção e Transmissão de Energia Elétrica*.
- Mendonça, M. et al. (2019d). Programação diária da operação considerando incertezas de defluência de usinas a montante operadas por outras empresas. *Seminário Nacional de Produção e Transmissão de Energia Elétrica*.
- Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308--313.
- Pedroso, L. G., Diniz-Ehrhardt, M. A., and Martínez, J. M. (2011). Derivative-free methods for non-linear programming with general lower-level constraints. *Computational & Applied Mathematics*, 30(1).
- Powell, M. J. (1969). A method for nonlinear constraints in minimization problems. *Optimization*, pages 283--298.
- Rios, L. M. and Sahinidis, N. V. (2013). Derivative-free optimization: A review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247--1293.
- Rockafellar, R. T. (1973). A dual approach to solving nonlinear programming problems by unconstrained optimization. *Mathematical programming*, 5(1):354--373.
- Shi, Z. J. and Shen, J. (2005). New inexact line search method for unconstrained optimization. *Journal of Optimization Theory and Applications*, 2:425--446.
- Takahashi, R. (2007a). Otimização escalar e vetorial - volume i. <http://150.164.25.15/~taka/index.html>.
- Takahashi, R. (2007b). Otimização escalar e vetorial - volume i. <http://150.164.25.15/~taka/index.html>.
- Vieira, D. A. G. and Lisboa, A. C. (2014). Line search methods with guaranteed asymptotical convergence to an improving local optimum of multimodal functions. *European Journal of Operational Research*, 235(1):38 -- 46.
- Vieira et al., D. A. G. (2012). Multicriteria optimization with a multiobjective golden section line search. *Mathematical Programming*, 131(1-2):131--161.
- Wolfe, P. (1969). Convergence conditions for ascent methods. *SIAM Review*, 1:226--235.
- Yuan et. al, G. (2010). A line search algorithm for unconstrained optimization. *Journal of Software Engineering and Applications*, 3(5):503--509.

Ficha catalográfica: Biblioteca Profº Mário Werneck, Escola de Engenharia da UFMG